

Jorge Castro López

Estudio del reconocimiento del  
antígeno Tn por  
anticuerpos/lecitinas para el  
desarrollo de vacunas y nuevas  
herramientas de diagnóstico frente  
al cáncer

Departamento  
Bioquímica y Biología Molecular y Celular

Director/es  
Hurtado Guerrero, Ramon

<http://zaguan.unizar.es/collection/Tesis>



Reconocimiento – NoComercial – SinObraDerivada (by-nc-nd): No se permite un uso comercial de la obra original ni la generación de obras derivadas.

© Universidad de Zaragoza  
Servicio de Publicaciones

ISSN 2254-7606

Tesis Doctoral

ESTUDIO DEL RECONOCIMIENTO DEL ANTÍGENO  
TN POR ANTICUERPOS/LECITINAS PARA EL  
DESARROLLO DE VACUNAS Y NUEVAS  
HERRAMIENTAS DE DIAGNÓSTICO FRENTE AL  
CÁNCER

Autor

Jorge Castro López

Director/es

Hurtado Guerrero, Ramon

**UNIVERSIDAD DE ZARAGOZA**  
Bioquímica y Biología Molecular y Celular

2020



# ESTUDIO DEL RECONOCIMIENTO DEL ANTÍGENO T<sub>n</sub> POR ANTICUERPOS/LECTINAS PARA EL DESARROLLO DE VACUNAS Y NUEVAS HERRAMIENTAS DE DIAGNÓSTICO FRENTE AL CÁNCER

Memoria presentada por **Jorge Castro López**

licenciado en Bioquímica, para optar al grado de Doctor en Ciencias.

Zaragoza, octubre de 2019



D. RAMÓN HURTADO GUERRERO, investigador ARAID del Instituto de Biocomputación y Física de Sistemas Complejos (BIFI) de la Universidad de Zaragoza,

CERTIFICA:

Que la tesis doctoral Estudio del reconocimiento del antígeno Tn por anticuerpos/lectinas para el desarrollo de vacunas y nuevas herramientas de diagnóstico frente al cáncer ha sido realizada, bajo su dirección, por el licenciado JORGE CASTRO LOPEZ, en el Instituto de Biocomputación y Física de Sistemas Complejos y que reúne, a su juicio, las condiciones requeridas para optar al grado de Doctor en Ciencias por la Universidad de Zaragoza.

Zaragoza, octubre de 2019.

La realización de esta tesis doctoral ha sido posible gracias a la financiación del Ministerio de Economía y Competitividad del Gobierno de España (BFU2010-19504, BFU2016-75633-P y CTQ2013-44367-C2-2-P), de la Agencia Aragonesa para la Investigación y Desarrollo (ARAIID) y de la Diputación General de Aragón (proyectos B89 y B18). Se ha colaborado también con los sincrotrones DLS, en Oxford (líneas I03 e I04) y ALBA, en Barcelona (línea BL-13 XALOC).

Durante los cuatro años de realización de esta tesis (desde abril de 2015 hasta abril de 2019, ambos incluidos), el doctorando ha sido beneficiario de una ayuda para contratos predoctorales por parte del Ministerio de Economía y competitividad del Gobierno de España (BES-2014-070941)



# Contenido

INTRODUCCION .....	9
1. La glicosilación de proteínas .....	10
2. Mucinas .....	13
2.1. Estructura de las mucinas.....	15
2.2. Relación de las mucinas con enfermedades.....	18
3. Antígenos Tn y STn.....	22
3.1. ¿Qué son?.....	22
3.1.1. Antígeno Tn.....	23
3.1.2. Antígeno STn .....	24
3.2. ¿Cómo y por qué se producen?.....	24
3.2.1. Antígeno Tn.....	25
3.2.2. Antígeno STn .....	26
3.3. Formas de tratamiento frente a cánceres que presentan el antígeno Tn y STn. 27	
3.3.1. Lectinas frente a los antígenos Thomsen-Friedenreich .....	28
3.3.2. Anticuerpos frente a antígenos Thomsen-Friedenreich.....	32
OBJETIVOS.....	35
MATERIALES Y MÉTODOS .....	37
LECTINA .....	38
2.1 PROCESO DE PURIFICACIÓN DE LA LECTINA SOYBEAN AGLUTTININ (SBA)      38	
2.1.1 Purificación de la Lectina (SBA).....	38
2.1.2 Cuantificación de proteínas por absorción UV-Visible .....	39
2.1.3 Electroforesis desnaturizante SDS-PAGE y tinción .....	41
2.1.4 Diálisis y concentración de la proteína.....	42
2.2 TECNICAS BIOFISICAS PARA ESTUDIOS DE INTERACCIÓN.....	44
2.2.1 Calorimetría de Titulación Isotérmica (ITC) .....	44

2.2.2	Resonancia magnética de diferencia de transferencia de saturación (STD)	46
2.2.3	Ensayo de unión Lectina -Enzima (ELLA).....	48
2.3	TÉCNICAS BIOFÍSICAS DE ANÁLISIS ESTRUCTURAL .....	49
2.3.1	Cristalización de proteínas.....	49
2.3.2	Co-cristalización de proteínas.....	50
2.3.3	Toma de cristales y congelación .....	51
2.3.4	Difracción de rayos X .....	52
2.3.5	Análisis de difracción .....	54
2.4	PROCESOS DE SÍNTESIS DE GLICOPÉPTIDOS.....	56
2.4.1	Síntesis SPPS, métodos Fmoc y Boc .....	56
2.4.2	Purificación de péptidos y glicopéptidos por HPLC .....	58
2.4.3	Disolución de los glicopéptidos .....	59
2.5	OTRAS TÉCNICAS PARA EL ANÁLISIS DE LAS ESTRUCTURAS MOLECULARES .....	59
2.5.1	Experimentos de 2D RMN (COSY, HSQC y NOESY).....	59
2.6	TECNICAS COMPUTACIONALES.....	61
2.6.1	Simulaciones de Dinámica Molecular .....	61
	PROTEINA 1SM3_ScFv.....	62
2.7	TÉCNICAS BIOQUÍMICAS .....	62
2.7.1	Elección del fragmento ScFv para su estudio .....	62
2.7.2	Clonajes del fragmento ScFv del anticuerpo 1SM3.....	64
2.8	CULTIVOS DE BIOMASA, EXPRESIÓN Y RECOLECCIÓN DE LA PROTEÍNA 1SM3_ScFv EN <i>E. coli</i> .....	73
2.8.1	Purificación de la proteína 1SM3_ScFv .....	75
2.9	CULTIVOS DE BIOMASA, EXPRESIÓN Y RECOLECCIÓN DE LA PROTEÍNA 1SM3_ScFv EN <i>P. PASTORIS</i> .....	80
2.9.1	OBTENCIÓN DE CÉLULAS COMPETENTES DE <i>PICHIA PASTORIS</i> Y TRANSFORMACIÓN POR ELECTROPORACIÓN .....	80
2.9.2	PRUEBA DE EXPRESIÓN DE COLONIAS POSITIVAS DE LA ELECTROPORACIÓN.....	82
2.9.3	Cultivo de biomasa, expresión y recolección de la proteína 1SM3_ScFv ...	84

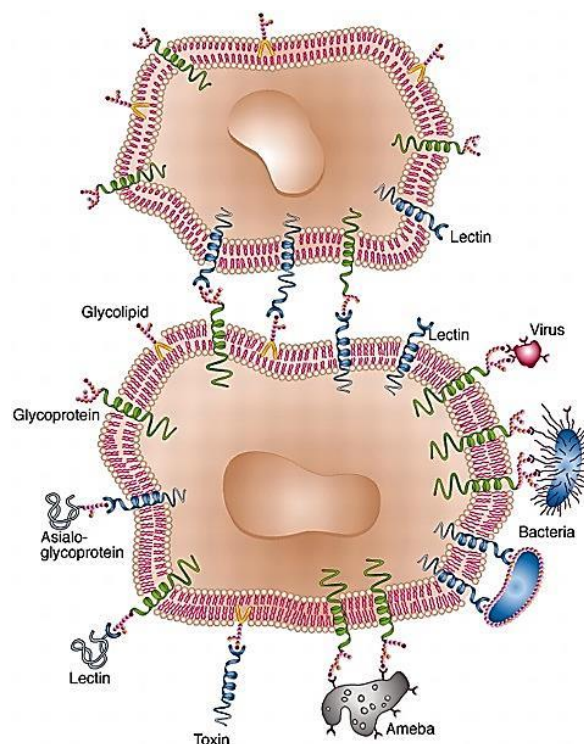
2.10	TECNICAS BIOFISICAS PARA EL ESTUDIO DE INTERACCIÓN .....	86
2.10.1	Calorimetría de Titulación Isotérmica (ITC) .....	86
2.10.2	Interferometria de Biocapa (BLI) .....	87
2.10.3	Ensayo ELISA .....	87
2.10.4	Espectroscopía de Fluorescencia .....	88
2.10.5	Microarrays .....	89
2.11	TÉCNICAS BIOFÍSICAS DE ANÁLISIS ESTRUCTURAL .....	91
2.11.1	Cristalización 1SM3_ScFv y co-cristalización con sus péptidos .....	91
2.11.2	Experimentos RMN .....	91
2.11.3	Dicroísmo Circular (DC) .....	91
2.11.4	IRID .....	93
2.12	TECNICAS COMPUTACIONALES .....	94
2.12.1	Cálculos DFT .....	94
2.12.2	Simulaciones de Dinámica Molecular .....	94
2.13	ENSAYOS EN CÉLULAS Y PRUEBAS EN PACIENTES .....	94
	ARTICULOS .....	95
	Article I: Detection of Tumor-Associated Glycopeptides by Lectins: The Peptide Context Modulates Carbohydrate Recognition .....	96
3.4.	ABSTRACT .....	97
3.5.	RESULTS AND DISCUSSION .....	99
3.6.	METHODS .....	111
	Article II: Deciphering the Non-Equivalence of Serine and Threonine O-Glycosylation Points: Implications for Molecular Recognition of the Tn Antigen by an anti-MUC1 Antibody** .....	122
3.7.	ABSTRACT .....	124
	Article III: Design of $\alpha$ -S-Neoglycopeptides Derived from MUC1 with a Flexible and Solvent-Exposed Sugar Moiety .....	132
3.8.	ABSTRACT .....	133
3.9.	INTRODUCTION .....	134
3.10.	RESULTS AND DISCUSSION .....	136
3.11.	CONCLUSIONS .....	149

Article IV: The Use of Fluoroproline in MUC1 Antigen Enables Efficient Detection of Antibodies in Patients with Prostate Cancer.....	179
3.12. ABSTRACT.....	180
3.13. INTRODUCTION .....	181
3.14. RESULTS AND DISCUSSION .....	184
3.15. CONCLUSIONS.....	190
Article V: Water Sculpts the Distinctive Shapes and Dynamics of the Tumor-Associated Carbohydrate Tn Antigens: Implications for Their Molecular Recognition .....	195
3.16. ABSTRACT.....	196
3.17. INTRODUCTION .....	197
3.18. RESULTS AND DISCUSSION .....	200
3.19. CONCLUSIONS.....	210
Article VI: Structure-Based Design of Potent Tumor-Associated Antigens: Modulation of Peptide Presentation by Single-Atom O/S or O/Se Substitutions at the Glycosidic Linkage .....	215
3.20. ABSTRACT: .....	216
3.21. INTRODUCTION .....	217
3.22. RESULTS AND DISCUSSION .....	218
3.23. CONCLUSIONS.....	230
DISCUSION Y CONCLUSIONES .....	238
REFERENCIAS .....	245
BIBLIOGRAFICAS .....	245
ABREVIATURAS.....	253
INFORMACIÓN GENERAL .....	256

# **INTRODUCCION**

## 1. La glicosilación de proteínas

La glicosilación de proteínas es un proceso bioquímico en el que se adiciona un monosacárido a una molécula de proteína (Li, 2019). Además, otras moléculas como carbohidratos, lípidos, ADN o moléculas pequeñas orgánicas pueden modificarse mediante la adición de un monosacárido (Schjoldager y Clausen, 2012.). Estos procesos se llevan a cabo por medio de las glicosiltransferasas y en el caso concreto de las proteínas esta modificación tiene un papel fundamental en la célula ya que participa activamente en las interacciones célula-célula, célula-matriz y célula-patógeno (Crocker, 1996; Feizi, 2000) (Figura 1.1). La glicosilación de proteínas es una de las modificaciones postraduccionales más abundantes y diversas en la naturaleza. Puede darse una vez las proteínas han sido sintetizadas por los ribosomas o bien puede darse en paralelo a la síntesis de la proteína cuando el ribosoma se encuentra asociado al retículo endoplásmico. Se estima que más de la mitad de todas las proteínas humanas están glicosiladas.



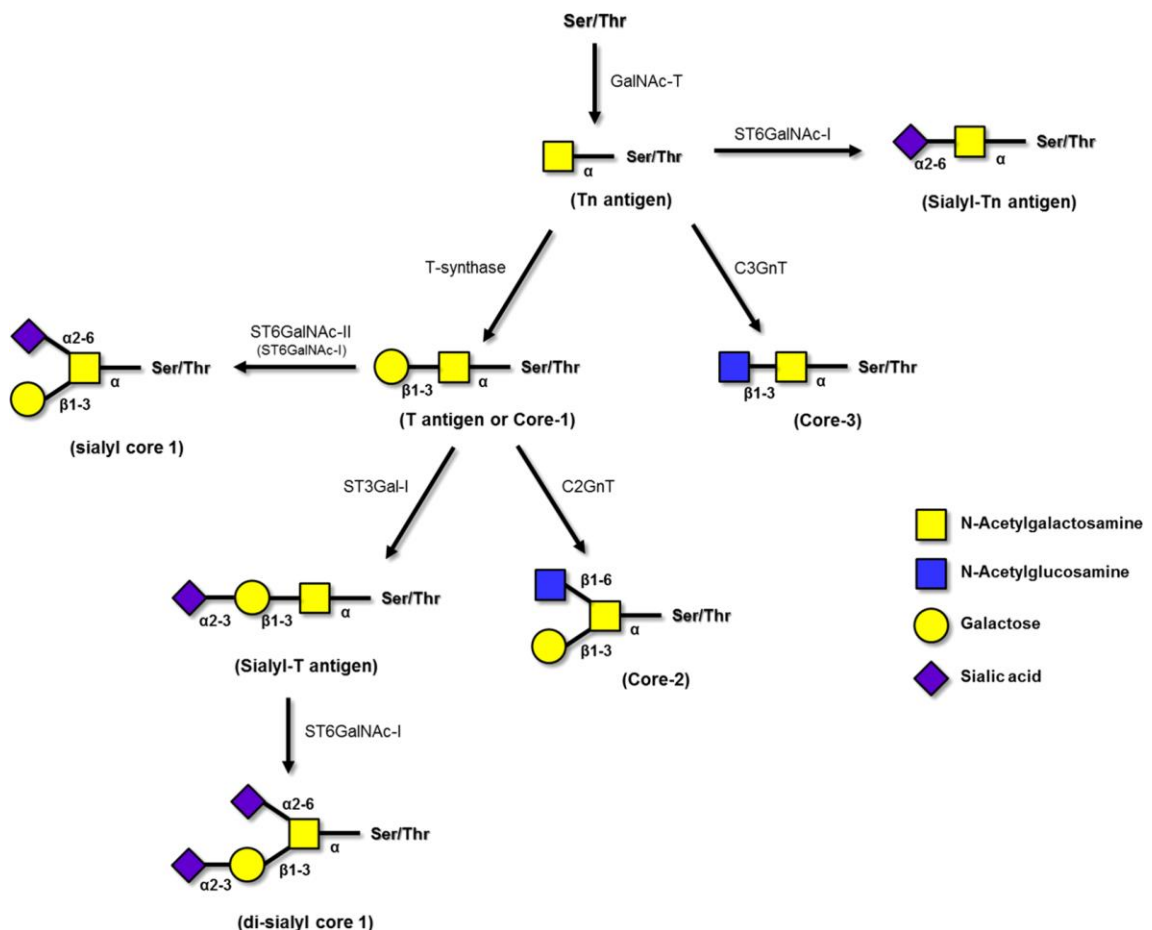
**Figura 1.1. Interacciones de los glicanos celulares.** Representación esquemática de los distintos tipos de interacciones que pueden experimentar los glicanos de una célula  
**Fuente:** Shanon y Lis, 2006

Hay varios tipos de glicosilación en proteínas: *O*-glicosilación y *N*-glicosilación. La principal diferencia es el tipo de residuo al que los carbohidratos se unen y el tipo de enlace que se forma. En el caso de la *O*-glicosilación se produce la transferencia de un azúcar al grupo hidroxilo de los aminoácidos treonina (Thr) y serina (Ser), mientras que para la *N*-glicosilación la transferencia del azúcar se hace al grupo amino de la amida de las asparaginas (Asn) (Kudelka, 2015). Cabe mencionar que también se ha encontrado que tanto los grupos hidroxilo de tirosinas como los que se forman en lisinas (hidroxilisinas) son también modificados por la adición de carbohidratos (Steentoft *et al.*, 2011). Como la biosíntesis de glicanos no es un proceso basado en plantillas como en el caso de la síntesis de proteínas, su expresión va a depender de varios factores: (a) equilibrio alcanzado por los niveles de expresión y actividad de las diferentes enzimas involucradas en el proceso de glicosilación; (b) localización/organización de estas enzimas; y (c) disponibilidad de moléculas precursoras de monosacáridos (Tucillo, 2014).

A mediados de la década de 1990 se produjo un aumento en el interés por el estudio de la glicosilación porque se observó una correlación significativa entre ciertos tipos de glicosilación alterada y el pronóstico del cáncer. De hecho, la glicosilación aberrante ocurre con frecuencia en el cáncer y juega un papel fundamental en la progresión del cáncer, la angiogénesis y la metástasis, el contacto entre células, la motilidad y la transición epitelial-mesenquimatosa (EMT) (Hakomori, 2001).

Estas alteraciones de la glicosilación comprenden la subexpresión o la sobreexpresión de estructuras de glicanos, así como la aparición de estructuras nuevas o truncadas. Sin embargo, a pesar de la diversidad de glicanos en la superficie celular, solo unas pocas estructuras distintas están asociadas con la transformación maligna y la progresión tumoral, como es el caso del incremento de la exposición de residuos de *N*-acetilgalactosamina (GalNAc) y de otros azúcares. Esto sugiere que los patrones específicos de glicosilación contribuyen a mecanismos precisos como el aumento de la función, la aptitud celular y la supervivencia (Varki, 2009).

Estos antígenos con GalNAc se denominan de diferente manera según el GalNAc esté sólo unido a residuos de Ser/Thr o bien vaya unido también a otros azúcares. De esta manera tenemos el antígeno T o Thomsen-Friedenreich (TF), el antígeno Thomsen-nouveau (Tn), el antígeno STn y el sialil T (ST) (Figura 1.2; ver más abajo en la introducción para una información más detallada sobre alguno de estos antígenos y cómo se descubrieron). Se ha propuesto que las moléculas que contienen galactosas terminales tienen la función de facilitar la unión de las células tumorales circulantes a las células del endotelio vascular por unión de estas galactosas a proteínas de unión a carbohidratos en la superficie de las células (Taylor, 2009; Wandall, 2014).



**Figura 1.2. Biosíntesis de las cadenas de O-glicanos.** Derivados de la glicosilación en Ser/Thr. Los antígenos de la familia Thomsen-Friedenreich también son conocidos como antígenos de carbohidratos asociados a tumores (TACA). (Loureiro, 2018)



Una de las familias más grandes de glicosiltransferasas (GTs) son las GalNAc-Ts, las cuales comprenden unas 20 isoenzimas en humanos que se encargan de transferir residuos de N-acetilgalactosamina (GalNAc) a proteínas y por tanto inician la O-glicosilación tipo GalNAc (en lo sucesivo, O-glicosilación). Las GalNAc-Ts pueden modificar una gran cantidad de proteínas ubicadas en la membrana plasmática, así como secretadas, y son las únicas GTs con un dominio lectina tipo ricina (las Rivas, 2019).

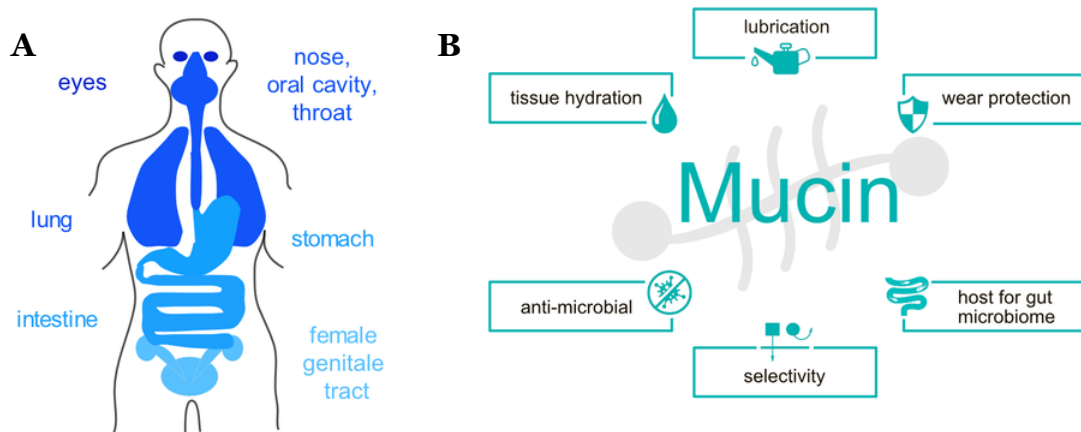
Las GTs se ubican preferentemente en el aparato de Golgi y en el retículo endoplásmico. Para algunas GTs como las GalNAc-Ts se han observado alteraciones de su localización desde el aparato de Golgi al retículo endoplásmico. Además, se ha demostrado que esta alteración se presenta en algunos cánceres y está asociada con la metástasis (Bard, 2013). La inhibición por tanto de estas GTs ubicadas erróneamente en el retículo endoplasmático es una estrategia que podría ser usada para el tratamiento de algunos cánceres (Varki, 2009).

Como veremos en los siguientes apartados, estas enzimas son las encargadas de sintetizar el antígeno Tn, marcador de un gran número de cánceres y que además es útil para el desarrollo de vacunas frente a estas enfermedades.

## 2. Mucinas

Las mucinas son glicoproteínas de alto peso molecular fuertemente glicosiladas (se estima que hasta un 50-80% de su masa total viene determinada por los carbohidratos unidos a las mismas) preferentemente en residuos de Ser/Thr. Participan en la regulación de diversas actividades celulares, otorgando protección fisiológica normal y lubricación a las superficies epiteliales (Hollingsworth, 2004; Senapati, 2008) (Figura 1.3). Hasta ahora se han identificado 21 mucinas humanas, designadas como MUC1 a MUC21, aunque no todas han sido secuenciadas completamente (Moniaux *et al.*, 2001; Perez-Vilar y Hill, 1999; Zhang *et al.*, 2003). Según su localización subcelular, las mucinas se agrupan en dos categorías diferentes, secretadas y unidas a la membrana.

Las mucinas secretadas (MUC2, MUC5AC, MUC5B, MUC6-8 y MUC19) carecen de un dominio transmembrana y se secretan en el espacio extracelular. Son gelificantes y sirven como el principal componente proteico de la capa mucosa que actúa como una primera línea de defensa para las superficies epiteliales contra un medio extracelular hostil.



**Figura 1.3. Mucinas en el cuerpo humano.** (A), Todos los órganos que contienen mucinas. (B) Funciones que desempeñan las mucinas en el organismo. **Fuente:** Prof. Dr. Oliver Lieleg <https://www.mw.tum.de/bme/mucine/>

Por el contrario, las mucinas unidas a la membrana (MUC1, MUC3, MUC4, MUC12-17 y MUC20) son proteínas de tipo I ancladas a la membrana con un dominio de transmembrana, una región extracelular en el N-terminal y una cola citoplasmática en el C-terminal. (Olsen, 2005). A diferencia de las mucinas secretadas, que solo interactúan con proteínas fuera de la célula, las mucinas unidas a la membrana pueden interactuar con proteínas presentes fuera o dentro de la célula a través de su dominio extracelular o su cola citoplasmática, respectivamente. También pueden interactuar con otras proteínas transmembrana a través de su región transmembrana. (Hollingsworth 2004; Kufe 2009)

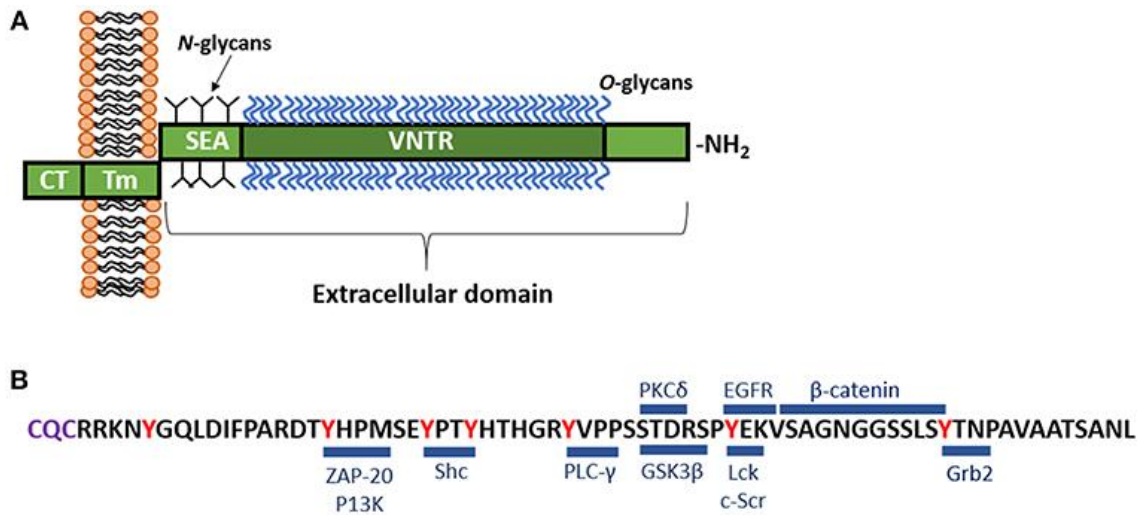
Aunque todas las mucinas están codificadas por diferentes genes, sus estructuras proteicas primarias comparten algunas similitudes características. Además de estas similitudes, las mucinas también albergan muchos dominios

funcionales únicos y motivos que son específicos de un individuo o grupo de mucinas (Hollingsworth, 2004; Senapati, 2008).

Numerosos estudios han demostrado que, debido a sus estructuras especializadas, diferentes mucinas interactúan con distintas proteínas. Estas diversas interacciones proporcionan funciones únicas de cada mucina. Es importante destacar que, en varios tipos de cáncer, tanto la expresión desregulada como las modificaciones estructurales de las mucinas pueden afectar a la progresión del cáncer. Por otro lado, el conocimiento de las vías de señalización por las que las mucinas de membrana promueven el cáncer debería facilitar la identificación de nuevas moléculas para el tratamiento del cáncer (Singh, 2006). De todas las mucinas, MUC1 es la mejor caracterizada con respecto a sus interacciones proteína-proteína. Sin embargo, investigaciones recientes han caracterizado muchas interacciones que involucran otras mucinas. Los estudios realizados en estas mucinas tratan de esclarecer las interacciones que éstas establecen a nivel molecular para facilitar el desarrollo de nuevas terapias frente al cáncer.

### **2.1. Estructura de las mucinas**

Las mucinas consisten en un esqueleto proteico, denominado "apomucina", cubierto con muchos oligosacáridos unidos principalmente a residuos de Ser/Thr y en menor medida a residuos de Asn. Las mucinas además de la glicosilación sufren otras modificaciones postraduccionales como la sulfatación, fosforilación, etc (Figura 1.4). Los primeros estudios de la estructura primaria de mucina fueron complicados debido a esta extensa glicosilación. El análisis de las estructuras de mucina mostró que la presencia de regiones de repetición en tándem (TR) ubicadas centralmente son el rasgo característico de todas las mucinas, que las diferencian de otras glicoproteínas unidas a la membrana.

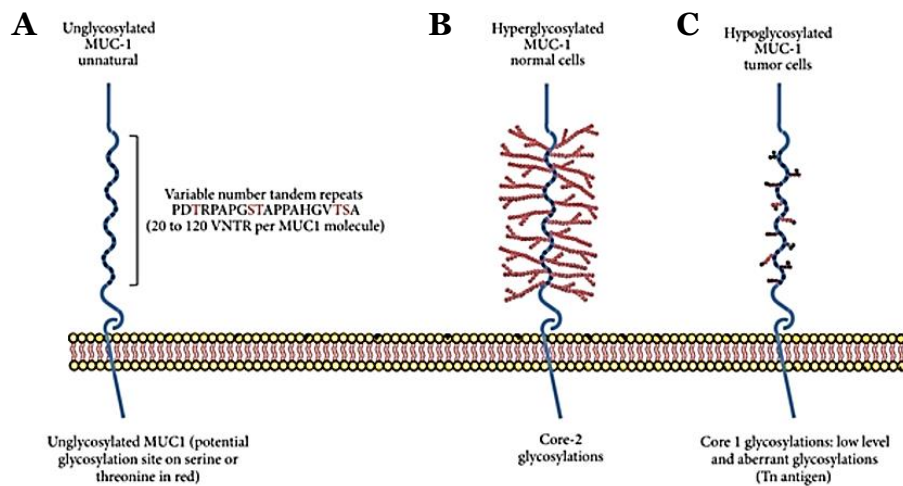


**Figura 1.4. Estructura de la mucina 1 (MUC1).** (A) Representación de los dominios estructuras característicos de la mucina 1. (B) Secuencia de aminoácidos del CT-MUC1 y sitios de unión reconocidos por otras proteínas. (Dhar y McAuley, 2019)

Debido a la presencia de muchos residuos de prolina, treonina y serina en la región TR, también se conoce como dominios PTS (Lang, 2007). Gendler y colaboradores (1987) fueron los primeros que mostraron que el número variable de polimorfismos de repetición en tándem (VNTR), encontrados en la secuencia de codificación de las mucinas, conduce a la expresión de polipéptidos de diferentes longitudes (Gendler, 1987). Sin embargo, se desconoce cómo estas variantes de mucina difieren funcionalmente entre sí. Los residuos de Ser y Thr de los TRs son sitios potenciales para la *O*-glicosilación, y los residuos de prolina podrían desempeñar un papel en el logro de una confirmación adecuada que permita el empaquetamiento cercano de las estructuras de carbohidratos. Las mucinas, como ya se ha mencionado anteriormente, son proteínas muy glicosiladas en donde las ramificaciones formadas por carbohidratos representan posibles sitios de unión para otras proteínas que interactúan con ellas a través de estos. El resto de las regiones de las mucinas alberga muchos motivos estructurales y dominios (Hollingsworth, 2004) que podrían ser importantes para el papel fisiopatológico de las mucinas. Se cree que los dominios de tipo EGF y los dominios de la cola citoplasmática (CT) de las mucinas dirigen las interacciones con diferentes proteínas (Figura 1.4). Los dominios de tipo mucina EGF podrían estar involucrados en la

heterodimerización con receptores ErbB. La presencia de sitios de fosforilación y otros motivos de interacción proteína-proteína, como la secuencia polibásica de aminoácidos, sugieren que los dominios de las CTs podrían interactuar con las quinasas y con otras proteínas asociadas al citoesqueleto. (Kufe, 2009; Senapati, 2008)

MUC1 ha sido la mucina de interés de esta tesis doctoral (Figura 1.5). Esta mucina fue la primera que se clonó, siendo una proteína transmembrana expresada en la superficie apical de la mayoría de las células epiteliales, incluidas las glándulas mamarias, y también en las células hematopoyéticas y en la médula ósea (Gendler, 2001). El dominio extracelular de MUC1 consiste en repeticiones en tándem que comprenden 20 aminoácidos de la secuencia HGVTSAPDTRPAPGSTAPPA con cinco posibles sitios de *O*-glicosilación y tres regiones relevantes dentro de esta repetición que corresponden con los sitios glicosilados: (a) la secuencia GVTSA, que es un sustrato eficaz para GalNAc-Ts (Wandall, 1999); (b) el fragmento PDTR, que es el dominio más inmunogénico de MUC1 y, en consecuencia, es un epítipo bien conocido reconocido por varios anticuerpos anti MUC1 (Karsten, 2004); y (c) la región GSTA, que se reconoce por diferentes anticuerpos y representa una herramienta potencial para su uso en aplicaciones de diagnóstico y terapéuticas (Huang, 2012). MUC1 se sobreexpresa en muchas formas de cáncer, en particular carcinomas de mama en donde se pierde la polarización apical (Girling, 1989), lo que hace que MUC1 se exprese en toda la superficie celular (Gendler, 2001). La glicosilación aberrante ligada a residuos de Thr y Ser de MUC1 también se observa en carcinomas (Fonteneau, 2013; Figura 1.5).



**Figura 1.5. Estructura de MUC1.** (A) Se muestra el esqueleto de la mucina 1. (B) Representación de una glicosilación de MUC1 normal. (C) Representación de una hipoglicosilación de MUC1 en células tumorales. (Fonteneau, 2013)

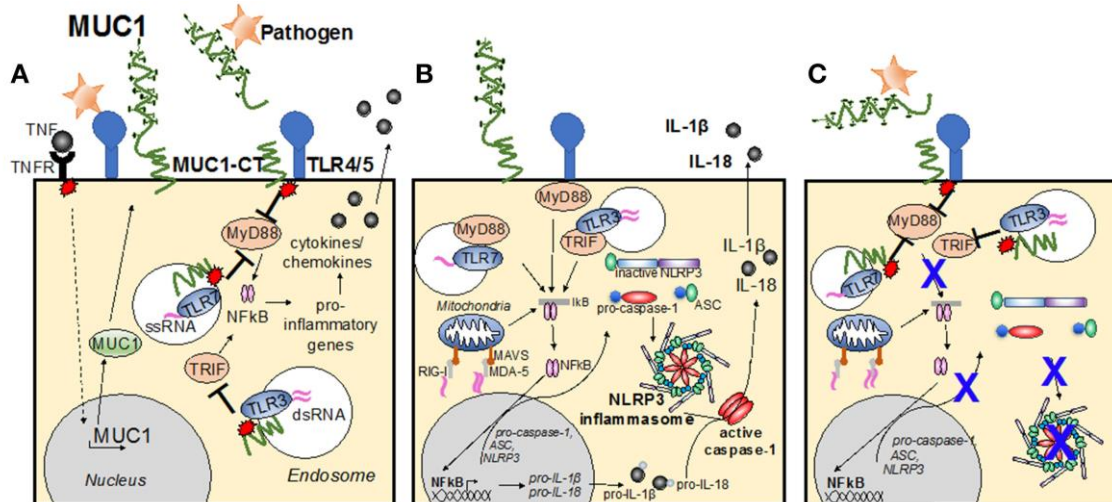
Por lo tanto, las interacciones entre las mucinas unidas a la membrana actuando como receptores y sus proteínas asociadas permiten la transmisión de una señal externa hacia el interior de una célula, lo cual tiene relevancia en la supervivencia, proliferación y diferenciación celular.

## 2.2. Relación de las mucinas con enfermedades

La biosíntesis del *O*-glicano se inicia mediante la transferencia de GalNAc a Ser o Thr de la cadena principal de la proteína, catalizada por alguna de las isoformas de las GalNAc-Ts. Pocos estudios han abordado cómo tiene lugar la glicosilación de *O*-GalNAc *in vivo*, aunque la gran cantidad de datos de los estudios *in vitro* son potencialmente aplicables *in vivo* (Steentoft, 2013). Una complicación con la interpretación de los datos *in vivo* es que los repertorios de GalNAc-Ts en las células no son del todo conocidos, al igual que las capacidades para la *O*-glicosilación que tienen. Nehrke originalmente abordó esto usando una construcción derivada de la proteína gp120 del VIH (Nehrke *et al.* 1998). Este péptido contiene un único aminoácido de treonina, que es glicosilado *in vitro* solo por GalNAc-T3 (Bennett, 1996). Usando este sistema fue posible

recapitular completamente *in vivo* la función de GalNAc-T3 encontrada *in vitro* (Nehrke *et al*, 1998).

Unos años más tarde, se logró encontrar que las mutaciones en el gen que codifica GalNAc-T3 causan calcinosis tumoral familiar, un trastorno metabólico autosómico recesivo grave (Topaz, 2004). Posteriormente se elucidó que el papel biológico más importante de la GalNAc-T3 es la *O*-glicosilación de la Thr178 del FGF23, que es un factor de crecimiento implicado en la regulación de la reabsorción del fosfato por parte del riñón (Clausen, 2006). Por tanto, esta función biológica específica permitió entender la relación directa entre mutaciones inactivantes de GalNAc-T3 y el desarrollo de calcinosis tumoral familiar (Clausen, 2006; Rafaelsen, 2014).



**Figura 1.6. Ejemplo de la influencia de MUC1 en las vías de señalización inflamatoria celular.** (A) Esquema general de como MUC1 media en la cascada de señalización para el bloqueo de la liberación de agentes inflamatorios como IL-1 $\beta$  and IL-18 (B) Se produce una cascada de reacciones desde la membrana y desde el endosoma. Un primer estímulo activa la ruta de transcripción NF-kB, que resulta en la expresión de los componentes del inflammasoma. Luego, un segundo estímulo hace que todo se ancle y active para secretar citoquinas que son secretadas por la célula. (C) MUC1-CT fosforilado recluta al TLR, limitando la activación del TLR para la cascada de señalización y bloqueando la regulación positiva de los componentes del inflammasoma. (Dhar y McAuley, 2019)

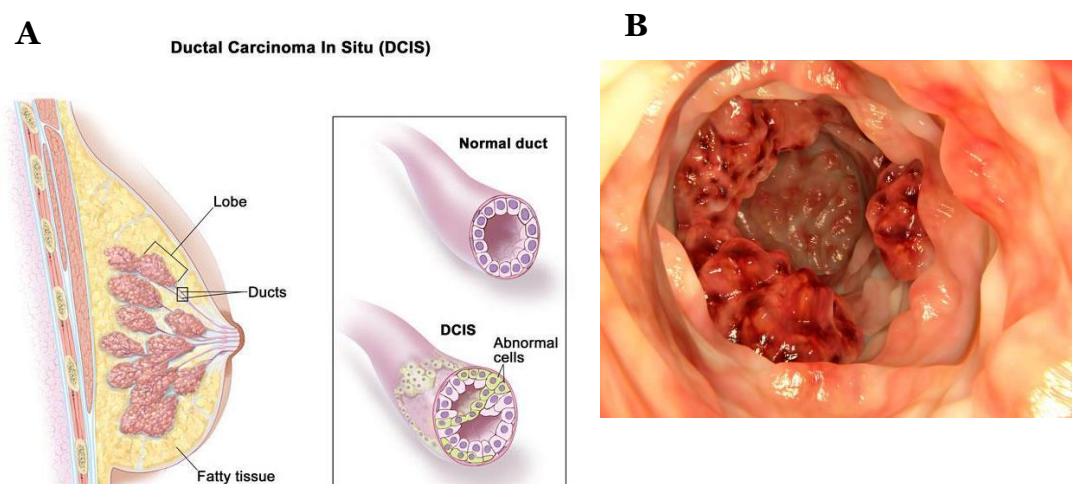
MUC1 ha demostrado su importancia cómo una de las defensas que el cuerpo posee frente a patógenos externos (Dhar y McAuley, 2019; Figura 1.6). La infección activa los TLR de la superficie celular provocando una cascada de reacciones que van a liberar mediadores inflamatorios como el TNF (citoquinas). MUC1 es el encargado de atenuar la inflamación. Al reconocer al patógeno con el dominio extracelular, MUC1 se fosforila en su cola CT, asociándose con el TLR y bloqueando la cascada de señales en el interior de la célula. También se ha observado una correlación entre el aumento de expresión de MUC1 en superficie y la liberación de la citocina antiinflamatoria IL-10 y el interferón tipo I (IFN) (Gaemers et al., 2001; Kato et al., 2014). Estos hechos muestran la importancia de MUC1 como barrera de protección frente a la infección.

No solo se ha observado la glicosilación aberrante en los dominios de tipo mucina en numerosas enfermedades, sino que también se ha descrito su papel principal en la patogénesis del cáncer (Chugh *et al.* 2015; Van Putten y Strijbis, 2017). La presencia aumentada de antígeno T en las zonas extracelulares de las mucinas de las células con adenocarcinoma de colon y colitis, frente a las células epiteliales de colon en condiciones fisiológicas, fue estudiada y documentada por Campbell *et al.*, en 1995. A partir de ese momento, numerosas publicaciones han mostrado la implicación de los glicanos aberrantes en distintas funciones de las células tumorales, como por ejemplo su adhesión celular (Hauselmann y Borsig, 2014), su motilidad, capacidad invasiva y de interacción con otras células. También se ha demostrado cómo estas células son capaces de evadir la respuesta inmunológica (Madsen *et al.*, 2013) evitando así ser eliminadas del organismo y proliferar.

En la leche materna y, por lo tanto, presumiblemente en el tejido mamario normal, los oligosacáridos más comunes se basan en la estructura del núcleo 2 (Gal $\beta$ 1-3[GlcNAc $\beta$ 1-6]GalNAc-Thr/Ser) (Amano *et al.*, 1991; Hanisch, 1989) que terminan en los epítomos de tipo Lewis (Hanisch, 1990). Los epítomos sialilados de tipo Lewis también se expresan en carcinomas de colon (Baekstrom, 1991). Tanto el cáncer de colon como el de mama (Figura 1.7) son de los que mas incidencia poseen en el mundo, por ello distintos grupos han centrado sus esfuerzos en su estudio. Los estudios de MUC1 expresados en la



línea celular de cáncer de mama humano T47D muestran que los glicanos son más cortos, están basados en el núcleo 1 (Gal $\beta$ 1-3GalNAc-Thr/Ser) y pueden estar muy sialilados (Hanisch, 1996; Hull, 1989; Lloyd, 1996). Sin embargo, algunas líneas celulares de cáncer de mama (MCF7, MDA 231 y ZR75-1) pueden expresar O-glicanos del núcleo 2 (Baekstrom, 2003; Muller y Hanisch, 2002). De acuerdo con esto, los estudios sobre carcinomas de mama primarios han demostrado que las glicosiltransferasas responsables de agregar ácido siálico a los glicanos del núcleo 1 están sobreexpresadas en los carcinomas de mama, mientras que la glicosiltransferasa C2GnT1, encargada de modificar el núcleo 1 hasta núcleo 2 por glicosiliación en  $\beta$ -1-6 de una GlcNAc, no se observó que estuviese afectada (Burchell *et al.* 1999).



**Figura 1.7. Cáncer de mama y de colon.** (A) Evolución del desarrollo del cáncer de mama por invasividad de las células aberrantes en los conductos lobulares. **Fuente:** <https://www.cancer.gov/news-events/cancer-currents-blog/2015/dcis-low-risk> (B) Imagen de una colonoscopia donde las células cancerígenas han proliferado. **Fuente:** <https://a2zhealthy.com/colon-cancer-symptoms/>

Como bien describieron Campbell y colaboradores en 1995 con los antígenos T, la glicosilación aberrante en las células tumorales está basada en glicanos de núcleo 1 y son ricos en antígenos de su misma familia, como son los antígenos T, ST y Tn. (Takeuchi, 2002, Lloyd, 1996; Springer, 1984). A continuación, vamos a ahondar en dichos antígenos. El conocimiento de sus características estructurales es crucial para un posible desarrollo de métodos de

diagnóstico, vacunas y tratamientos frente al cáncer. En concreto nos centraremos en los conocimientos actuales de los antígenos Tn y STn.

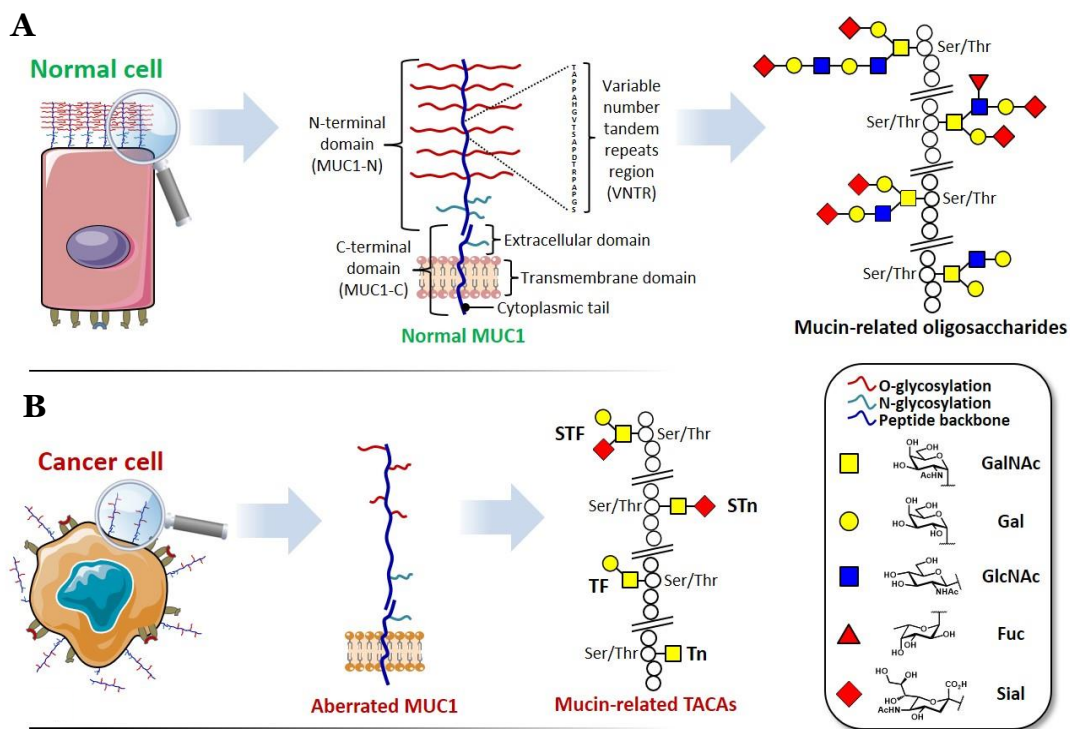
### 3. Antígenos Tn y STn

#### 3.1. ¿Qué son?

El descubrimiento de ambos antígenos fue un hito en la historia de la glicobiología, y su expresión inusual fue una de las primeras descripciones de unos antígenos relacionados con carbohidratos implícitamente relacionados con un trastorno humano. Los antígenos de la familia Thomsen-Friedenreich se denominaron inicialmente antígenos oncofetales, ya que generalmente son precursores de las cadenas complejas normales de *O*-glicanos (Springer 1984).

El antígeno T fue identificado por primera vez por Thomsen en 1930, en colaboración con Friedenreich, en una muestra de sangre contaminada por bacterias. Se observó la liberación del ácido siálico de las cadenas de glicanos por acción de la neuraminidasa de dichas bacterias, exponiendo el antígeno Tn en los hematíes. Este hecho producía que los anticuerpos anti-T presentes en el suero los reconociesen generando la hemaglutinación. Este hecho fue descrito en 1957 por Moreau como síndrome de Tn o síndrome de poliaglutinación, donde observó la expresión de los antígenos Tn y STn en pacientes de una rara anemia hemolítica que era independiente del grupo sanguíneo ABO (Loureiro, 2015).

El antígeno T se sobreexpresa en el cáncer de mama, colon, páncreas y pulmón, pero también se puede detectar en hasta el 40 % de los tejidos normales (Kudelka, 2015) y en los tejidos inflamados. Del mismo modo, la expresión de ST aumenta en gran medida en algunos tipos de cáncer en comparación con los tejidos normales, como el cáncer de mama y vejiga (Wang, 2015; Videira, 2009; Schneider, 2001), pero también se expresa en el epitelio normal, muchas proteínas séricas y leucocitos (Videira, 2008; Priatel, 2000). Sin embargo, los antígenos Tn y STn muestran un mejor patrón asociado al cáncer, con una expresión ausente o muy limitada en tejidos sanos adultos (Osako, 1993; Cao, 1995; David, 1992), y por lo tanto son más interesantes como objetivos terapéuticos (Figura 1.8).



**Figura 1.8. La glicosilación aberrante genera posibles dianas terapéuticas.** (A) Esquemas ampliados de la glicosilación de MUC1 en la superficie de las células de tejidos sanos. (B) Esquemas ampliados de la glicosilación que presenta MUC1 en células tumorales donde se muestran todos los posibles antígenos de la familia Thomsen-Friedenreich que se han estudiado como posible diana terapéutica. **Fuente:** <https://glycopedia.eu/e-chapters/Anti-cancer-Immunotherapy/MUC1-and-tumor-associated-carbohydrate-antigens-TACAs>

### 3.1.1. Antígeno Tn

El antígeno Tn es expresado en la mayoría de los carcinomas, como el de mama, páncreas, colon, pulmón y vejiga, siendo menos común en neoplasias hematológicas. En los tejidos normales, solo se expresa en el cerebro embrionario (Akita, 2001) y está protegido en el tejido adulto. El antígeno Tn es en realidad un precursor biosintético normal de los O-glicanos. Sin embargo, no es un producto terminal normal que se encuentre en células sanas. Sin embargo, se ha observado ocasionalmente en células adultas normales en el aparato secretor dónde se ha encontrado en proteínas secretadas (Osako, 1993).

La sobreexpresión de antígenos Tn promueve la proliferación de células cancerosas y la invasividad. De hecho, en el cáncer de mama, la acumulación de proteínas portadoras de Tn en lamelipodios promueve la adhesión celular, la motilidad celular y la invasividad (Gill, 2013). Sorprendentemente, el antígeno Tn también se detecta en las primeras etapas del desarrollo del tumor y puede servir como un biomarcador, ya que su expresión está asociada con tumores invasivos y altamente proliferativos, y metástasis (Itzkowitz, 1992).

### 3.1.2. Antígeno STn

El antígeno Sialyl-Tn (STn) es también un antígeno de carcinomas que se expresa temprano en la carcinogénesis. A diferencia de Tn, STn no es un precursor biosintético normal, lo que significa que su expresión es necesariamente patológica (Kudelka, 2015).

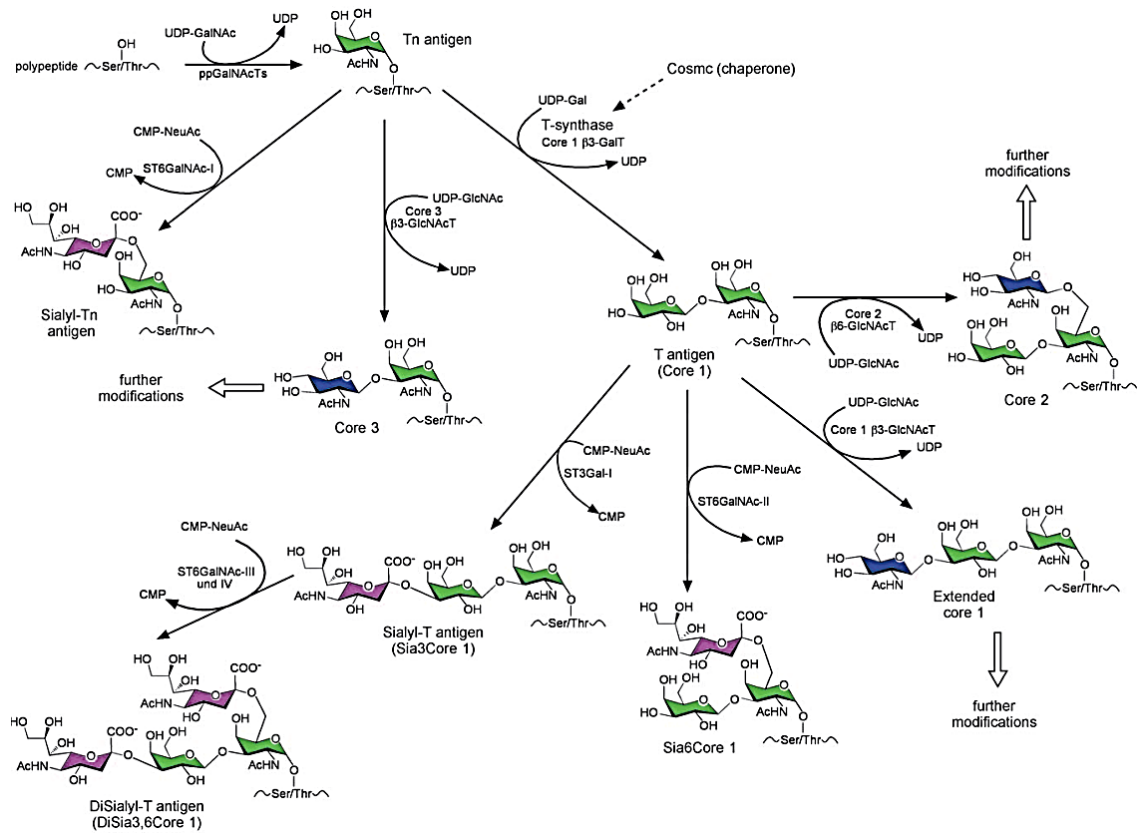
En algunos tejidos normales, como el de colon, el residuo de ácido siálico del STn puede ser *O*-acetilado, enmascarando así el STn y, por lo tanto, su reconocimiento por los anticuerpos anti-STn (Itzkowitz, 1989).

La expresión de STn modula un fenotipo maligno que induce un comportamiento celular más agresivo en las células de carcinoma gástrico y de mama, como la disminución de la agregación de células y el aumento de la adhesión, migración e invasión de la matriz extracelular (Pinho, 2007; Julien, 2005 y 2006; Tamura, 2014).

## 3.2. ¿Cómo y por qué se producen?

Como ya se ha introducido, la *O*-glicosilación de tipo mucina es una modificación postraduccional esencial en organismos superiores (ausente en levaduras, plantas y bacterias). Aunque sólo el 1% del genoma codifica genes implicados en la biosíntesis de esta modificación post-traduccional (Apweiler et al., 1999; Lowe y Marth, 2003), más del 80% de las proteínas que pasan a través de la vía secretora son modificadas por la maquinaria enzimática responsable de la misma (Gill et al., 2011; Bennet et al., 2012). Este proceso se inicia con la adición de un residuo de N-acetilgalactosamina (GalNAc) al grupo hidroxilo de la cadena lateral de una Ser o Thr de la proteína sustrato. Dicha adición, llevada a cabo en el aparato de Golgi en condiciones normales (Rottger et al., 1998; Gill

et al., 2010; Hang y Bertozzi, 2005; Ten Hagen et al., 2003) y, según se ha postulado más recientemente, en el retículo endoplasmático en condiciones tumorales (Gill et al., 2010; Chia et al., 2016), la realiza algún miembro de la gran familia de GalNAc-Ts, clasificadas en el grupo GT27 de la base de datos de enzimas que procesan carbohidratos, CAZy Database (<http://www.cazy.org>; Lombard et al., 2013). El producto de esta reacción, en el que el donador es el UDP-GalNAc, es el ya mencionado antígeno Tn (Hang y Bertozzi, 2005; Gill et al., 2016; Chia et al., 2016).



**Figura 1.9. Esquema de las distintas rutas de síntesis del antígeno Tn y sus derivados.** Esquema ampliado con las enzimas implicadas y las estructuras que adicionan los distintos oligosacáridos en la cadena (Ju *et al*, 2011)

### 3.2.1. Antígeno Tn

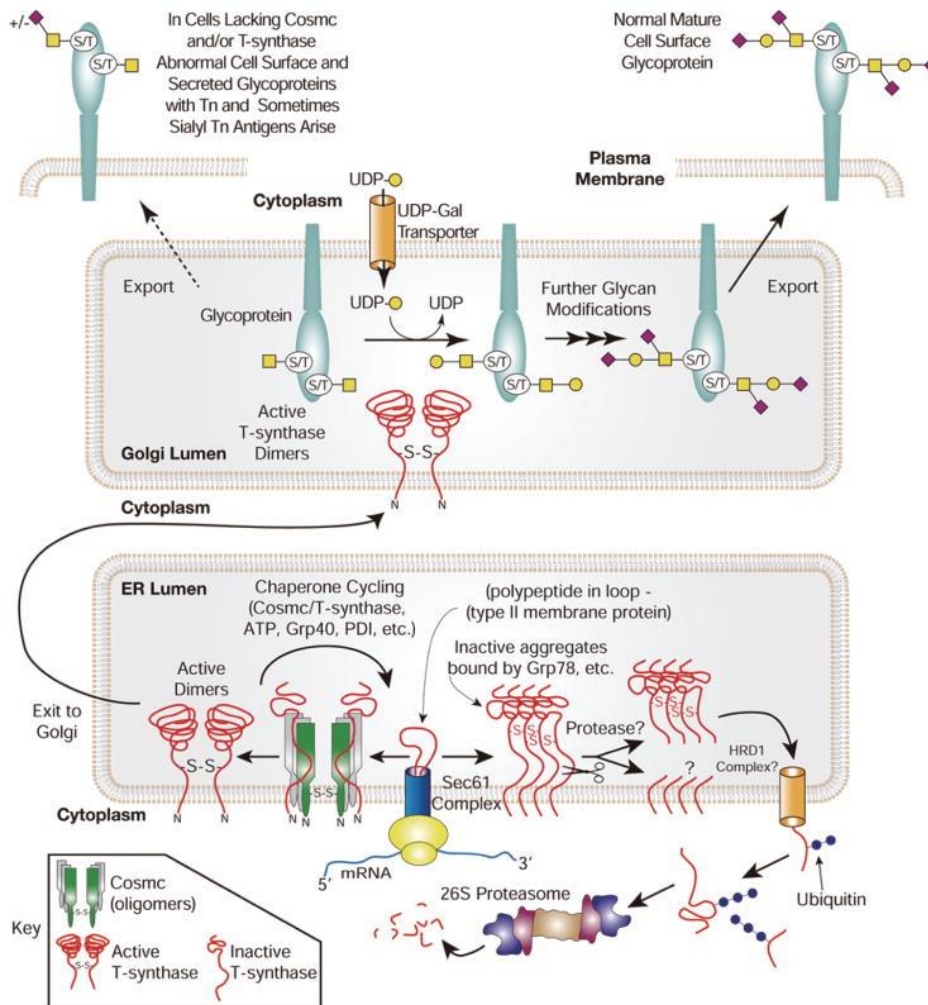
El antígeno Tn, normalmente se extiende por la 1,3-galactosiltransferasa C1GalT1, también denominada T-sintasa o núcleo-1 sintasa, o por el núcleo-3 1,3-N-acetilglucosaminatransferasa (C3GnT) para formar el antígeno T (núcleo-1) y estructura core-3, respectivamente (Figura 1.9). Su sobreexpresión en

cáncer se genera por varios motivos: pérdida de las enzimas T-sintasa o C3GnT (Vavasseur, 1995; Yang, 1994), la expresión alterada de las mismas o la reubicación de las GalNAc-Ts desde el Golgi al retículo endoplásmico (Gill, 2013; Chia, 2016). Además, los defectos en Cosmc, una chaperona esencial para la actividad de C1GalT1, debido al silenciamiento epigenético o las mutaciones presentes en la misma, también pueden conducir a un aumento de la expresión del antígeno Tn en el cáncer (Ju, 2008; Mi, 2012). Cosmc, con ubicación en el retículo endoplásmico (RE), tiene como función plegar a C1GalT para prevenir su agregación y degradación proteosómica (Aryal, 2010). La ausencia de Cosmc provoca que la enzima C1GalT1 se agregue y asocie a otras chaperonas, como Grp78, promoviendo su degradación a través del proteosoma (Figura 1.10).

### 3.2.2. Antígeno STn

El antígeno STn (ácido siálico- $\alpha$ -2-6-GalNAc-Ser/Thr) es un derivado del Tn que además contiene un grupo siálico unido covalentemente a GalNAc. (Figura 1.9). A menudo se coexpresa con Tn y, por lo tanto, los mecanismos que producen una sobreexpresión de Tn probablemente se apliquen también a STn. Se han demostrado también mecanismos específicos para la sobreexpresión de STn, que incluyen la regulación positiva de ST6GalNAc-I, la reubicación de Cosmc del Golgi al retículo endoplásmico y también la pérdida de grupos O-acetil por parte de STn, en algunos cánceres (Julien, 2001; Marcos, 2011, Vazquez-Martin, 2004; Ferreira, 2013).

Todos los factores expuestos en este apartado por los cuales se producen estos antígenos han dado lugar a la hipótesis de que los antígenos Tn y STn son una estructura aberrante en la superficie de las células cancerosas y que, por consiguiente, debieran ser reconocidas por el sistema inmune del paciente a través de la formación de anticuerpos. Esta hipótesis por tanto ha llevado a esfuerzos para desarrollar inmunoterapias y vacunas basadas en la existencia de estos antígenos Tn y STn en diferentes cánceres (Dorkuno, 1998; Karsten, 2004; Loureiro, 2015).



**Figura 1.10. Cosmc, una chaperona para la T-sintasa o C1GalT1.** La expresión normal de Cosmc y T-sintasa conduce a la expresión normal de O-glicanos y estructuras complejas en las glicoproteínas de la membrana plasmática, mientras que la falta de Cosmc y T-sintasa conduce a la expresión de antígeno Tn y STn. (Ju, 2018)

### 3.3. Formas de tratamiento frente a cánceres que presentan el antígeno Tn y STn.

Se sabe que solo unas pocas glicoproteínas presentan antígenos Tn, T, STn y sialil-T (ST) en tejidos malignos (Yu 2007). La mucina MUC1 y CD44v6 muestran antígenos STn y ST en cánceres de colon, gástricos y de mama (Baldus, 1998; Burdick, 1997; Singh, 2001; Storr, 2008). MUC2 es un importante portador de glicanos truncados en el cáncer gástrico (Conze 2010). La sobreexpresión de STn en el cáncer de mama y gástrico se asocia con la sobreexpresión de MUC1, CD44 y ST6GalNAc1 (Ozaki, 2012; Julien, 2006). Aunque CD44v6 se expresa en algunos tipos de epitelios sanos, se observa una mayor expresión en carcinomas de células escamosas y adenocarcinomas,

incluidos los de mama, pulmón, colon y carcinomas pancreáticos (Hofmann, 1991; Ponta 2003; Wai 2004). Curiosamente, los niveles séricos de osteopontina, un ligando de CD44, que en sí mismo es un portador STn, se han detectado en pacientes con cáncer y se correlacionan con un pronóstico desfavorable (Wai, 2004).

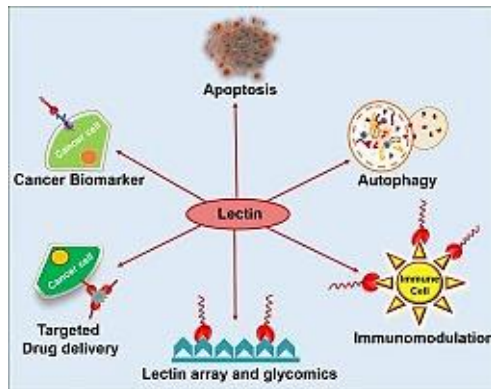
Como ya se ha expuesto en anteriores apartados, debido a la asociación de los antígenos Tn y STn con varios tipos de cáncer, así como a la progresión tumoral y la metástasis (Osako, 1993; Cao, 1995; David, 1992), se han realizado esfuerzos para desarrollar métodos de detección y tratamiento contra estas estructuras glicosiladas. Para lograr este fin, se han utilizado principalmente proteínas de reconocimiento de azúcares, como son las lectinas, y en mayor medida anticuerpos. El hecho de que los pacientes con cáncer puedan producir autoanticuerpos provocó su uso como biomarcadores para la detección temprana del cáncer (Wandall, 2010) o la respuesta a la terapia. Curiosamente, los anticuerpos encontrados en pacientes con cáncer no han demostrado ser eficaces para activar el sistema inmunitario y eliminar las células malignas (Wolfert and Boons, 2013). Por lo tanto, es esencial producir anticuerpos más efectivos que ataquen específicamente las células cancerosas.

### 3.3.1. Lectinas frente a los antígenos Thomsen-Friedenreich

Las lectinas son un grupo de proteínas presentes tanto en el reino animal, vegetal como en bacterias y virus, que se unen a azúcares con un cierto grado de especificidad (Hadjialirezaei, 2017)

Pueden unirse a uno o más residuos de carbohidratos, normalmente de 2 a 15 monosacáridos siendo los más comunes D-manosa, D-galactosa, D-glucosa, L-fucosa, N-acetil-D-glucosamina, N-acetil-D-galactosamina, glucosamina y galactosamina. Se han utilizado en terapias contra el cáncer de formas variadas (Bhutia *et al*; Figura 1.11), aunque su acción puede depender en algunos casos de la asociación a otras moléculas, como las proteínas del complemento (Hadjialirezaei, 2017).



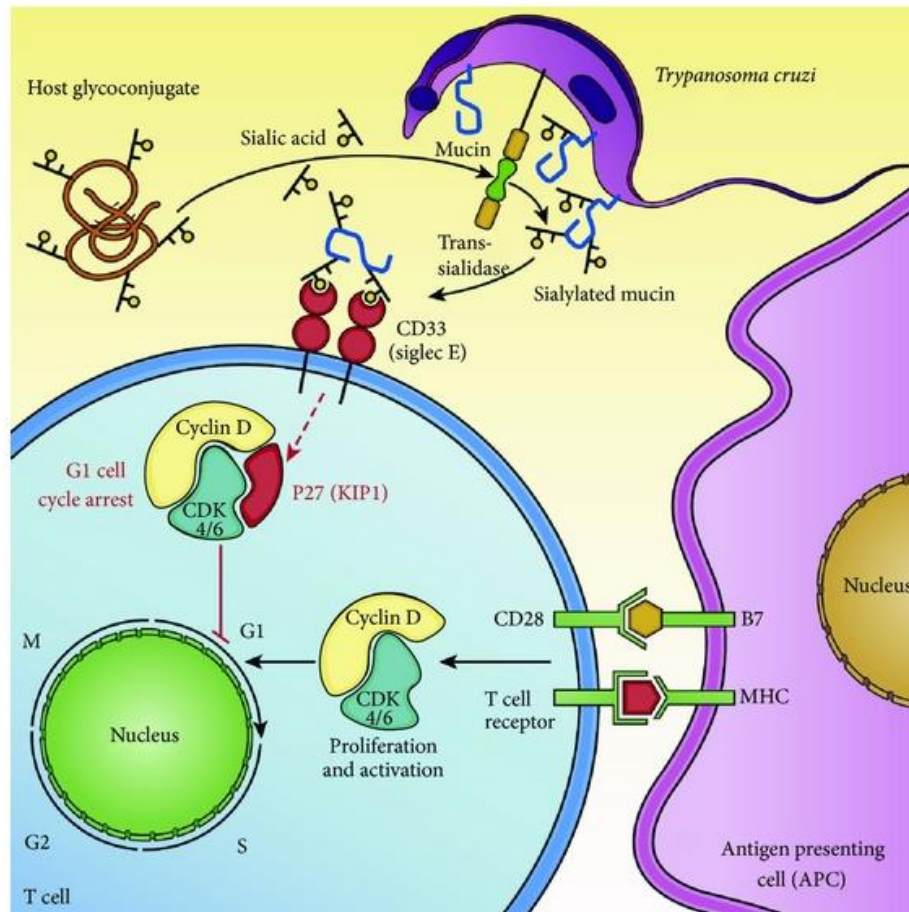


**Figura 1.11. Usos de las lectina en terapias de cáncer.** Pueden usarse para casos como apoptosis por muerte celular dependiente de autofagia, transporte de medicamentos, inmunomodulación y chips de lectinas. **Fuente:** (Bhutia *et al.*, 2019)

Algunos de los ejemplos de lectinas que se han publicado frente a estos antígenos son la lectina de *Helix pomatia* (HPA) y la lectina B<sub>4</sub> de *Vicia villosa* (VVA-B<sub>4</sub>), ambas lectinas de plantas que reconocen el antígeno Tn (Ju 2011; Bard 2013). Por otro lado, las lectinas que se utilizan para reconocer al antígeno STn suelen ser de origen animal, como lectinas de unión a manosa (MBL), MGL o KLH (Hadjialirezaei 2017). Suelen ser proteínas transportadoras del antígeno que, en ocasiones, se utilizan como tal a la hora de generar respuesta inmunológica junto con un adyuvante como QS-21 (Gilewski, 2007; Sletmoen, 2017; Burchell, 2015). Existen otras superfamilias, como son las SIGLECs y las galectinas, que también reconocen a dichos antígenos debido a su localización en la superficie celular (Takamiya, 2013) y son reguladoras de procesos inflamatorios y de respuesta inmune (Ruvolo, 2016).

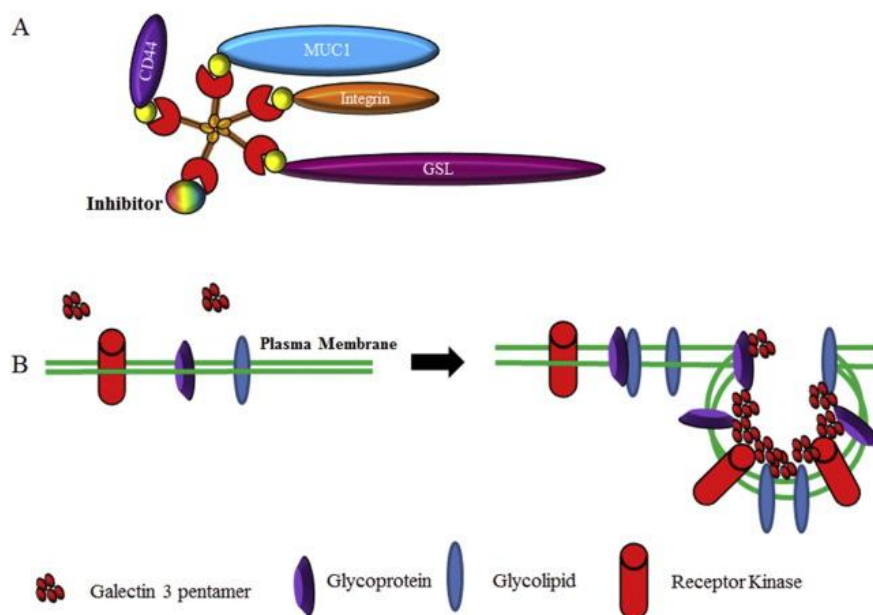
Las dos lectinas más comunes que se utilizan para explorar la expresión del antígeno Tn son VVA-B<sub>4</sub> y HPA (Tollefsen SE, 1983, Newman, 1977). VVA-B<sub>4</sub> se ha utilizado para explorar la expresión del antígeno Tn en diferentes células cancerosas, incluyendo las del páncreas, mama, pulmón y próstata. HPA ha sido ampliamente utilizado desde la década de 1980 para estudiar la expresión del antígeno Tn. Se utilizó acoplada a una sonda fluorescente para determinar al antígeno Tn en eritrocitos (Tollefsen SE, 1983). Sin embargo, se observaron ciertos problemas, ya que no solo se unía al antígeno Tn, sino que

HPA también se une al antígeno Cad ( $\text{NeuAc}\alpha 2\text{-3}[\text{GalNAc}\beta 1\text{-4}]\text{Gal}\beta 1\text{-R}$ ), al antígeno de grupo sanguíneo A ( $\text{Fuc}\alpha 1\text{-2}[\text{GalNAc}\alpha 1\text{-3}]\text{Gal}\beta 1\text{-R}$ ) y al glucolípido Forssman (Newman, 1977). Prácticamente la totalidad de las lectinas muestran este comportamiento promiscuo, por lo que estos estudios deben hacerse con cautela y complementarse en paralelo con otros métodos (por ejemplo, a través de ensayos ELLA, microarrays, ensayos ELISA, etc).



**Figura 1.12. Modelo que representa el efecto inhibitorio de las sialoglicoproteínas de *Trypanosoma cruzi* sobre la activación de las células T.** Diagrama esquemático que muestra como la sialilación de oligosacáridos unidos a O-glicanos de tripomastigotes, al interactuar con Siglec E (CD33), conducen a la inducción de p27/Kip1, un miembro de la familia de inhibidores de CDK que regula negativamente la transición de G1 a S, por lo que amortigua las respuestas inmunes mediadas por células T al inducir la detención del ciclo celular. (Morrot, 2013)

Siglec-15 reconoce preferentemente el antígeno STn. Se expresa en macrófagos asociados a tumores (TAM) y en diversos tejidos de carcinoma humano, incluidos pulmón, hígado y recto (Takamiya, 2013). Las lectinas SIGLECs pertenecen a una superfamilia de inmunoglobulinas de membrana que se unen al ácido sialico y están implicadas en procesos de transducción de señales inhibitorias (Figura 1.12). La unión de las células mieloides a través de Siglec-15 a STn en las células tumorales dio como resultado un aumento de la secreción de TGF- $\beta$  en el microambiente tumoral que se asocia con la progresión del cáncer. También existen ciertos tipos de proteínas que reconocen a varios de los antígenos Thomsen-Friedenreich como es el caso de la galectina 3 (Figura 1.13). La expresión de esta galectina también aumenta en sueros de pacientes de varios tipos de cáncer y se asocia con un mayor riesgo de metástasis (Iurisci, 2000; Vereecken, 2006) debido a que su función engloba procesos muy variados, como adhesión celular o endocitosis (Ruvolo, 2016).



**Figura 1.13. La endocitosis mediada por glicolípidos de receptores como CD44 involucra a la galectina 3.** (A). La galectina 3 puede formar pentámeros donde cada monómero puede unirse a una glicoproteína, glicolípidos o inhibidor por separado. (B). La galectina 3 puede promover la formación de vesículas endocíticas de clatrina cuando la galectina se une a varias glicoproteínas y glicolípidos de membrana (Ruvolo, 2016)

Los miembros de esta familia de lectinas, a diferencia de las SIGLECs, son inmunomoduladores solubles y se unirán a la galactosa que está unida a N-acetilglucosamina en  $\beta$ -1,3 o  $\beta$ -1,4. Participan modulando vías de señalización, tanto intra como extracelularmente.

En esta tesis doctoral se presentará la caracterización biofísica de la unión de la lectina *Soybean agglutinin* (SBA) frente a glicopéptidos no modificados y modificados que mimetizan a la familia de antígenos Thomsen-Friedenreich, además de la determinación de los aspectos moleculares de cómo las lectinas seleccionadas reconocen a estos glicopéptidos. Esta investigación pretende asentar las bases para la síntesis de futuras lectinas mucho más específicas con aplicaciones en el diagnóstico del cáncer (Poiroux, 2011).

### 3.3.2. Anticuerpos frente a antígenos Thomsen-Friedenreich

Los epítomos Tn y STn han sido utilizados como diana para el estudio de su interacción con anticuerpos desde que se les relacionó con los procesos de progresión del cáncer. A pesar de su importancia, estos antígenos han mostrado ser poco inmunogénicos. Los esfuerzos previos de desarrollo de anticuerpos utilizaron glicoproteínas purificadas de muestras de cáncer y adyuvante de Freund, o bacterias inactivadas por calor recubiertas de mucina, para la inmunización de ratones (Prendergast, 2017). Estos enfoques han dado como resultado el desarrollo de varios anticuerpos monoclonales (mAb) anti-STn murinos, incluidos B72.3 (Schlom, 1989) (y su anticuerpo sucesor CC49) (Katari, 1990), TKH2 (Kjeldsen, 1988; Yuhne, 2013) y HB-STn1 (clon 3F1)(Cao 1996) (Kjeldsen, 1988) y muchos otros (Loureiro, 2015), así como otros tanto que reconocen el antígeno Tn como son los anticuerpos 1SM3 (Dokurno, 1998), AR20.5 (Movahedin, 2017), 5E5 (Sabbatini *et al*, 2007; Posey 2016) y MLS 128 (Nakada, 1992)

La mayoría de anticuerpos que se están desarrollando al respecto están asociados a MUC1 (Karsten, 2004), aunque en algunos casos no es así. El anticuerpo monoclonal MLS 128 (Nakada, 1990) dirigido frente al antígeno Tn, proviene de la glicoforina A. Esta proteína se encuentra localizada en las membranas de los eritrocitos. Al igual que MUC-1, esta proteína posee una gran cantidad de oligosacáridos basados en la estructura GalNAc-Ser/Thr. Nakada *et*

*al.* concluyó en su estudio que este anticuerpo no necesitaba más de 3 residuos consecutivos de GalNAc-Ser/Thr para un reconocimiento óptimo.

En cuanto a los anticuerpos dirigidos a MUC-1, a pesar de estar basados en la misma secuencia señal, presentan diferencias de acción como ocurre con el 214D4, 1SM3 y 5E5. El 214D4 reconoce el MUC1 humano independientemente de su patrón de glicosilación, el mAb 1SM3 reconoce las glicoformas dependientes de diferenciación (todo el core 2 y sus predecesores, los antígenos Tn y T) y el 5E5 reconoce exclusivamente las glicoformas asociadas al cáncer de MUC1 (Tn y STn asociados a MUC1) (Sorensen, 2006). Aunque los anticuerpos 1SM3 y 5E5 muestran reactividad asociada al cáncer, el 1SM3 también reconoce, aunque en mucha menor medida, el tejido humano sano como se ha demostrado en el riñón y el timo (Burchell, 1987). El 1SM3 se descubrió por primera vez en pacientes con cáncer de mama, y no es sólo el anticuerpo más estudiado de todos si no que, además, fue el primer ejemplo de un anticuerpo específico que reconoce preferentemente el antígeno Tn presente en las MUC1 (Karsten, 2004).

El anticuerpo 1SM3 muestra una alta selectividad frente a la mucina MUC1 asociada a carcinomas. La alta especificidad del anticuerpo 1SM3 hace que sea una herramienta potencialmente útil en el diagnóstico y tratamiento del cáncer de mama. Por lo tanto, es de gran interés el entender a nivel molecular la especificidad de la interacción entre el 1SM3 y glicopéptidos derivados de MUC1. En particular, este anticuerpo reconoce el epítipo PDT\*RP donde T\* es una treonina unida covalentemente a GalNAc. Estos estudios podrían permitir la obtención de anticuerpos de mayor afinidad mediante el uso de herramientas de biotecnología.

Los anticuerpos monoclonales y policlonales han demostrado ser herramientas útiles en la identificación de la glicosilación específica de tumor, aunque los primeros enfoques inmunológicos para la detección del antígeno Tn se basó en antisueros anti-Tn (Prendergast, 2017). En 1985, se desarrollaron, por primera vez, anticuerpos monoclonales murinos (mAb) contra el antígeno Tn (Springer *et al.*, 1984). La especificidad de los anticuerpos ha sido un problema con el uso de diferentes mAbs para el antígeno Tn. Algunos

anticuerpos monoclonales para el antígeno Tn presentan una reacción cruzada con otros glicanos que contienen GalNAc, tales como el antígeno del grupo sanguíneo A. Uno de los que más se utiliza hoy en día para unir al antígeno Tn es el anticuerpo de ratón anti-Tn (CA3638, clon 12A8-C7-F5), purificado por Springer *et al.* (Dokurno, 1998).

Para poder desarrollar futuras vacunas es necesario en primer lugar comprender las interacciones entre el anticuerpo y el antígeno Tn. En esta tesis se ha decidido usar un fragmento ScFv del anticuerpo 1SM3 para poder entender las bases moleculares de la unión antígeno-anticuerpo. Este fragmento se escogió porque contiene la zona de reconocimiento del antígeno Tn y porque se ha demostrado que es relativamente fácil expresar este tipo de fragmentos en organismos como *E. coli* o *Pichia pastoris*. Con estudios exhaustivos biofísicos usando este fragmento del 1SM3, junto con la determinación de la base molecular de cómo este fragmento reconoce al antígeno Tn, se podrían sintetizar derivados del antígeno Tn que puedan inducir la formación de anticuerpos más potentes. Estos anticuerpos a su vez podrían ser útiles para generar vacunas y herramientas de diagnóstico frente a diferentes tipos de cáncer.

# **OBJETIVOS**

Teniendo en cuenta los estudios anteriores a este trabajo, el objetivo general que se planteó al inicio de la misma fue profundizar en el conocimiento de las bases moleculares de cómo anticuerpos y lectinas reconocen al antígeno Tn presente en MUC1. Además, pretendíamos esclarecer y mejorar el reconocimiento antigénico a través de la obtención de mejores anticuerpos o péptidos modificados de MUC1 que pudiesen ser usados para el desarrollo de vacunas y de herramientas de diagnóstico frente al cáncer.

Dentro de este objetivo general se plantearon los siguientes objetivos específicos, que a su vez fueron tratados en los artículos correspondientes:

- Elucidación de las bases moleculares de la interacción entre la lectina SBA y glicopeptidos del tipo Tn, derivados del esqueleto aminoacídico de MUC1 (Madariaga, **Castro-López et al**, 2014).
- Caracterización de las bases moleculares de cómo el 1SM3-ScFv reconoce a los antígenos Tn que contengan Ser o Thr (**Castro-López**, 2015).
- Diseño racional de nuevos antígenos Tn en los que el oxígeno del hidroxilo de Thr/Ser es reemplazado por azufre y contenga “linkers” entre el aminoácido y el carbohidrato para la obtención de mejores vacunas (Rojas-Ocariz, **Castro-López et al**, 2016).
- Modificación de la Pro del APDT\*RP para mejorar su reconocimiento por parte de los anticuerpos anti-Tn con el fin de usarlos para el diagnóstico de cánceres asociados al antígeno Tn (Somovilla, **Castro-López et al**, 2017).
- Estudio de la relevancia de las moléculas de agua en el reconocimiento de los anticuerpos anti-Tn frente al antígeno Tn conteniendo Ser o Thr (Bermejo, **Castro-López et al**, 2018).
- Diseño racional de nuevos antígenos Tn en los que el oxígeno del hidroxilo de Ser/Thr es reemplazado por S/Se con el fin de obtener mejores anticuerpos anti-Tn (Compañón, **Castro-López et al**, 2019).



# **MATERIALES Y MÉTODOS**

## LECTINA

Todas las técnicas detalladas en este apartado son las empleadas en el artículo I, referente a la lectina (SBA). Las aportaciones experimentales del doctorando en dicho artículo han consistido en la purificación de la muestra, la obtención y resolución de todas las estructuras cristalinas que se publican por cristalografía de rayos X y la realización de ensayos biofísicos de afinidad (ITC). Son, por ello, éstas las que se detallan con mayor concreción, si bien aparecen recogidas el resto de las técnicas empleadas para la obtención del conjunto de datos que han dado lugar al artículo. Los detalles de estas últimas aparecen con más precisión en el apartado correspondiente al artículo.

### **2.1 PROCESO DE PURIFICACIÓN DE LA LECTINA SOYBEAN AGLUTTININ (SBA)**

#### 2.1.1 Purificación de la Lectina (SBA)

##### 2.1.1.1 *Solubilización de la muestra*

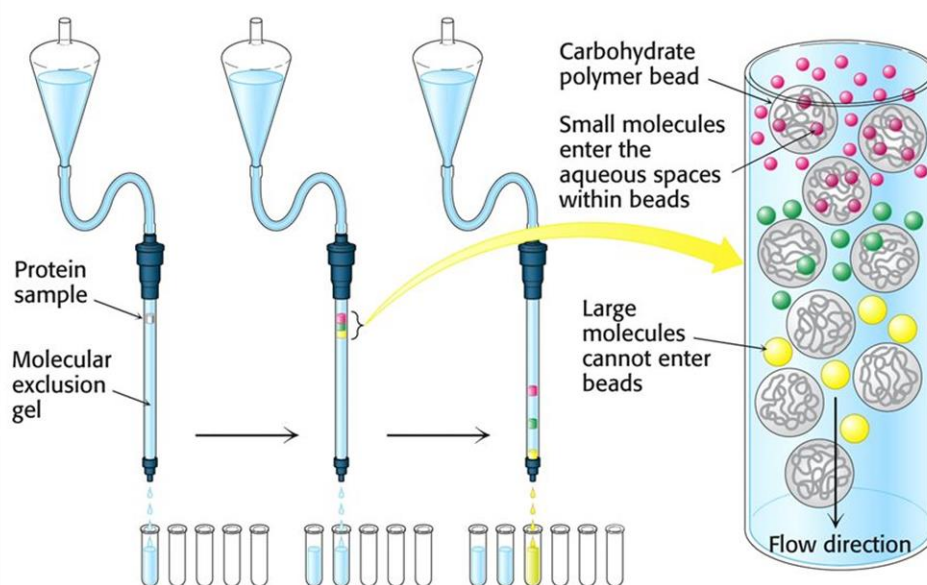
La lectina “Soybean Agglutinin” (SBA) se disolvió en 2,5 ml del tampón 25 mM TRIS pH 7,5, 150 mM NaCl y se filtró a través de un filtro de 0,45  $\mu\text{m}$  de tamaño de poro para eliminar posibles agregados. Para evitar pérdidas de proteína por adsorción en el filtrado, se procedió a humedecer previamente la membrana del filtro con 5 ml del tampón en el que se encuentra la proteína disuelta. Se arrastraron los posibles restos de muestra filtrando otros 0,5 ml más de tampón hasta alcanzar un volumen final de 3,0 ml.

##### 2.1.1.2 *Cromatografía de Exclusión Molecular*

La cromatografía de exclusión molecular es una técnica que permite separar moléculas en función de su tamaño (Figura 2.1). La matriz está compuesta por un material poroso que forma canales, reteniendo en su interior durante más tiempo las moléculas más pequeñas. Esta cromatografía nos va a permitir separar tanto posibles impurezas como agregados de proteína que podrían dificultar la posterior cristalización de la misma (van der Laan *et al.*, 1989). Por lo tanto, esta columna ha sido el último paso de purificación de todas las proteínas en esta tesis, antes de los posteriores intentos de cristalización.

En el caso de la lectina, se cargaron 3 ml de la muestra en la columna de exclusión molecular HiLoad® 16/600 Superdex® 75 pg (GE Healthcare Life Sciences) previamente equilibrada con el mismo tampón. La inyección de la muestra se realizó a 2 ml/min, y la elución de la misma se detectó mediante la medida de absorbancia a 280 nm (Creighton 1997).

## CROMATOGRAFÍA DE EXCLUSIÓN MOLECULAR



**Figura 2.1. Esquema del proceso de cromatografía de exclusión molecular.** Esta técnica permite separar moléculas en función de su tamaño. Las de menor peso (representadas en la figura como esferas de color rosa) quedan retenidas en la matriz de la columna, mientras que las de mayor tamaño (esferas amarillas) no caben por los poros de la misma y eluyen antes.

**Fuente:** <https://lidiakonlaquimica.wordpress.com>.

### 2.1.2 Cuantificación de proteínas por absorción UV-Visible

Esta técnica aprovecha la característica de absorción de la luz UV de las proteínas con aminoácidos aromáticos (tirosina y especialmente triptófano) que presentan un pico a 280 nm debido a sus enlaces  $sp^2$  en los dobles enlaces de los anillos aromáticos (Creighton 1997). Se utilizan servidores informáticos,

PROTPARAM en nuestro caso, para calcular el coeficiente teórico de las proteínas en función de su secuencia de aminoácidos y poder realizar el cálculo de su concentración en disolución (Noble y Bailey 2009; Ecuación 1).



**Figura 2.2. Espectrofotómetro de microvolúmenes.** Fluorómetro DeNovix DS-11 FX utilizado para la medida y cálculo de la concentración de las proteínas en la tesis. **Fuente:** <https://unitedtrading.com.pe/producto/espectrofotometro-ds-11-fx/>

La cuantificación de la muestra se ha llevado a cabo tras cada paso de purificación en cada una de las proteínas de esta tesis. La muestra de lectina fue cuantificada por su medida en un espectrofotómetro de microvolumen (DeNovix® DS-11 FX) (Figura 2.2). Siendo el coeficiente de extinción teórico de la lectina dado en PROTPARAM de 40450 mL mg<sup>-1</sup> cm<sup>-1</sup>, el propio equipo aplica la ecuación de Lambert-Beer (Ecuación 1) para el cálculo de la concentración en la muestra medida.

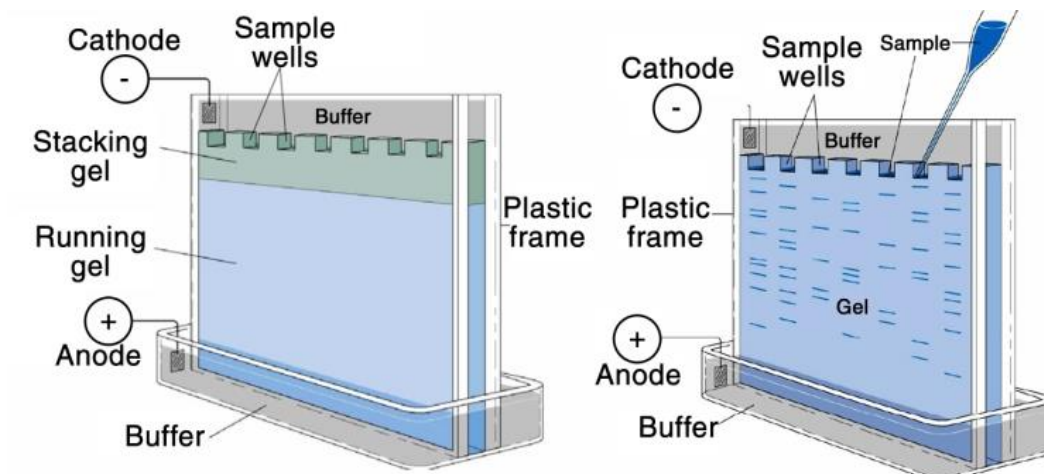
$$Abs = c * l * \epsilon$$

**Ecuación 1.** Ecuación de Beer-Lambert utilizada para calcular la concentración de proteína por medida de la absorbancia a 280 nm. Donde Abs es la absorbancia obtenida, c (mol l<sup>-1</sup>) la concentración que se quiere calcular, l la longitud del paso óptico, en este caso 0,1 cm, y  $\epsilon$  (l mol<sup>-1</sup> cm<sup>-1</sup>) el coeficiente de extinción molar.

### 2.1.3 Electroforesis desnaturalizante SDS-PAGE y tinción

La presencia y pureza de las diferentes proteínas de interés se comprobó mediante electroforesis SDS-PAGE (Figura 2.3) de las fracciones recogidas, previamente hervidas 5 minutos a 100 °C.

La electroforesis desnaturalizante en geles de poliacrilamida es una técnica clásicamente empleada para la separación y visualización de proteínas durante y tras la purificación de las mismas (Fishman y Moyer, 1942). El fundamento de la misma reside en la diferente migración de las muestras desnaturalizadas y cargadas negativamente en función de su peso molecular al ser sometidas a una diferencia de potencial eléctrico.



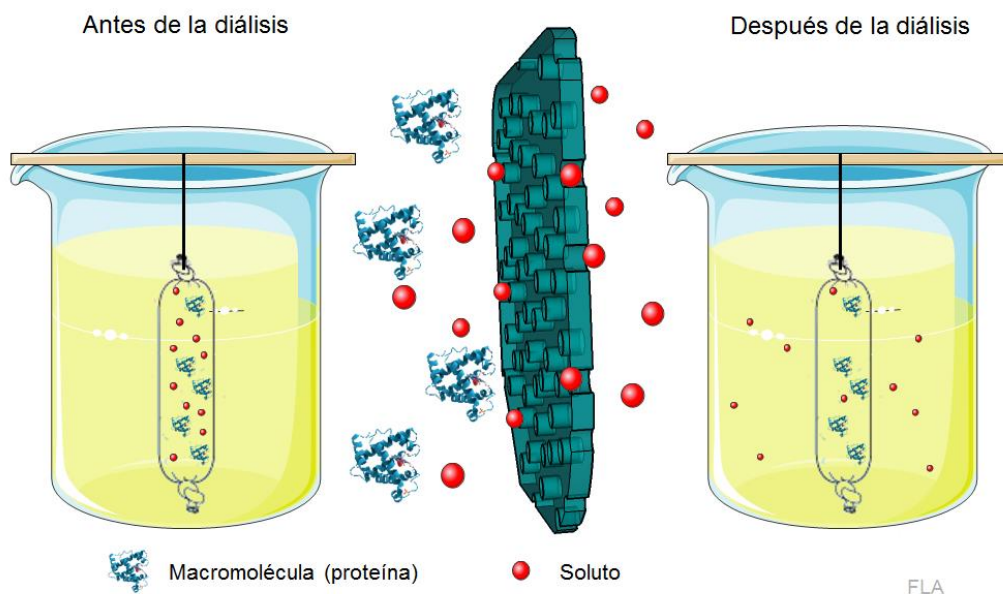
**Figura 2.3. Diagrama de una electroforesis desnaturalizante.** Gel desnaturalizante SDS-PAGE utilizado para resolver las bandas de proteína obtenidas tras cada paso de purificación. **Fuente:** J. Muntané Relat Introducción a la investigación básica RAPD 2010 VOL 33 N3 Mayo - Junio 2010

Para realizar un gel SDS-PAGE se requieren dos tipos de geles de poliacrilamida. Un gel denominado “de compactación” que estará en la zona superior ocupando un tercio del total y sirve para compactar todas las muestras cargadas en los pocillos. En segundo lugar, se sitúa el gel donde las muestras se resolverán llamado “de resolución”. Los componentes son los siguientes: 40 % Acrilamida, 1,5 M TRIS pH 8,8 (resolución)/ 1 M TRIS pH 6,8 (compactación), 20 % SDS, 10 % APS y 0,01 % TEMED, en distintas proporciones según el peso

de la proteína para resolverse a unos tamaños u otros. Para separar proteínas grandes se usan porcentajes de acrilamida pequeños, y viceversa. La presencia del detergente SDS va a desnaturalizar las proteínas, así como la fase de hervido de las muestras, permitiendo que éstas migren según su peso molecular en el gel. Este componente se añade junto a un agente de tinción que permite seguir el frente mientras migra. Es también necesario añadir un marcador de pesos moleculares (BLUEstain™ Protein Ladder de Thermo Scientific™) en un carril contiguo para contrastar la migración de esas bandas control con tu muestra.

En las distintas purificaciones descritas en esta tesis se han empleado geles de acrilamida desnaturalizantes, conteniendo un 10% o un 12% de acrilamida en la capa de resolución. La visualización de las bandas proteicas se ha llevado a cabo por tinción con Azul de Coomassie y posterior destinción con una disolución de metanol y ácido acético (40% MetOH y 10% ácido acético), tras la que sólo quedan en el gel las bandas correspondientes a la proteína presente en el mismo.

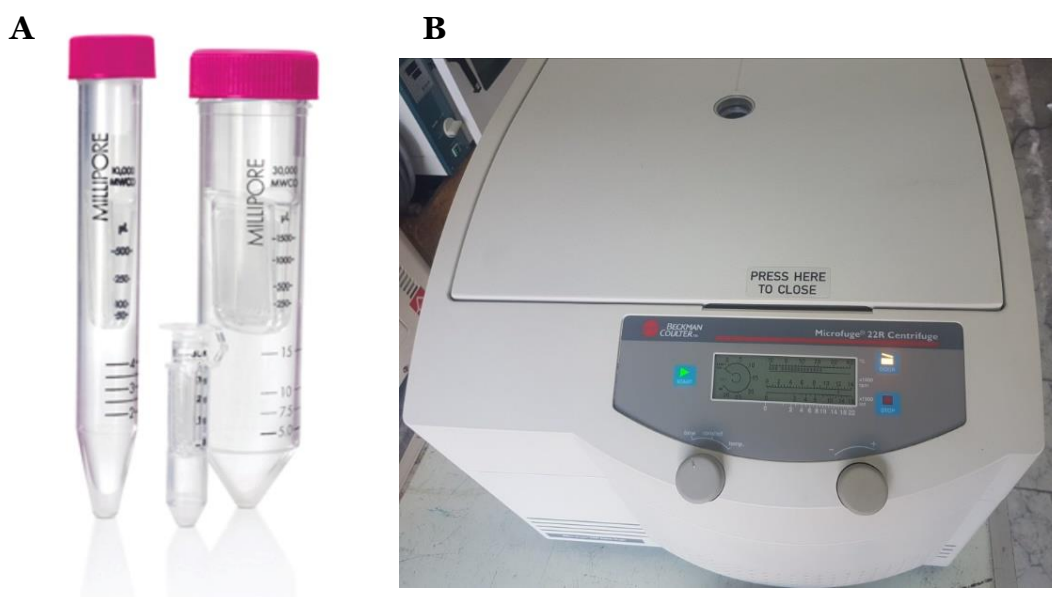
#### 2.1.4 Diálisis y concentración de la proteína



**Figura 2.4. Esquema de una diálisis por intercambio en solución.** El soluto atraviesa la membrana mientras la proteína no lo hace. **Fuente:** <http://tustareas.lat/index.php/biologia/item/3926-cromatografia-e-isotopos>

Las fracciones de proteína purificada se dializaron en un tampón de diálisis 25 mM TRIS pH 7,5 a 4°C. Las membranas de diálisis permiten realizar un cambio de tampón mediante difusión osmótica, logrando retirar la sal de nuestra muestra sin que la proteína escape de ella (Figura 2.4).

Posteriormente, se concentró la muestra de proteína por ultrafiltración en concentradores Centricon con un corte de poro de 10 kDa (Amicon Ultra-15 centrifugal Filter, Merck Millipore; Figura 2.5) hasta alcanzar una concentración aproximada de 16 mg/ml. Se empleó para ello una centrífuga Microfuge® 22R (Beckman Coulter®), y la muestra se conservó a 4° C en hielo para su uso inmediato. Si la proteína se quería preservar para usos más allá de 48 horas después de su purificación, el método de conservación fue la congelación en nitrógeno líquido y el almacenamiento en un congelador de -80 °C (New Brunswick Scientific™ ultra low temperature freezer).



**Figura 2.5. Tipos de Centricon y centrífuga.** (A) Ejemplos de Centricon utilizados en la concentración de muestra en la tesis en función del volumen. (B) Equipo de centrifugación donde se colocan los centricones. **Fuente:** <https://www.fishersci.es/shop/products/emd-millipore-amicon-ultra-2-0-centrifugal-filter-units-5/p-3748116> y [https://articulo.mercadolibre.com.mx/MLM-615834350-centrifuga-beckman-coulter-\\_JM](https://articulo.mercadolibre.com.mx/MLM-615834350-centrifuga-beckman-coulter-_JM)

## 2.2 TECNICAS BIOFISICAS PARA ESTUDIOS DE INTERACCIÓN

### 2.2.1 Calorimetría de Titulación Isotérmica (ITC)

La Calorimetría de titulación isotérmica (ITC) es una técnica biofísica utilizada para determinar los parámetros termodinámicos de interacciones en solución. Se utiliza con mayor frecuencia para estudiar la unión de pequeñas moléculas (tales como compuestos medicinales) a macromoléculas grandes (proteínas, ADN, etc.). (Freyer y Lewis 2008)

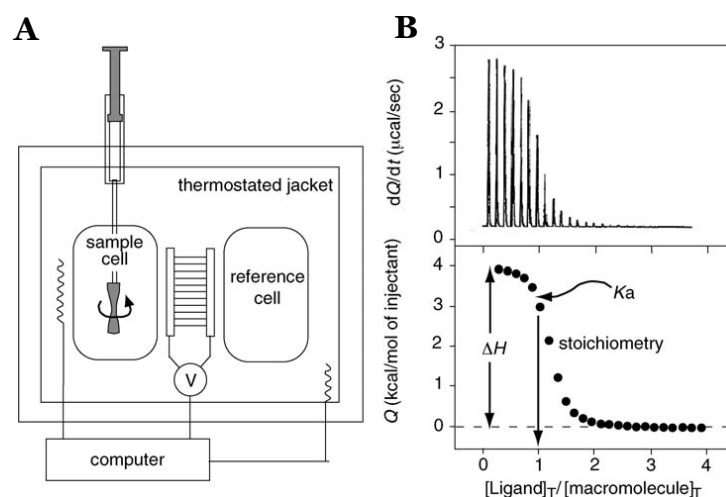
La interacción entre la lectina SBA y los péptidos de interés se caracterizó con un calorímetro MicroCal AutoITC200 (MicroCal, GE Healthcare Life Sciences).

Dentro de la calorimetría podemos distinguir varias técnicas. Sin embargo, es la calorimetría de titulación isotérmica (ITC) la que más comúnmente se utiliza en los estudios de asociación/disociación de ligandos con macromoléculas.

El procedimiento se basa en una serie de inyecciones del ligando sobre la macromolécula, que provocarán absorción o desprendimiento de calor debido a la interacción entre ellos (Brown 2009).

Como el nombre de la técnica indica, se precisa que el procedimiento sea isoterma, por lo que el mecanismo aportará o retirará energía para mantener la temperatura constante en el aparato. Se mide la cantidad de energía en  $\mu\text{cal/s}$  necesarias para mantener constante la diferencia de temperatura entre la muestra y la celda de referencia. Un perfil de una curva ejemplo se observa en la figura 2.6. En nuestro caso, se pensaba que se obtendría un perfil tipo sigmoidal donde se observaría un punto de inflexión que coincide con el ratio molar ( $n$ ). Este parámetro representa la estequiometría del complejo formado entre la proteína y el ligando. Además, esta técnica nos permite observar otros parámetros como la constante de asociación ( $K_B$ ) y la variación de entalpía ( $\Delta H$ ), que se pueden obtener directamente de un solo experimento. El procesamiento de los datos consiste en la integración del área de cada uno de los picos.





**Figura 2.6. Esquema del proceso de ITC.** (A) Esquema de la estructura interna del equipo de calorimetría. (B) Ejemplo de patrón obtenido tras el ensayo de ITC para analizar los parámetros de afinidad. Arriba se muestran las variaciones de calor recogidas por el equipo y debajo se muestran los valores normalizados para obtener los valores termodinámicos. (Velázquez-Campoy, 2004)

Se realizó un cribado de tampones para determinar el más adecuado para la recogida de datos de nuestra muestra, dado que el tampón en ocasiones puede apantallar la señal o producir ruido de fondo, lo que a su vez repercute negativamente en los resultados y el cálculo de las constantes de afinidad. Otro de los requisitos fue la comprobación de las concentraciones reales de la proteína y el péptido. La relación para determinar la constante de afinidad depende de la estequiometría, pero la concentración a añadir de las mismas debe determinarse para obtener una señal mínima que permita realizar posteriormente los cálculos de afinidad.

Una vez establecidos estos valores, el sistema estaba listo para usarse con las siguientes condiciones en cada pocillo (Tabla 2.1).

**Tabla 2.1. Condiciones de concentración y volumen del pocillo para el experimento de ITC con lectina SBA y sus glicopéptidos.**

Proteína SBA	Péptidos
40-120 $\mu\text{M}$ (300 $\mu\text{l}$ )	2-3 mM (300 $\mu\text{l}$ )

Esta técnica nos permitió determinar las constantes de afinidad péptido-SBA para cada uno de los glicopéptidos utilizados en el artículo I, valorando así la importancia que tiene la estructura peptídica y el azúcar en la unión.

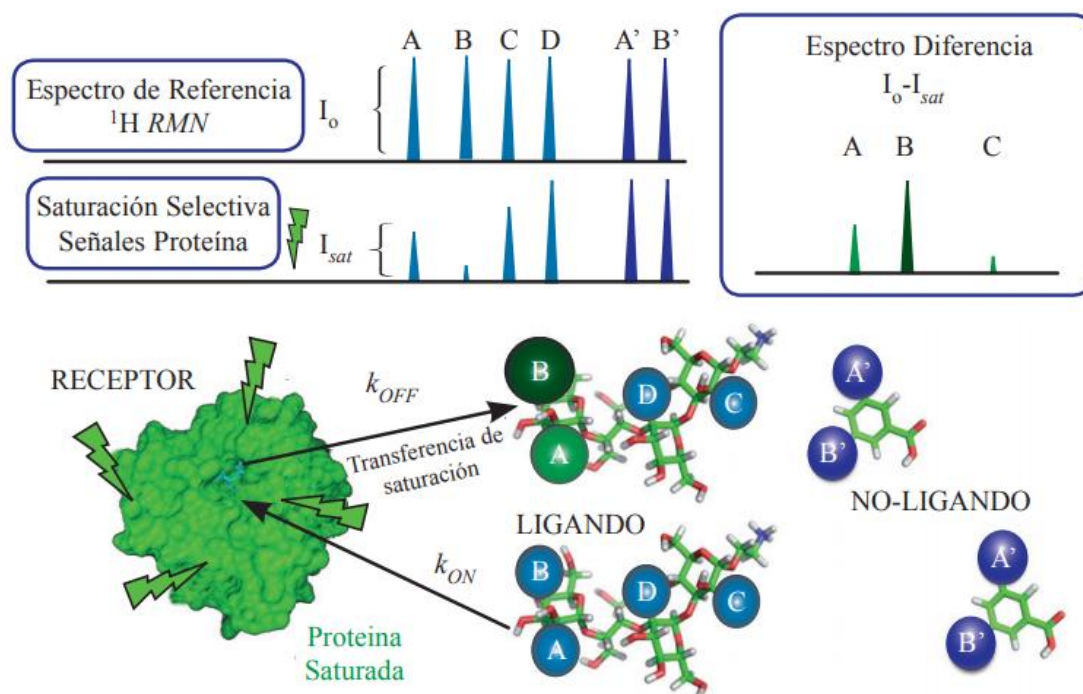
### 2.2.2 Resonancia magnética de diferencia de transferencia de saturación (STD)

Esta variante de Resonancia Magnética Nuclear (RMN) permite detectar y caracterizar interacciones receptor-ligando en disolución, observando para ello las señales de la molécula pequeña (ligando), que presenta propiedades espectroscópicas apropiadas para estudios a alta resolución (Angulo y Nieto, 2011).

La base de un experimento de STD es la siguiente (Angulo y Nieto, 2011; Figura 2.7):

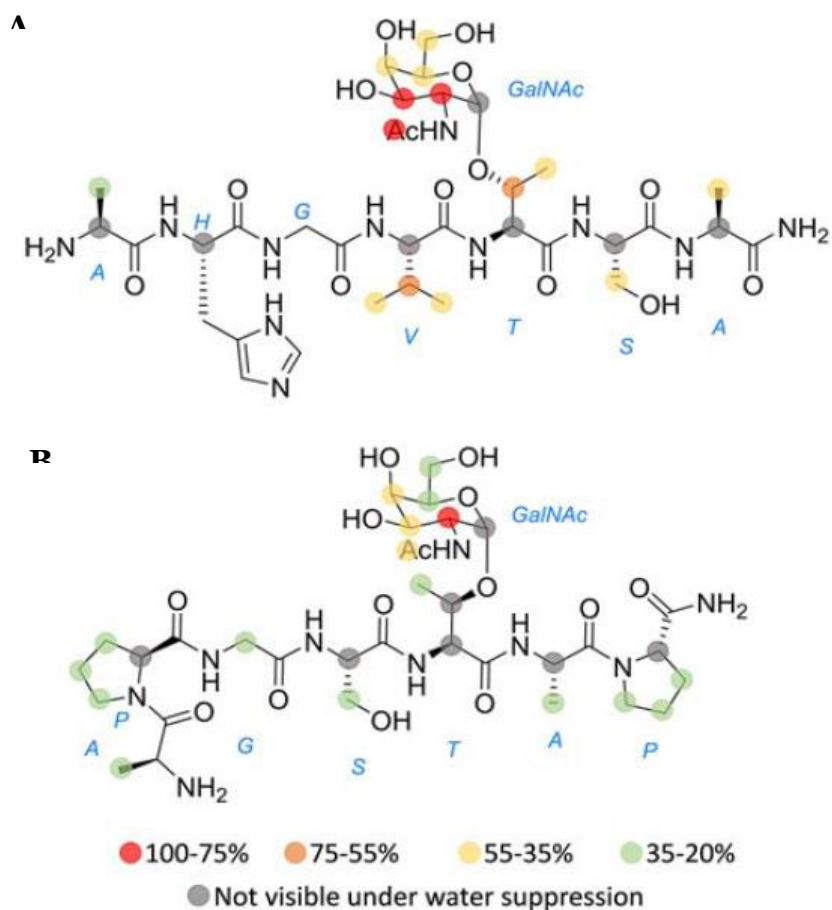
- En primer lugar, se realiza un espectro monodimensional  $^1\text{H}$  estándar de la mezcla de proteína y un exceso de ligando a una frecuencia de irradiación alejada de la frecuencia de resonancia del ligando o de la proteína. Dicho espectro, llamado off-resonance, mostrará las intensidades de referencia del ligando (las de la macromolécula no se ven).
- Se realiza un segundo experimento en el que la proteína se satura de forma selectiva irradiando, por ejemplo, en la frecuencia de los carbonos alifáticos. La saturación selectiva se transfiere al conjunto de la proteína por el denominado efecto NOE intramolecular (NOEs), que se ve favorecido por el gran tamaño de las macromoléculas. De igual modo, esta saturación se transfiere desde la proteína hasta aquellas zonas del ligando que estén directamente unidas a la misma. Todas estas señales saturadas no aparecerán en el espectro.
- La diferencia o resta matemática de las señales del segundo espectro respecto a las del primero es lo que nos proporciona el espectro STD, que solamente contiene las señales asignables a los protones del ligando que

están en contacto directo con la superficie de la proteína. El valor de sus intensidades serán un reflejo de la proximidad del protón a dicha superficie.



**Figura 2.7. Representación esquemática del experimento 1D STD NMR en una muestra.** Se realizan dos experimentos: un 1D  $^1\text{H}$  NMR estándar, y un 1D  $^1\text{H}$  NMR en el cual la proteína ha sido presaturada selectivamente. Las moléculas que interaccionen con la proteína recibirán parte de dicha saturación por NOE-intermolecular en el estado enlazado. La diferencia entre ambos espectros revela exclusivamente las señales del ligando. **Fuente:** Jesús Angulo, Pedro M. Nieto 2011

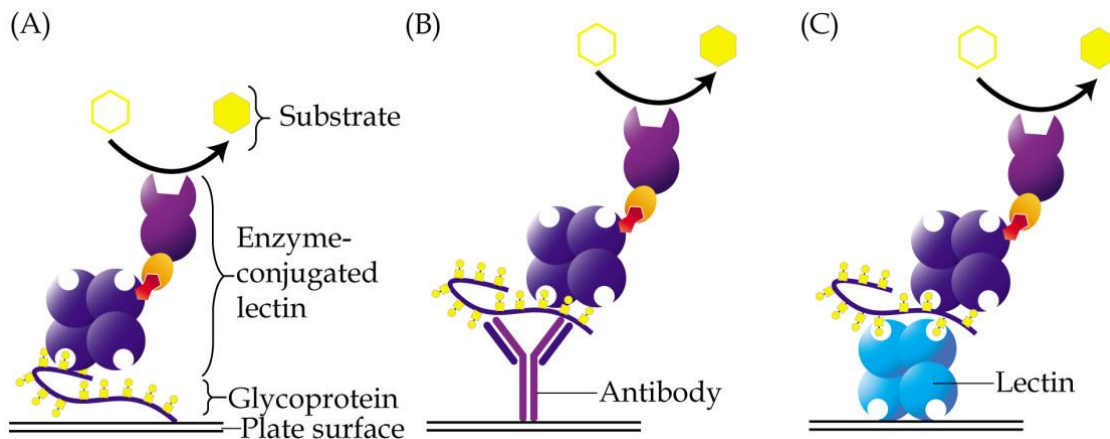
Los experimentos de STD nos van a permitir observar cómo es la interacción de los glicopéptidos a nivel atómico. En la figura 2.8 vemos un ejemplo del tipo de datos que se obtienen una vez procesados los espectros, donde los colores indican el diferente grado de interacción proteína-ligando que ha tenido lugar para cada uno de los átomos de la molécula de ligando.



**Figura 2.8.** Mapeo de los epitopos de los glicopéptidos **1** (A) y **3** (B) unidos a lectina SBA deducidos por experimentos de STD **Fuente:** Figura S16 del material suplementario SI (artículo I)

### 2.2.3 Ensayo de unión Lectina -Enzima (ELLA)

El ensayo de unión Lectina-Enzima (ELLA) consiste en fijar en la placa el carbohidrato/glicopéptido que el conjugado debe reconocer (Figura 2.9). Este conjugado es muy eficiente debido a que se eluye de las propias células. Cuando se utilizan lectores de placas de ensayos ELISA, permite una mejor detección y cuantificación relativa de unidades de carbohidratos de superficie específicas que la que puede obtener la inmunofluorescencia estándar con lectinas conjugadas con fluoresceína (Hashim, 2017).



**Figura 2.9. Ensayo de unión Lectina-Enzima.** Distintas aproximaciones del ensayo de interacción. La imagen A representa la interacción directa con el carbohidrato que reconoce el conjugado. La imagen B representa el ensayo híbrido, con un anticuerpo que une el glicopéptido para presentarlo al conjugado. La imagen C representa el ensayo denominado “sándwich”, que utiliza una lectina para anclarse a la superficie de la placa y presentar el carbohidrato al conjugado. **Fuente:** Hashim, 2017

Este ensayo se desarrolló con el objetivo de detectar unidades específicas de carbohidratos en la superficie de células en suspensión. Se utiliza también en ensayos de inhibición ya que permite utilizar muestras de suero.

## 2.3 TÉCNICAS BIOFÍSICAS DE ANÁLISIS ESTRUCTURAL

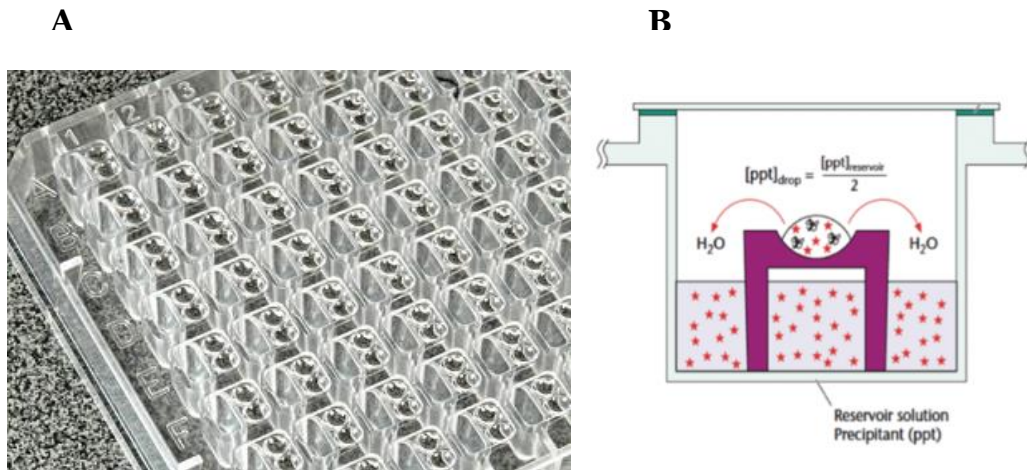
### 2.3.1 Cristalización de proteínas

La cristalización de proteínas es un proceso de ordenamiento regular en el espacio tridimensional por parte de las moléculas de proteína pura para formar un cristal. Este proceso se da en unas situaciones de concentración y equilibrio específicos para cada proteína, que hay que optimizar con cribados de distintas condiciones. Obtener estas estructuras cristalinas supone una serie de dificultades experimentales (Müller, 2017) pero nos va a permitir, gracias a su difracción por rayos X y su posterior procesamiento resolver las estructuras tridimensionales y caracterizar estructuralmente las proteínas (Figura 2.12). La cristalización de proteínas es la técnica más empleada para la caracterización estructural de las mismas, responsable de más del 86% de las estructuras depositadas en el Protein Data Bank (PDB) (Dessau y Modis, 2011).

### 2.3.2 Co-cristalización de proteínas

El procedimiento habitual que se ha utilizado para la obtención de cristales de cada una de las proteínas en esta tesis ha sido la difusión de vapor en gota sentada (Davies y Segal, 1976) (Figura 2.10). Se realizó una co-cristalización de los glicopéptidos con la lectina (SBA). Esta técnica consiste en añadir conjuntamente tanto el glicopéptido como la proteína para formar el complejo. El método habitual de cristalización suele ser, como se ha explicado en el anterior apartado, utilizar proteínas puras. La co-cristalización tiene la peculiaridad de usar un complejo que actúa a la hora de ordenarse en el espacio, como una proteína homogénea, lo que nos permitirá ver su interacción tras la difracción y resolución de su estructura.

Se utilizaron distintos cribados comerciales para la obtención de cristales. JCSG HTS, Classic I, Basic, PEG-SALT (Jena Bioscience), AmSO<sub>4</sub> suite (QUIAGEN), Proplex, Midas y Morpheus (Molecular Dimensions) fueron los elegidos, con 96 condiciones cada uno.



**Figura 2.10. Cristalización de proteínas.** (A) Placa de cristalización para la difusión de vapor en gota sentada. **Fuente:** Fisher Scientific. (B) Esquema de la difusión de vapor en gota sentada (McPherson, 2001).

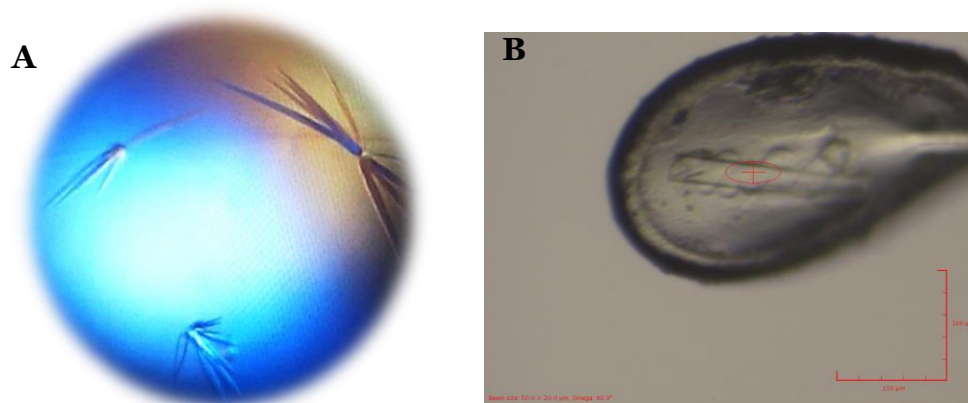
Se utilizaron placas de 96 pocillos (Axygem™ 96 well), con capacidad de dos gotas por pocillo (gota A y B). En cada pocillo se cargaron 60 µl de solución del agente precipitante y dos gotas. La gota A consistió en 1 µl de solución con proteína (lectina 8,5 mg/ml + péptido n°1 a 10 mM) y 1 µl de solución del agente precipitante, mientras que la gota B consistió en 1 µl de solución con proteína (lectina 8,5 mg/ml + péptido n°2 a 10 mM) y 1 µl de solución del precipitante. Para el péptido 3 (ver artículo en la sección de resultados) se usó una placa nueva que solamente contuvo la lectina y este péptido en la gota A.

Una vez preparadas, las placas se sellaron con cinta adhesiva transparente para evitar su evaporación y se almacenaron en una incubadora a 18°C (Formulatrix) para el almacenaje de placas de cristalización, acoplada a un reproductor de imágenes. Se tomaron fotos de las placas a tiempos de 0 h, 12 h, 1 día, 2, 3, 5, 8, 13 y 21 días.

### 2.3.3 Toma de cristales y congelación

Se obtuvieron cristales de lectina en complejo con el glicopéptido 2 en soluciones de 20-30% PEG 3350, 0,2 M NaBr y 20-30% PEG 400 como aditivo. La propia solución madre fue usada posteriormente como crioprotector.

Una vez obtenidos los cristales, debemos transportarlos para que puedan ser difractados y así obtener la estructura del complejo. En esta operación se utilizaron *crioloops* (Hampton®), hebras sintéticas superenrolladas formando una hélice que permiten manipular el cristal (Figura 2.11). Gracias a esta herramienta, se capturó el cristal de la condición donde ha crecido en la placa, para incubarlo breves segundos en un crioprotector (PEG 400 en el caso de la lectina), evitando así que se deteriore en su transporte con nitrógeno líquido hasta el lugar donde van a ser difractados. El crioprotector consta, generalmente, de las mismas condiciones en las que se desarrolló el cristal adicionando un 20% de un alcohol de pequeño tamaño (PEG menores de 1000 Da de peso molecular, etilenglicol, glicerol, etc) (Parkin y Hope, 1998; Kwong y Liu, 1999; Bujacz *et al.*, 2010).

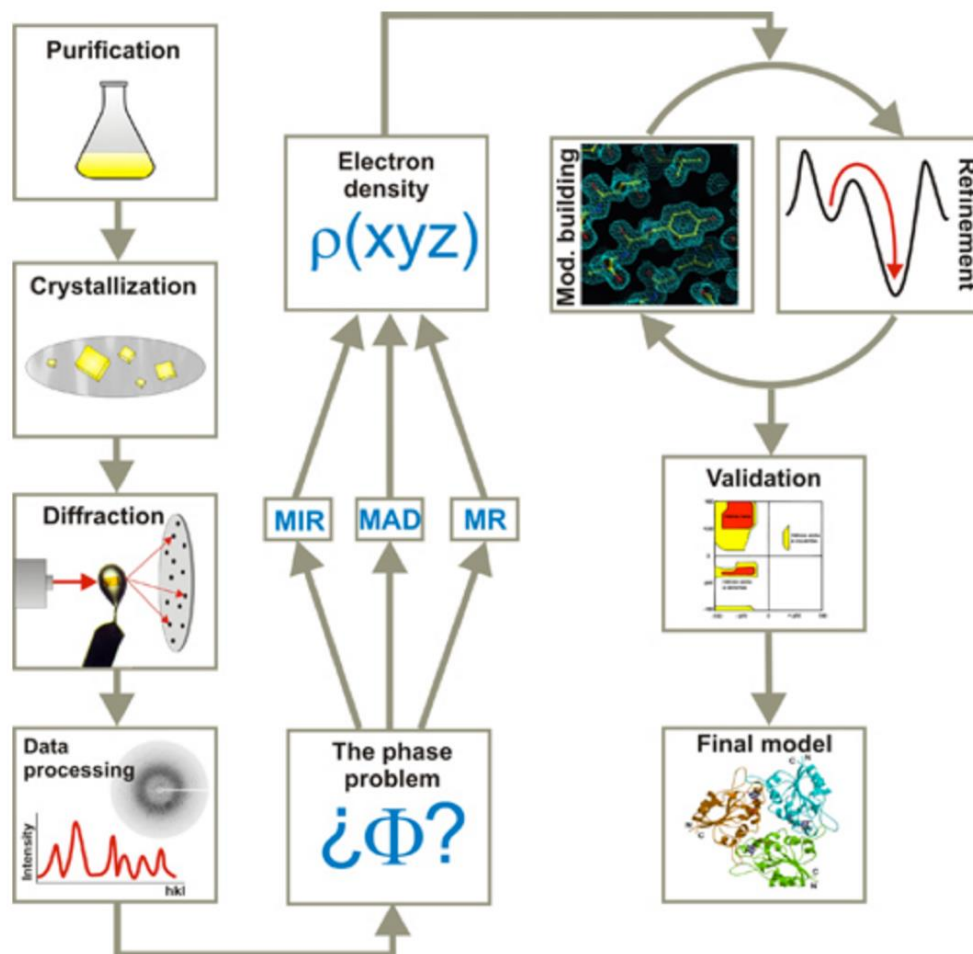


**Figura 2.11. Cristales de proteína.** (A) Foto de cristales de SM3 generados por difusión de vapor en gota sentada (B) Cristal durante el proceso de enfoque del haz de rayos X en un *crioloop*.

#### 2.3.4 Difracción de rayos X

La difracción es un fenómeno óptico, que se produce cuando las ondas encuentran un obstáculo de longitud similar a la longitud de onda incidente. La radiación electromagnética de rayos X consiste en ondas que oscilan entre 0,1 Å a 100 Å de longitud de onda ( $1 \text{ Å} = 10^{-10} \text{ m}$ ), que son emitidas por los electrones excitados que vuelven a estados atómicos energéticamente basales. Será necesario el uso de la radiación del mismo orden de magnitud que los objetos atómicos para calcular sus características atómicas. En los sincrotrones se generan rayos X con una longitud de onda de aproximadamente 1 Å. En las proteínas, la longitud de los enlaces químicos covalentes entre el carbono, el nitrógeno y el oxígeno varía de 1,24 Å (C = O) a 1,53 Å (C-C $\alpha$ ); estas dimensiones están muy cerca de la longitud de onda de rayos X y, por tanto, este tipo de radiación nos permitirá "ver" la disposición tridimensional de los átomos de la proteína.



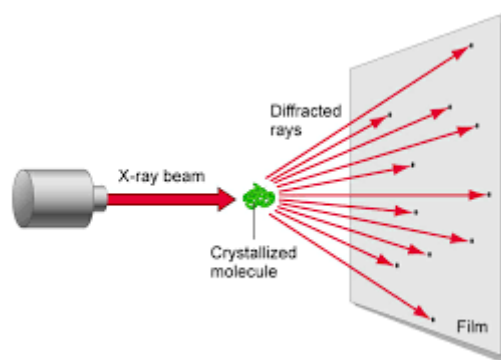


**Figura 2.12 Esquema de la técnica de la difracción de rayos X de un cristal de proteína hasta obtener su estructura tridimensional.** Primero es necesario la obtención de un cristal de proteína para que sea difractado por rayos X y así obtener un patrón de difracción. Posteriormente los datos serán procesados y la estructura de la proteína se podría obtener o por reemplazamiento molecular o técnicas de faseado (Párraga *et al.*, 2010).

Se requiere de un sincrotrón para acelerar los electrones hasta una velocidad cercana a la de la luz, que al cambiar su trayectoria liberarán radiación electromagnética X. Son mucho más precisos y potentes que los de un laboratorio común porque pueden producir longitudes más bajas y de mayor intensidad. Los datos obtenidos de los cristales en esta tesis doctoral han sido recogidos en varios sincrotrones. Uno de ellos fue el sincrotrón ALBA, en la línea BL-13 XALOC, localizado en Barcelona. El otro fue el sincrotrón Diamond

Light Source (DLS), en las líneas I03 e I04, ubicado en Oxford. Ambos presentan características similares, si bien la potencia y capacidad de procesamiento del segundo son mucho mayores, por lo que los datos se pueden tomar a mayor velocidad.

Los cristales se ordenan en redes cristalinas, que son matrices ordenadas de celdillas unidad repetidas continuamente en el espacio. Esta celdilla se define por tres distancias ( $a$ ,  $b$ ,  $c$ ) y tres ángulos ( $\alpha$ ,  $\beta$ ,  $\gamma$ ). Los índices de Miller son un conjunto de tres números ( $h$ ,  $k$ ,  $l$ ) que se utilizan para definir una familia de planos especificando las pendientes de los planos. También se utilizan para describir los “spots” que surgen debido a la difracción de estos planos, como veremos a continuación.



**Figura 2.13. Proceso de obtención de datos por difracción.** Difracción de un cristal de proteína por rayos X que debe ser analizado para resolver la estructura tridimensional de la proteína. Hoy en día, los datos se recogen en un detector que computa los valores obtenidos. Antiguamente esta labor se realizaba a mano, los spots se recogían en películas que luego debían de ser interpretados y analizados manualmente.  
**Fuente:**[http://www.biochem.arizona.edu /classes/bioc462/462bh2008/462bhonorsprojects/462bhonors2006/cooperj/protein%20structure%20analysis.htm](http://www.biochem.arizona.edu/classes/bioc462/462bh2008/462bhonorsprojects/462bhonors2006/cooperj/protein%20structure%20analysis.htm)

### 2.3.5 Análisis de difracción

Tras la difracción, se obtiene el patrón de difracción compuesto de puntos o “spots” que debe ser interpretado para obtener el mapa de densidad electrónica de la proteína, o en nuestro caso, del complejo proteína-glicopéptido

(Figura 2.13). La posición de los “spots” en el espacio depende de la celdilla unidad y la intensidad de los “spots” depende de la disposición de los átomos. Se integran las intensidades de los “spots” de todas las imágenes obtenidas con el fin de obtener un grupo espacial probable.

La densidad de electrones en un punto  $x, y, z$  puede calcularse a partir de la transformada de Fourier sobre la medida de los factores de estructura  $hkl$ :

$$\rho(x, y, z) = 1 / V \sum_{hkl} |F_{hkl}| \exp[-2\pi i (hx + ky + lz) + i\alpha(hkl)]$$

**Ecuación 2.** Transformada de Fourier para calcular el mapa de densidad electrónica de un cristal de proteína

La otra mitad de la información, la fase ( $\alpha$ ) no se encuentra ya que no se puede determinar directamente. Se trata del problema de fase que no puede ser determinado por ningún detector y requiere del uso de métodos experimentales o de cálculo. Este problema se resuelve por reemplazamiento molecular usando estructuras parecidas previamente ya obtenidas como fue nuestro caso. Se utiliza una estructura homóloga como molde o “template” para resolver el problema de las fases y obtener un modelo que se ajuste a los mapas de densidad obtenidos.

Los datos se procesaron con XDS (Kabsch, 2010) y el software CCP4 (Winn *et al.*, 2011), posteriormente se usó “SCALA” para compilar los mismos según sus índices  $hkl$ . La estructura fue resuelta mediante reemplazamiento molecular usando “Phenix” (Adams, 2010) y el PDB 1SBF como “template”. Posteriormente se usó “Coot” (Emsley y Cowtan, 2004), y “REFMAC5” (Murshudov *et al.*, 2011) para la construcción y refinamiento de la lectina. Los modelos finales se validaron con “PROCHECK” (Laskowski, 1993) y las estadísticas de los modelos se presentan en la tabla S4 del material suplementario I (SI). Finalmente, las coordenadas y factores estructurales de cada estructura se depositaron en el World Wide Protein Data Bank (wwPDB)

con el número de entrada 4d69. La unidad asimétrica de estos cristales contiene 12 moléculas de lectina.

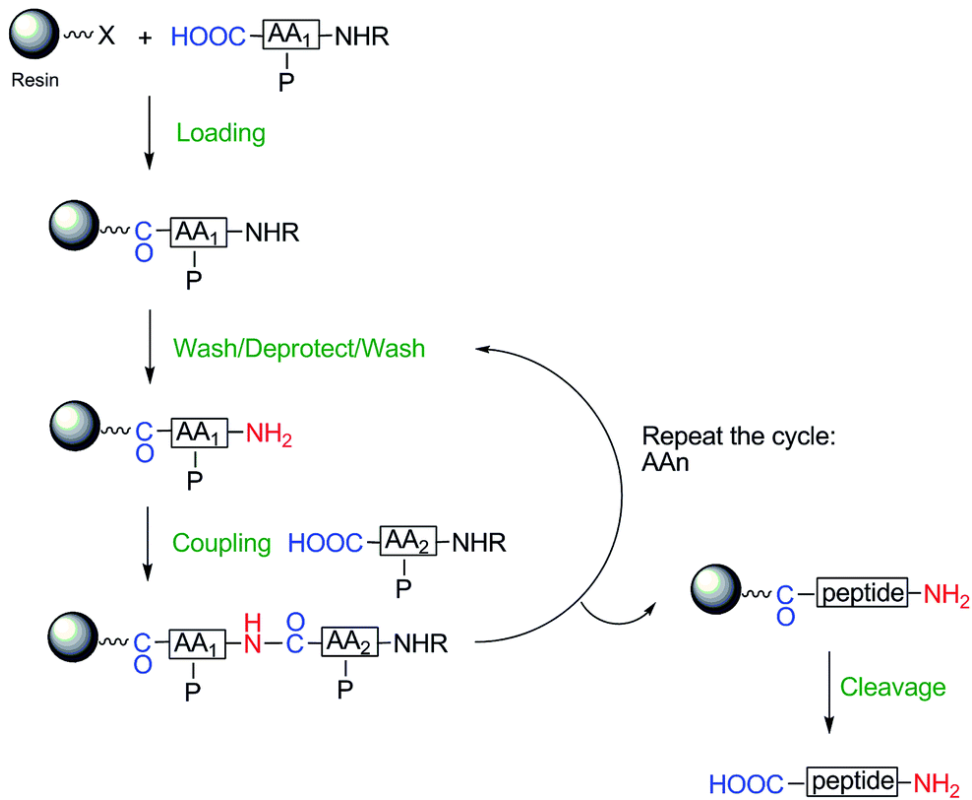
## 2.4 PROCESOS DE SÍNTESIS DE GLICOPÉPTIDOS

### 2.4.1 Síntesis SPPS, métodos Fmoc y Boc

Los diferentes péptidos y glicopéptidos empleados en la tesis han sido diseñados de forma racional, tal y como se explica en el desarrollo y en la información suplementaria correspondiente a cada uno de los artículos. Todos los glicopéptidos utilizados en esta tesis se generaron por síntesis en fase sólida (SPPS) usando la estrategia Fmoc.

En la síntesis SPPS, cada uno de los aminoácidos que va a reaccionar se encuentra unido a una esfera sólida para construir la cadena peptídica (Figura 2.14). A diferencia de la síntesis proteica en ribosomas, esta se produce desde el extremo carboxilo hacia el amino terminal. Por ello supone múltiples ventajas como sintetizar péptidos naturales difíciles de expresar en bacterias, incorporar aminoácidos no proteicos, modificar el esqueleto de péptidos y proteínas y sintetizar proteínas a partir de D-aminoácidos. Puede llevarse a cabo de dos estrategias diferentes: Fmoc y Boc.

Esta síntesis consiste en una serie de ciclos acoplamiento-lavado-desprotección-lavado, donde el extremo amino desprotegido de un aminoácido reacciona con el carboxilo de otro, produciendo así la desprotección de su grupo amino para seguir así la cadena. Tras finalizar la síntesis, se realizó la liofilización de los péptidos, generando sales que deben ser rehidratadas en el buffer TRIS y ajustadas a un pH adecuado (entre 7-8) para la estabilidad de la proteína, como se detalla en el apartado 2.4.3.



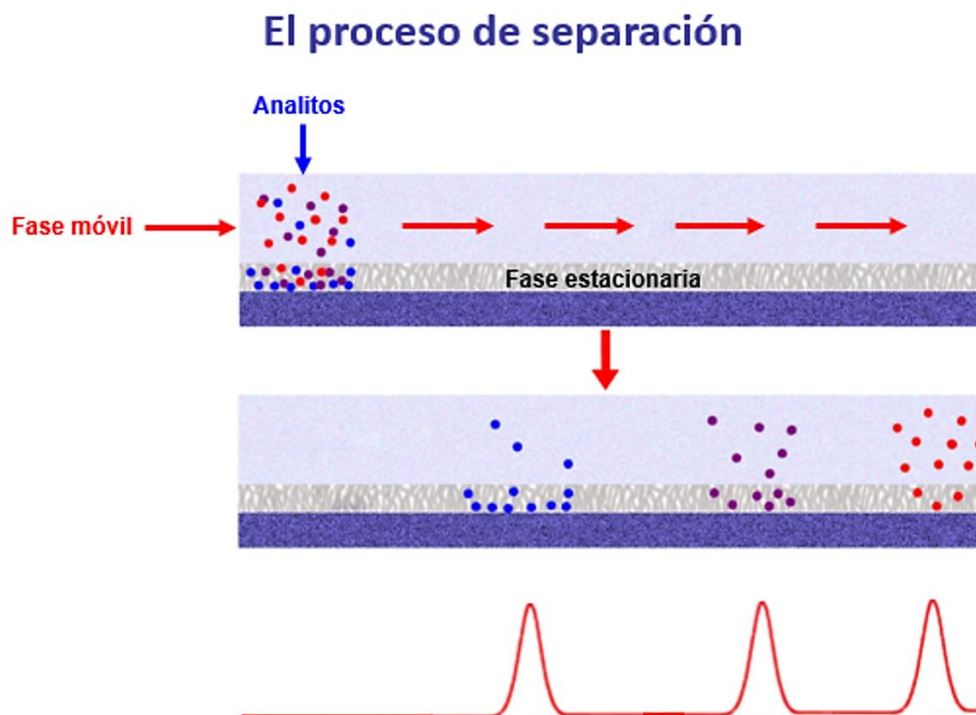
**Figura 2.14. Esquema de la síntesis de glicopéptidos SPPS.** El primer aminoácido se ancla a una resina para posteriormente ir añadiéndole el resto hasta generar el péptido deseado. (J.M. Palomo, 2014)

La estrategia Fmoc, utiliza grupos protectores basados en 9-fluorenilmetoxicarbonilo (Fmoc) unidos a los grupos amino del aminoácido. Fmoc se suele utilizar frente a otras estrategias porque tiene unas condiciones más suaves (sensible a bases como la piperidina) para liberar el grupo protector y puede ser utilizado para cuantificar la unión y eficiencia de la reacción al poder ser detectado por medida de UV.

La estrategia Boc, utiliza grupos protectores basados en terc-butiloxicarbonil (boc), sensibles a ácidos como el trifluoroacético (TFA). La aplicación de un protocolo de neutralización in situ (Schnolzer *et al*, 1992. Merrifield, 1986) ha hecho que la química BOC sea la estrategia seleccionada para aumentar la eficiencia en secuencias peptídicas difíciles o más largas incluso implementado en sintetizadores automatizados.

#### 2.4.2 Purificación de péptidos y glicopéptidos por HPLC

A diferencia de nuestras proteínas, para purificar las muestras en la síntesis de los péptidos y glicopéptidos en esta tesis se utilizó la cromatografía líquida de alta resolución (HPLC). En esta técnica se busca separar los componentes de una muestra y, al igual que en la purificación de proteínas por cromatografía de exclusión, se inyecta la muestra por una matriz con la que cada uno de los componentes va a interactuar en distinta medida, produciendo así la separación de los mismos (Figura 2.15). La elección de una buena fase móvil (líquido) va a ser de suma importancia para que la muestra interactúe con la fase estacionaria (matriz). La técnica de HPLC se ha utilizado para purificar los productos de la reacción tras la síntesis química de los compuestos.



**Figura 2.15. Esquema del proceso de separación por HPLC.** Las muestras que interactúan con la fase estacionaria, resultan más retenidas que el resto. La composición de la fase móvil también influye en las interacciones que se generan con la fase estacionaria. **Fuente:** <https://www.notijenck.com.ar/notas/que-es-la-hplc-y-como-funciona>

### 2.4.3 Disolución de los glicopéptidos

Todos los péptidos sintetizados se reciben puros y liofilizados. Debido a su carácter generalmente muy ácido, incompatible con el rango de pH en el que las proteínas son estables, se disuelven en una disolución 25 mM TRIS/HCl, pH 7,5 y se ajusta el pH con unos microlitros de disolución 2 M NaOH y papel indicador de pH (Fisher) hasta dejarlos a una concentración final de 150 mM de compuesto o péptido y un pH entre 7 y 8.

## 2.5 OTRAS TÉCNICAS PARA EL ANÁLISIS DE LAS ESTRUCTURAS MOLECULARES

### 2.5.1 Experimentos de 2D RMN (COSY, HSQC y NOESY)

La RMN de alta resolución constituye una de las técnicas de mayor utilidad de las empleadas en química para la determinación de estructuras moleculares. Cabe destacar que el premio Nobel de Física de 1952 fue para Felix Bloch y Edward Purcell por observar las resonancias nucleares al tratar de medir el momento magnético del  $^{14}\text{N}$ . (Emsley and Feeney, 1995)

Estos experimentos, por lo tanto, requieren de un equipo especializado denominado espectrómetro. Este equipo consta de varias partes: un imán que genere un campo magnético estable para poder definir la frecuencia de resonancia de cada núcleo; una sonda, dentro del imán, responsable de emitir y recibir las radiofrecuencias; una consola que genere los pulsos de radiofrecuencia; y el ordenador que analice todos los datos obtenidos.

Aunque la RMN puede utilizarse en muestras sólidas, generalmente se utilizan experimentos de RMN en disolución ya que permite el estudio de estructuras tridimensionales en disolución, lo que hace que esta técnica se pueda complementar esta técnica con la resolución de estructuras cristalinas por rayos X.

Se utilizaron los siguientes experimentos, cada uno específico para conocer una información distinta de la estructura de los péptidos como se detallará en la sección de resultados.

- COSY es el acrónimo del experimento “Espectroscopía de correlación” y nos permite determinar espines acoplados entre sí, o dicho de otra forma, cómo de cerca se encuentran dos átomos según los enlaces que los separan (correlación homonuclear). Se trata de una técnica de RMN de dos dimensiones que consiste en dos pulsos espaciados en un tiempo  $t_1$ . Tras realizar el segundo pulso se recoge la señal para obtener un espectro bidimensional.

Generalmente muestra la frecuencia de un solo isótopo en ambos ejes, aunque existen técnicas para recoger más. Esto genera un patrón de “manchas” en los que se observan dos tipos de picos: diagonales o cruzados. Los picos diagonales aparecen en la diagonal, como su nombre indica, y representan los valores obtenidos del patrón de resonancia del isótopo de H. Por el contrario, los picos cruzados aparecen fuera de la diagonal, lo que implica un acoplamiento entre pares de núcleos.

- HSQC es el acrónimo de “Espectroscopía de correlación cuántica simple heteronuclear” y detecta correlaciones entre núcleos diferentes separados por un enlace (correlación heteronuclear). Esta técnica funciona transfiriendo la magnetización desde el primer núcleo (generalmente el protón) al segundo núcleo (generalmente el heteroátomo) usando una secuencia de pulsos. Este primer paso se realiza porque el protón tiene una mayor magnetización de equilibrio y, por lo tanto, este paso crea una señal más fuerte. La magnetización luego evoluciona y luego se transfiere de regreso al primer núcleo para su observación.
- NOESY es el acrónimo del experimento “Espectroscopía de mejoramiento nuclear de Overhauser” y nos va a permitir determinar que hidrógenos de la molécula se encuentran cercanos entre sí, independientemente de los enlaces (correlación a través del espacio), gracias a solapar dos espectros de resonancia entre sí. Requiere el uso de un equipo de resonancia nuclear para poder obtener el diagrama que nos muestre los acoplamientos de los H. Este diagrama se muestra como un cuadro marcado con “manchas” en distintas regiones que representan las zonas de acoplamiento de dichos hidrógenos.



Todos estos experimentos de NMR constituyen una valiosa herramienta para identificar detalles estructurales de una molécula en disolución.

## 2.6 TECNICAS COMPUTACIONALES

### 2.6.1 Simulaciones de Dinámica Molecular

Las simulaciones de dinámica molecular se han convertido en una herramienta imprescindible para la investigación de las biomoléculas. Nos permiten “dar vida” a las biomoléculas para analizar distintas de sus características, ya sea darnos datos sobre la dinámica natural de biomoléculas en solución, asignar promedios térmicos de algunas de sus propiedades moleculares e incluso explorar que conformaciones de la molécula o complejo son accesibles térmicamente (Hansson 2002; Yu 2004)

Si combinamos estas simulaciones con ensayos reales que restrinjan las posibilidades, como es el caso de estructuras cristalinas de rayos X o espectroscopía de RMN, se conseguirá mejorar y refinar las estructuras obtenidas de forma experimental (Voordijk 2000). En esta tesis, las simulaciones han tenido un gran peso a la hora de obtener nuestros resultados. En todos y cada uno de los artículos, esta técnica es utilizada tanto en simulaciones de dinámica molecular sin restricciones como simulaciones de dinámica molecular con restricciones de tiempo.

Hemos utilizado los datos obtenidos a partir de la resolución de nuestras estructuras cristalinas para restringir las simulaciones y así obtener datos que nos permitan conocer mejor como los péptidos se posicionan e interactúan con la lectina.

Del mismo modo, en las simulaciones con restricción de tiempo se han utilizado datos de los ensayos NOE para realizar limitaciones referentes al péptido. Los ensayos RMN permiten analizar mejor el comportamiento del glicopéptido en la lectina.

## PROTEINA 1SM3 ScFv

Todas las técnicas a continuación que se van a detallar a partir de ahora describen la metodología empleada en los artículos II, III, IV, V y VI, relacionados con la la proteína 1SM3\_ScFv. Las aportaciones propias a los artículos han consistido en todas las técnicas bioquímicas de las construcciones, su expresión y purificación. Además, se han obtenido y resuelto todas las estructuras cristalinas que se publican (cristalografía de rayos X), así como algunos otros ensayos como son ensayos de afinidad (ITCs, espectroscopia de fluorescencia) y análisis de la estructura secundaria (DC) de la proteína y sus ligandos.

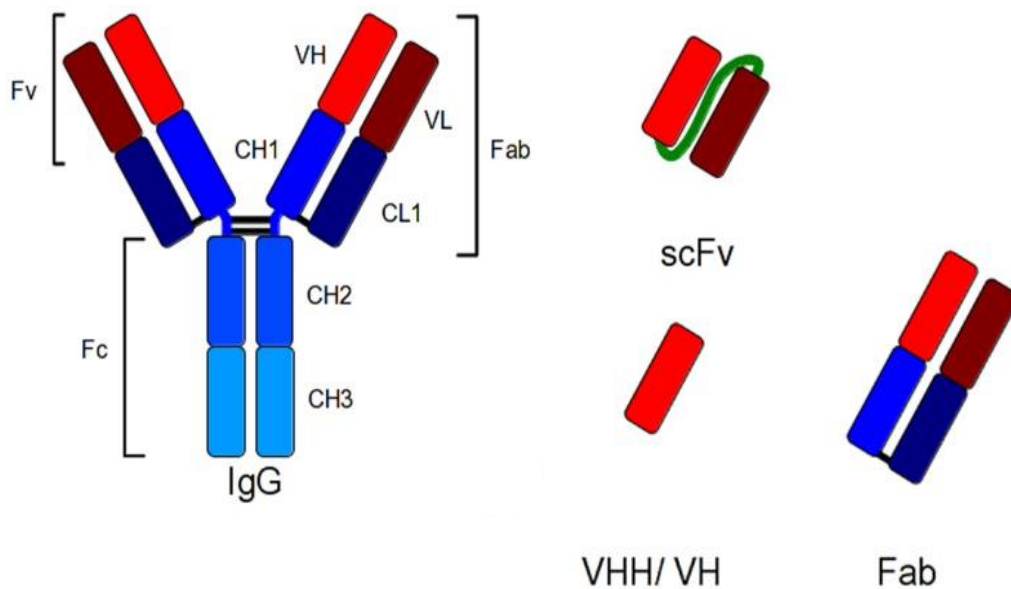
### 2.7 TÉCNICAS BIOQUÍMICAS

#### 2.7.1 Elección del fragmento ScFv para su estudio

La molécula de anticuerpo es modular y se pueden extraer dominios separados por medios bioquímicos o genéticos. Su fragmentación da como resultado características fisicoquímicas alteradas de estas moléculas terapéuticas. Por ejemplo, el tamaño más pequeño de los fragmentos de anticuerpos permite la penetración en los tejidos inaccesibles para los mAbs de tamaño completo (Yokota, 1992; Jain, 1990). La afinidad del anticuerpo también es crítica y, si es demasiado alta, restringirá su capacidad de penetrar un tumor (Adams, 2001). Otra ventaja del uso de estos fragmentos de anticuerpos es que pueden resultar más fáciles y menos costosos de fabricar debido a la falta de glicosilación y tener un tamaño relativamente pequeño, lo que permite su uso de sistemas de expresión procariotas. Sin embargo, los fragmentos carecen del dominio Fc que sirve tanto para estabilizar los anticuerpos de tamaño completo como para permitir el reciclaje mediado por FcR. Como consecuencia, los fragmentos se degradan rápidamente en los humanos y tienen vidas medias circulantes cortas (Cumber, 1992) aunque también hay formas de aprovechar esta circunstancia. En terapias contra el cáncer, a la hora de obtener imágenes de tumores, que la vida media de estos anticuerpos sea prolongada va a provocar un contraste deficiente, dificultando su visualización. Por lo tanto, que la interacción entre el Fc y su receptor FcR sea breve, así como la rápida eliminación de estos anticuerpos radiomarcados por parte del riñón, será favorable en la resolución de las imágenes (Brekke,

2003). Se han desarrollado varias estrategias para extender la vida media de los fragmentos, incluida la conjugación con proteínas como la albúmina (Holt, 2008) y la PEGilación (Jain, 2008), que se aplicó al aprobado por la FDA del anti-TNF $\alpha$  Fab, certolizumab pegol. Sin embargo, las ventajas de biofabricación de la producción de fragmentos pueden perderse si se requiere PEGilación porque el proceso puede resultar costoso y técnicamente desafiante.

Se contemplan tres tipos generales de fragmentos, fragmentos de unión a antígeno (Fab), fragmentos variables de cadena sencilla (ScFv) y "tercera generación" (3G), cada uno de los cuales representa una ola sucesiva de tecnología de fragmentos de anticuerpos.



**Figura 2.16. Esquema de un anticuerpo.** La región ScFv que se ha usado en esta tesis proviene del anticuerpo 1SM3. **Fuente:** Nelson, Antibody fragments, 2010 Jan-Feb; 2(1): 77–83 PMID: 20093855

- Fragmentos de unión a antígeno: los agentes Fab son la clase más antigua de terapias con fragmentos de anticuerpos monoclonales (mAb), demostrada por el hecho de que las ocho terapias con fragmentos que ingresaron al desarrollo clínico antes de 1995 eran Fabs y por lo tanto los más utilizados (Nelson, 2009).

- Fragmentos variables de cadena simple: los fragmentos variables de cadena sencilla (ScFv) son moléculas recombinantes en las que las regiones variables de cadenas de inmunoglobulina ligera y pesada que codifican dominios de unión a antígeno se manipulan en un único polipéptido. En general, las secuencias  $V_H$  y  $V_L$  están unidas por una secuencia enlazadora flexible, y se generan una serie de variantes para optimizar la afinidad y la estabilidad de la unión (Bird 1988). Se empezaron a desarrollar a partir de 1995 como un método innovador y prometedor de terapia frente a cáncer, llegando varios de ellos al desarrollo clínico un tiempo después (Chames, 2009) (Dreier, 2002).

- Moléculas de Tercera Generación ( $V_{HH}$ ): otro enfoque para reducir el tamaño de las proteínas recombinantes derivadas de inmunoglobulina de unión a antígeno ha sido "miniaturizar" los mAbs de tamaño completo mediante la eliminación de dominios considerados no esenciales para la función (Nelson, 2010). Estos se describieron ya en 1988 aunque hasta la actualidad solo un puñado de estos anticuerpos "miniaturizados" han entrado en desarrollo clínico. (Hu, 1996)

A pesar de que los fragmentos Fab son los más comúnmente utilizados, casi tantas terapias ScFv han entrado en desarrollo clínico (19) como con los propios Fabs (23), lo que representa el 40% de los fragmentos evaluados clínicamente (Nelson 2009). Este hecho genera un gran interés por el uso de esta opción de expresión los anticuerpos fragmentados, ya que son más fáciles de expresar que los anticuerpos Fab además de la ventaja del tamaño inferior (Figura 2.16). Por estos motivos, utilizar el fragmento ScFv del anticuerpo 1SM3 fue nuestra elección para esta tesis.

### 2.7.2 Clonajes del fragmento ScFv del anticuerpo 1SM3

Se partió de la construcción del gen *1SM3\_ScFv* sintetizado comercialmente por la empresa GenScript (EEUU) y que consiste exclusivamente en el fragmento Fv del anticuerpo, contenido en el vector pUC57. Posee un sitio de corte para la proteasa de PreScission (PP) y una cola de histidinas que facilite su posterior purificación (Tabla 2.2).

**Tabla 2.2. Secuencia de la proteína 1SM3\_ScFv con los aminoácidos clave señalados.** Cola de histidinas en rojo y corte de la proteasa PreScission en verde

KREAEAEFGG <sup>10</sup>	SHHHHHHHH <sup>20</sup>	HHHSGGSAWS <sup>30</sup>	HPQFEKGGSL <sup>40</sup>	EVLFOGPGSQ <sup>50</sup>	VQLQESGGGL <sup>60</sup>
VQPGGSMKLS <sup>70</sup>	CVASGFTFSN <sup>80</sup>	YWMNWVRQSP <sup>90</sup>	EKGLEWVAEI <sup>100</sup>	RLKSNNYATH <sup>110</sup>	YAESVKGRFT <sup>120</sup>
ISRDDSKSSV <sup>130</sup>	YMQMNNLRAE <sup>140</sup>	DTGIYYCTGV <sup>150</sup>	GQFAYWGQGT <sup>160</sup>	TVTSSSSGG <sup>170</sup>	GGSGGGGGSS <sup>180</sup>
GSSDIVVTQE <sup>190</sup>	SALTTSPGET <sup>200</sup>	VTLTCRSSTG <sup>210</sup>	AVTTSNYANW <sup>220</sup>	VQEKPDHLFT <sup>230</sup>	GLIGGTNNRA <sup>240</sup>
PGVPARFSGS <sup>250</sup>	LIGDKAALTI <sup>260</sup>	TGAQTEDEAI <sup>270</sup>	YFCALWYSNH <sup>280</sup>	WVFGGGTKLT <sup>290</sup>	VLG

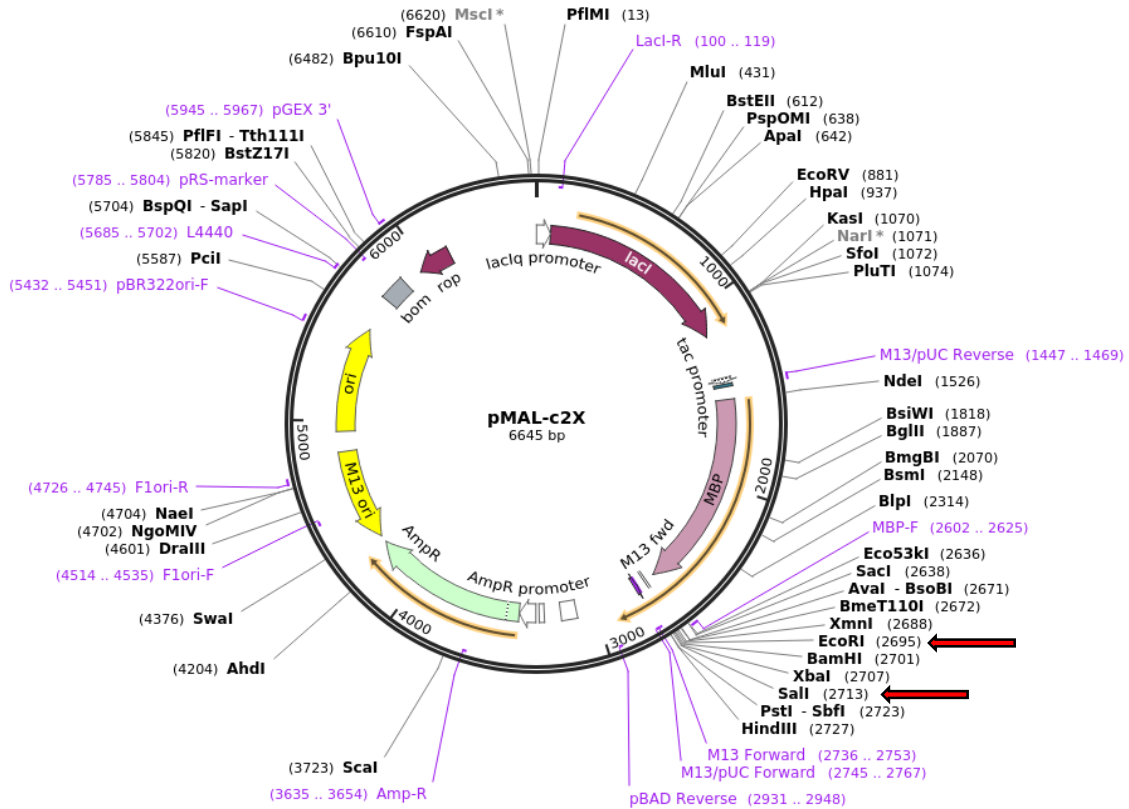
A partir de su secuencia se siguieron tres posibles estrategias en cuanto a su clonaje. Se realizaron dos clonajes distintos para la expresión en *Escherichia coli* en los plásmidos pMALc2x y pET32a, y un clonaje para la expresión en *Pichia pastoris* en el plásmido pPICZaA.

#### 2.7.2.1 Clonaje del fragmento ScFv del anticuerpo 1SM3 en pMALc2x y pET32a

Se utilizaron los plásmidos pMALc2x y pET32a.

El vector pMALc2x tiene un tamaño de 6645 pares de bases, resistencia a ampicilina y posee el gen de la proteína de unión a maltosa, obteniendo así nuestra proteína como proteína de fusión (Figura 2.17).

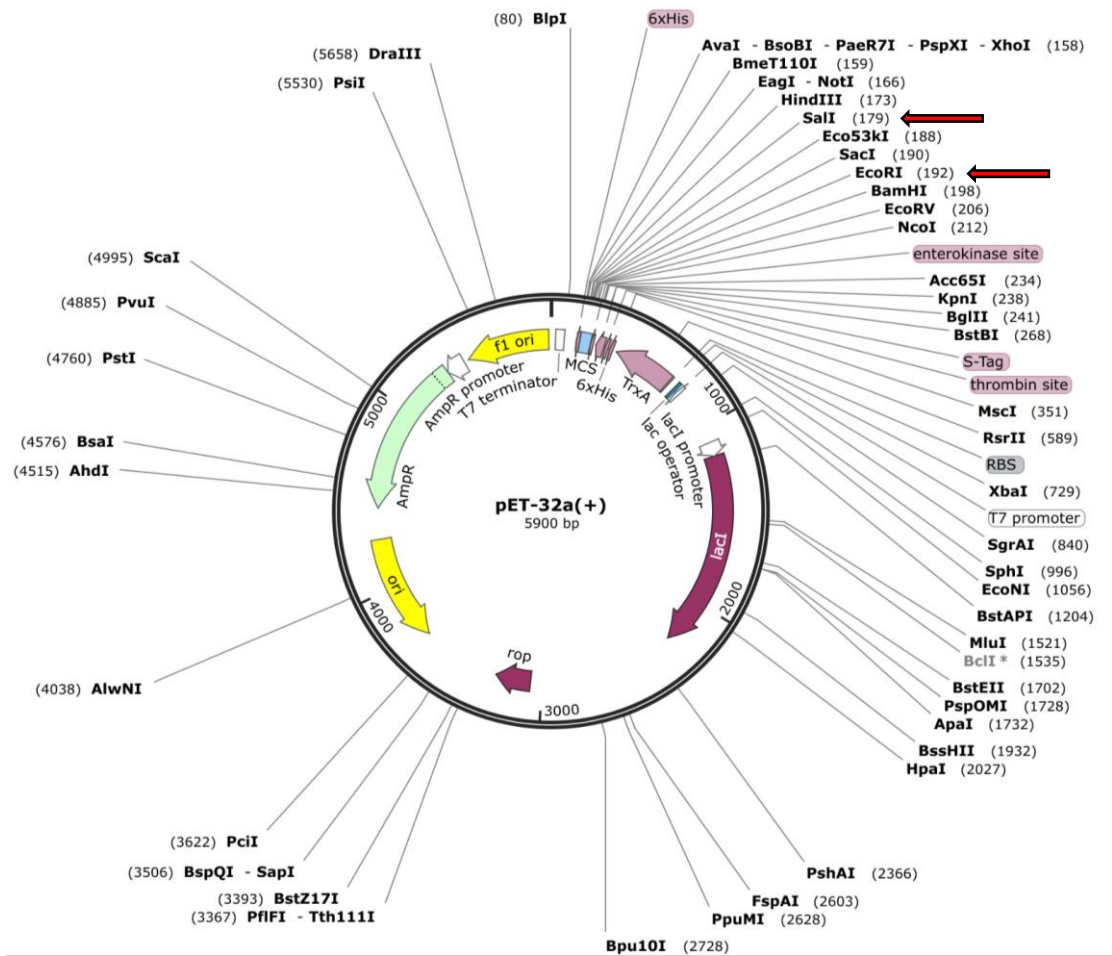
El vector pET32a posee un tamaño de 5900 pares de bases, con gen de resistencia a ampicilina y nos permitirá la obtención de la proteína como proteína de fusión unida a la proteína de tiorredoxina codificada en el gen *trxA* (Figura 2.18).



**Figura 2.17. Mapa del plásmido pMALc2x.** Vector utilizado para el clonaje de nuestra proteína 1SM3\_ScFv. **Fuente:** <https://www.addgene.org/75286/sequences/>

Se digirieron los dos vectores (pUC57-1SM3 y pMALc2x, también pET32a puesto que el corte será el mismo en ambos plásmidos de expresión elegidos) con las enzimas de restricción *EcoR*-I y *Sal*-I. De este modo podemos elegir el sentido en el que se unirá el inserto en el plásmido. En las reacciones de digestión se utilizó el tampón CutSmart (10X, New England Biolabs) y éstas se incubaron a 37 °C durante toda la noche.

Pasado ese tiempo, el volumen total de cada digestión se cargó en un gel al 1% (m/v) de agarosa en tampón TBE (50 mM TRIS pH 8, 50 mM ácido bórico, 1 mM EDTA). Las muestras se adicionaron con tampón de carga 6X (0,25% azul de bromofenol y 30 % glicerol en TE (1 M TRIS pH 8, 0,25 M EDTA pH 8)) y las bandas se tiñeron con SYBR® Safe DNA gel stain (Invitrogen). El marcador de peso molecular utilizado fue el 1 Kb Plus DNA Ladder (Invitrogen). Los geles se corrieron a un voltaje constante de 100 V durante 40 minutos y las bandas se observaron mediante la exposición del gel con luz azul.



**Figura 2.18. Mapa del plásmido pET-32a.** Vector donde se clonó nuestra proteína *1SM3\_ScFv*. **Fuente:** [https://www.snapgene.com/resources/plasmidfiles/?set=pet\\_and\\_duet\\_vectors\\_\(novagen\)&plasmid=pET-32a\(%2B\)](https://www.snapgene.com/resources/plasmidfiles/?set=pet_and_duet_vectors_(novagen)&plasmid=pET-32a(%2B))

A continuación, se extrajeron del gel de agarosa las bandas correspondientes al vector pMALc2x digerido y a la construcción *1SM3\_ScFv* liberada del vector pUC57 mediante el kit comercial Accuprep Gel Purification (Bioneer). Después, se realizó la ligación de ambas extracciones, agregando inserto y vector a la reacción de ligación en proporciones molares aproximadas 3:1. Se empleó utilizando la T4 DNA ligasa (Takara). Como control negativo se realizó la misma reacción de ligación, pero sin añadir el inserto. Las reacciones se incubaron toda la noche en un baño de agua a 23°C.

El producto de ligación se transformó en células competentes de *E. coli* DH5α mediante el siguiente protocolo: se añadieron 33 µl de la ligación diluida en agua miliQ (1:3 v/v), a una alícuota de 150 µL de células. Las células se

incubaron en hielo durante 30 minutos, posteriormente se sometieron a un choque térmico (42 ° C durante 1 minuto) y se incubaron en hielo durante dos minutos. Pasado este tiempo, se agregaron 500 µl de medio SOC (2% triptona, 0.5% extracto de levadura, 0.05% NaCl, 2.5 mM KCl, 10 mM MgSO<sub>4</sub> y 25 mM glucosa) y las células se dejaron recuperar 1 hora a 37 °C y 250 rpm. Después, se tomaron 100 µl de estas células para sembrarlas en placas de LB agar (2% LB, 1,5% Agar) con 100 µg/ml de ampicilina y el resto se centrifugó a 2500 g durante 5 minutos, se retiraron 400 µl de sobrenadante y se resuspendieron las células en el volumen restante, el cual se utilizó para sembrarlas en placas de LB agar con 100 µg/ml de ampicilina. Cabe destacar que la cepa de células de *E. coli* DH5α se utiliza para replicar el plásmido en gran cantidad ya que su aceptación por el ADN es muy buena, se duplican cada 30 min y no poseen resistencia a la ampicilina. Las placas se incubaron a 37 °C toda la noche.

**Tabla 2.3. Componentes para la reacción de PCR en colonia**

<b>Componente</b>	<b>Concentración inicial</b>	<b>Concentración final</b>	<b>Volumen</b>
<b>H<sub>2</sub>O mQ</b>			8 µl
<b>Taq master mix 2X (BIORON) (contiene la polimerasa termoestable, su correspondiente tampón, MgCl<sub>2</sub> y los dNTPs)</b>	2x	1x	10 µl
<b>Cebadores</b>	10 µM c / u	0,2 µM c / u	1 µl c / u
<b>Forward pMAL 2501/ Reverse pMAL 2890</b>			
<b>Volumen Final</b>			20 µl

Para comprobar que las colonias transformadas poseían el inserto del gen 1SM3\_ScFv en sus vectores, se realizó una PCR en colonia tomando como muestra las colonias que mayor tamaño poseían en placa, y lo más aisladas del



resto para evitar colonias satélite (crecen en ausencia de antibiótico alrededor de las colonias con plásmido de resistencia), de menor tamaño. Para cada PCR en colonia se utilizaron los componentes recogidos en la tabla 2.3.

Cebadores pET32a: pET32a tromb R-S FOR /T7 terminator REV

Cebadores pPICZ $\alpha$ A: AOX-1 FOR/REV

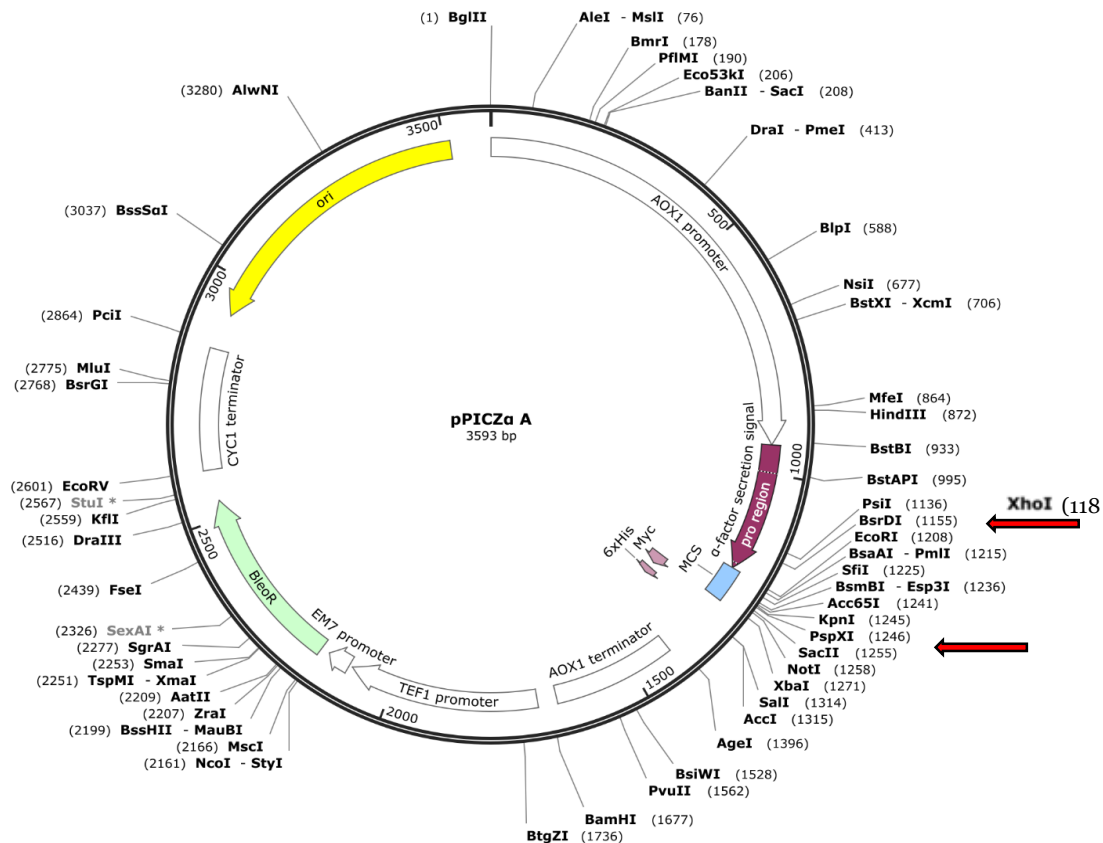
El programa de la PCR en colonia utilizado consiste en:

- Un primer paso de 2 min a 94°C (apertura de horquilla de DNA)
- 30 ciclos de: 20 segundos a 94 °C, seguidos de 30 segundos a 50 °C y 2 minutos a 72 °C (acción de la polimerasa por desnaturalización de la doble hebra, hibridación de los nucleótidos y elongación de la cadena)
- Por último, 1 solo ciclo de 5 minutos a 72 °C (extensión de la cadena)

Los productos de PCR se analizaron mediante separación en un gel de agarosa al 1% y, aquellos vectores que habían incorporado el gen 1SM3\_ScFv, se mandaron a secuenciar a la empresa valenciana Sistemas Genómicos SL (<https://www.sistemasgenomicos.com/>), para comprobar que el clonaje era correcto.

#### 2.7.2.2 Clonaje del fragmento ScFv del anticuerpo 1SM3 en pPICZ $\alpha$ A

El plásmido que se utilizó para la expresión en *P. pastoris* fue pPICZ $\alpha$ A, con un tamaño de 3593 pares de bases, con gen de resistencia a zeocina y la peculiaridad de poseer el promotor de la enzima alcohol oxidasa (AOX1) (**Figura 2.19**).



**Figura 2.19.** Vector pPICZα A donde se clonó el gen *ISM3*\_ScFv para su expresión en *Pichia Pastoris*. **Fuente:** [https://www.snapgene.com/resources/plasmid-files/?set=yeast\\_plasmids&plasmid=pPICZ\(alpha\)\\_A](https://www.snapgene.com/resources/plasmid-files/?set=yeast_plasmids&plasmid=pPICZ(alpha)_A)

Se utilizó el mismo protocolo que para los clonajes dirigidos a la expresión en *E. coli* a excepción de las enzimas de restricción usadas para el corte, que en este caso fueron *Xho*-I/*Sac*-II para digerir tanto plásmido como inserto, señaladas en la imagen, así como placas de LB-“low salt” (LB-LS) agar (1% triptona, 0,5% extracto de levadura, 0,5% NaCl y 1,5% agar) con zeocina (25 µg/ml) a la hora de plaquear las células.

La secuenciación mostró un aminoácido alterado debido al corte realizado por la enzima de restricción *Xho*-I. Se realizó una mutagénesis dirigida para corregir este error usando los cebadores pPICZαA\_CORRECTION\_XHOISITE\_FOR y su reverso (Tabla 2.4), así como el resto de componentes necesarios para llevar a cabo la reparación de la mutación (Tabla 2.5).

**Tabla 2.4. Cebadores utilizados para reparar el aminoácido alterado en el clonaje de *1SM3\_ScFv*.**

<b>pPICZaA_CORRECTION_XHOISIT E_FOR</b>	GCTAAAGAAGAAGGGGTATCTCTCGAGAAGAG AGAAGCAG
<b>pPICZaA_CORRECTION_XHOISIT E_REV</b>	CTGCTTCTCTTTCTCGAGAGATACCCCTTCTTC TTTAGC

**Tabla 2.5. Componentes de la reacción de PCR para las mutagénesis**

<b>Componente</b>	<b>Concentración componente</b>	<b>Concentración final</b>	<b>Volumen</b>
<b>Tampón de carga de la polimerasa (Thermo Scientific)</b>	<b>5x</b>	<b>1x</b>	<b>10 µl</b>
<b>Mezcla de dNTPs</b>	2 mM c / u	0,2 mM c / u	5 µl
<b>DMSO</b>	<b>100 %</b>	<b>3 %</b>	<b>1'5 µl</b>
<b>Cebadores</b>	10 µM c / u	0,4 µM c / u	2 µl cada uno
<b>ADN Polimerasa Phusion Hot Start II High-Fidelity (Thermo Scientific)</b>	<b>2 U / µL</b>	<b>18 mU / µl</b>	<b>0,45 µl</b>
<b>ADN Molde</b>	100 ng / µl	2 ng / µl	1 µl
<b>H2O mQ</b>			<b>28,05 µl</b>
<b>Volumen final</b>			50 µl

El programa utilizado para llevar a cabo la amplificación del ADN consistió en:

- Un primer ciclo de 2 minutos a 98°C.
- 30 ciclos de: 20 segundos a 98°C, 1 minuto a 50°C y, 18 minutos a 72°C
- Un último ciclo de 5 minutos a 72°C.

La verificación de la correcta amplificación se realizó mediante una electroforesis en gel de agarosa al 1 %.

Una vez comprobada la amplificación, se procedió a la transformación con 6 µl de ADN de células competentes de *Escherichia coli* DH5α, y las bacterias transformadas se sembraron en placas de medio LB-LS agar conteniendo 25 µg/ml de zeocina (Invivo Gen), que se incubaron a 37 °C durante una noche. Al día siguiente se escogieron aleatoriamente tres colonias de la placa, cada una de las cuales se estirió en una placa nueva y se empleó como inóculo de un precultivo de 10 ml de LB-LS líquido con la misma cantidad de zeocina, que se dejó incubando a 37 °C y 250 rpm durante una noche. Los precultivos crecidos se centrifugaron al día siguiente durante 5 minutos a 3800 x g y 4 °C (Eppendorf Centrifuge 5810R), se desechó el sobrenadante y se extrajo el ADN plasmídico de las células siguiendo para ello las instrucciones del protocolo del kit comercial Accuprep Plasmid Mini Extraction de Bioneer (Figura 2.20). La elución final de cada muestra se realizó en 65 µl de H<sub>2</sub>O Milli-Q (Millipore) a 65 °C y el ADN resultante se cuantificó mediante absorción a 260 nm en un nanodrop NanoVue Plus (GE Healthcare Life Sciences). De nuevo, se requirió de una comprobación por secuenciación de cada gen mutado, llevada a cabo por la empresa valenciana Sistemas Genómicos S.L. (<https://www.sistemasgenomicos.com/>).

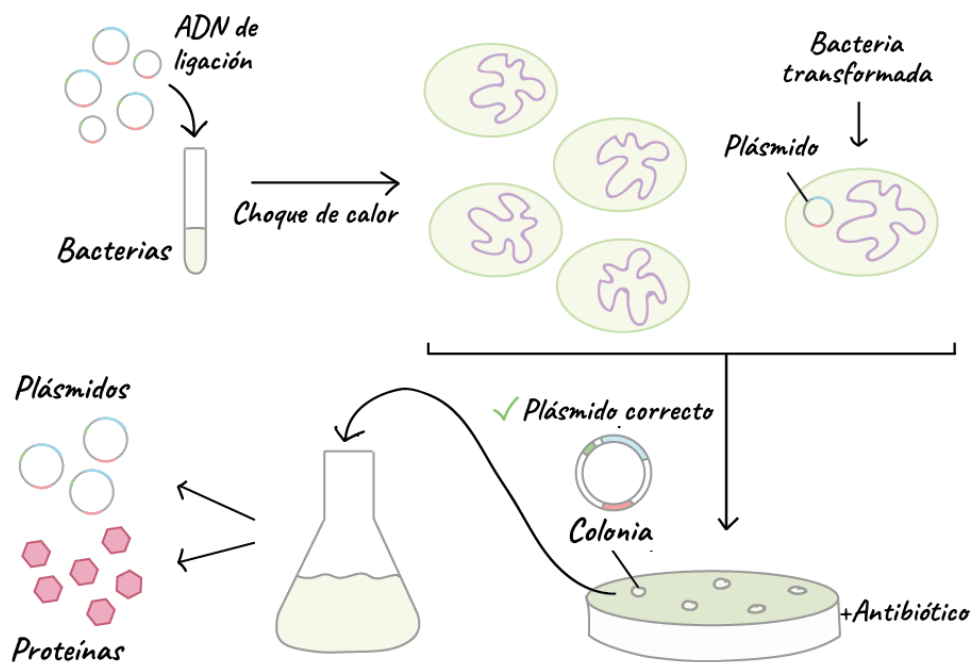


**Figura 2.20. Protocolo de extracción del ADN plasmídico.** Esquema de los pasos que requiere el proceso según el protocolo del kit comercial Accuprep Plasmid Mini Extraction de Bioneer. **Fuente:** <https://bioneer.com.au/Files/Product-Images/Single/Nano Plus Plasmid Mini fo2.aspx>

## 2.8 CULTIVOS DE BIOMASA, EXPRESIÓN Y RECOLECCIÓN DE LA PROTEÍNA 1SM3\_ScFv EN *E. coli*

Para la expresión de la proteína 1SM3\_ScFv con los plásmidos pMALc2x y pET32a, se utilizó la cepa resistente a tetraciclina Rosetta-gami (DE3) de *E. coli*. Esta cepa posee el gen de la ARN polimerasa T7 y su promotor, que posee una expresión inducida por IPTG. Otra de las características por las que se usa es que combina la expresión de siete tARNs raros en *E. coli* con la formación mejorada de puentes disulfuro.

En primer lugar, el clon pMALc2x/pET32a conteniendo la construcción 1SM3-ScFv se transformó en células competentes de Rosetta-Gammi (DE3). El protocolo de transformación fue el mismo que se detalló anteriormente, con la variación de que las células transformadas se sembraron en una placa de medio LB agar adicionado con ampicilina (100 µg/ml) y tetraciclina (12 µg/ml) (Figura 2.21).



**Figura 2.21. Esquema de transformación y selección de colonias para expresión de proteínas.** El ADN que proviene de la ligación del plásmido de interés con nuestro inserto de 1SM3\_ScFv se introduce en las células, que se seleccionan en placa por la resistencia al antibiótico que el plásmido les ha otorgado a las células, para así utilizarlas en la expresión de la proteína. **Fuente:** <https://es.khanacademy.org/science/biology/biotech-dna-technology/dna-cloning-tutorial/a/bacterial-transformation-selection>

A continuación, se creció un cultivo de células *E. coli* transformadas con el clon seleccionado en medio 2xYT (16 g/L Triptona, 10 g/L extracto de levadura, 5g/L NaCl, pH 7) estéril. Para ello se prepararon 50 ml de precultivo por cada litro de cultivo, complementado con el antibiótico determinado (mismos antibióticos y concentraciones que los usados para adicionar anteriormente el medio de la siembra de las colonias transformadas) y, se inoculó con una colonia de *E. coli* transformada. El precultivo se dejó crecer toda la noche a 37 °C, 250 rpm. Después, se añadió la concentración de antibiótico correspondiente al medio de cultivo y se inoculó cada litro de medio con 50 ml de precultivo crecido. Las células se incubaron a 37°C y 250 rpm

hasta que alcanzaron una densidad óptica de  $OD_{600} = 0,6$ . Posteriormente, se indujo la expresión proteica añadiendo 1 mM IPTG al medio e incubando las células a 18°C y 250 rpm.

Pasadas 20 horas de inducción, se centrifugaron las células a 9000 rpm (rotor JA10 con capacidad para 6 frascos de 500 ml de una centrífuga Beckman Avanti J-26 XP), 20 minutos, 4°C. Se desechó el sobrenadante y se recogió el pellet con el tampón A (25 mM TRIS pH 7,5, 10 mM Imidazol, 150 mM NaCl). Por último, se centrifugaron las células a 4000 rpm, 1 h, 4°C, se descartó el sobrenadante y se conservó el pellet.

En el siguiente paso, el pellet se resuspendió en tampón A en proporción 1:1 (m/v) y se mezcló hasta conseguir una consistencia homogénea. Posteriormente, se añadió benzonasa a la mezcla (0,5  $\mu$ l/10 ml de volumen total), lisozima (1 mg/ml de volumen total) y un cóctel de inhibidores de proteasas (1 vial de 500  $\mu$ L /50 ml de volumen total; de modo que el cóctel queda a una concentración final de 1  $\mu$ M PMSF, 10  $\mu$ M benzaminidina, 0,5  $\mu$ M leupeptina). Se incubó la mezcla durante 30 minutos a 37°C. Seguidamente, se sonicaron en hielo las células para romper las membranas celulares y liberar el contenido celular. Se aplicaron 10 pulsos de intensidad 80 %, y de 30 segundos de duración (con 30 segundos de reposo entre pulso y pulso). Después, se centrifugaron las células a 19500 rpm (rotor JA20), 20 minutos, 4°C. Tras la centrifugación se filtró el sobrenadante por un filtro de 0,45  $\mu$ m y se conservó en hielo.

#### 2.8.1 Purificación de la proteína 1SM3\_ScFv

Cada clonaje necesitó un protocolo de purificación diferente debido a las diferencias en las construcciones y a su expresión por parte de los plásmidos que las contienen. La proteína se purificó mediante la técnica de cromatografía líquida (FPLC). Todas las columnas utilizadas estaban acopladas a un sistema ÄKTA purifier (GE Healthcare Life Sciences) (Figura 2.22).



**Figura 2.22.** Sistema ÄKTA purifier (GE Healthcare Life Sciences). Permite un control de presión y seguimiento de la proteína por UV, así como un flujo constante lo que facilita ampliamente la purificación de proteínas respecto al uso de bombas peristálticas o mediante el uso de la gravedad. **Fuente:**<https://www.gelifesciences.com/en/us/solutions/protein-research/products-and-technologies/protein-purification-systems>

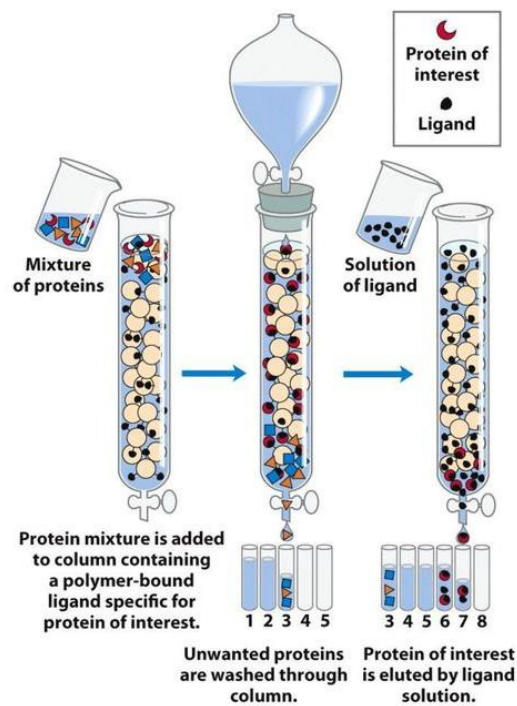
#### 2.8.1.1 Purificación de la proteína 1SM3\_ScFv a partir del clonaje en pMALc2x

La construcción obtenida por el clonaje en pMALc2x requirió un primer paso de purificación por cromatografía de afinidad. Como se ha visto en la secuencia de la proteína, la construcción posee colas de histidinas, por lo que se emplearon columnas de afinidad HisTrap™ HP (GE Healthcare), que contienen iones níquel o cobalto y se unen específicamente a estos aminoácidos. En este caso utilizamos columnas de afinidad que contenían níquel.

La cromatografía de afinidad se basa en la afinidad de unión de las proteínas por la matriz de una determinada columna. Las partículas de dicha pueden tener, como en el caso de las HisTrap, un grupo químico unido covalentemente llamado ligando, y que es el que se une a la macromolécula o



proteína de interés. Cuando se carga una mezcla de proteínas en la columna, aquéllas que tengan afinidad por este ligando se quedarán unidas a las partículas de resina, retardando su migración a través de la columna. A la hora de la elución, se añade un tampón con una alta concentración de sal (que debilita las uniones inespecíficas al aumentar la fuerza iónica de la disolución) o con un ligando más afín que la muestra por la matriz. De esta forma, el ligando libre compite con el ligando unido a las partículas de la columna, liberando la proteína de la matriz; el producto proteico que eluye de la columna se encuentra a menudo unido al ligando usado para la elución. (Lehninger 5<sup>a</sup> Ed. 2009) (Figura 2.23)



**Figura 2.23.** Cromatografía de afinidad. Esquema de como las proteínas se unen a su ligando en la matriz de la columna, para después ser eluidas por gradiente del mismo por competición en la fase líquida. **Fuente:** Lehninger 5<sup>a</sup> Ed. 2009

La muestra de proteínas se cargó en una columna de níquel previamente equilibrada con el tampón A (filtrado por filtro 0.45  $\mu\text{m}$  y desgasificado) y, se pasó el mismo tampón para lavar las proteínas no unidas. La proteína de interés

no se quedó retenida mediante la unión al níquel por sus colas de histidinas, por lo que tampoco se eluyó lavando la columna con el tampón B (25 mM TRIS pH 7,5, 500 mM imidazol, 150 mM NaCl, filtrado por filtro 0,45  $\mu$ m y desgasificado) con un gradiente de 15 minutos hasta alcanzar el 100% del tampón B. La proteína se localizó en el pellet y por tanto estaba en cuerpos de inclusión. Debido a esto no se pudo purificar a partir del sobrenadante y usando la cromatografía indicada porque toda la proteína se expresó como cuerpos de inclusión. La cromatografía se llevó a cabo a un flujo de 2 ml/min y la presencia de proteína se visualizó mediante el método Bradford así como con geles desnaturizantes de acrilamida SDS-PAGE para analizar las distintas fracciones obtenidas.

El mencionado método Bradford no es sino un ensayo colorimétrico en el que el reactivo, que contiene azul brillante de Coomassie, y la proteína presente en la muestra forman un complejo coloreado. Se puede emplear de forma cualitativa o cuantitativa, en lugar de llevar a cabo la cuantificación de proteínas por absorción de UV-visible (280 nm). A diferencia de ésta, permite el uso de agentes reductores, su absorción se realiza a 470 nm (color azul) y tiene un límite de detección para mayores concentraciones (20-2000  $\mu$ g/ml frente a 0.1-100  $\mu$ g/ml en la absorbancia 280nm) lo que hace que resulte más rápido para medidas “in situ”, que no requieren diluciones. (Johnson 2018; <http://www.labome.es/method/Protein-Quantitation.html>)

Se comprobó que la proteína de interés estaba presente en el pellet mediante la electroforesis SDS-PAGE de las fracciones recogidas. Las muestras se mezclaron con Tampón de carga 4X (24 % (v/v) de 1 M TRIS pH 6,8, 8 % (m/v) de SDS, 40 % (v/v) de glicerol, 10 % (m/v) de Azul de bromofenol y 10 % (v/v) de  $\beta$ -mercaptoetanol en agua destilada) y se hirvieron a 100 °C durante 5 minutos. Los geles se corrieron a 120 V durante 90 minutos. Como marcador de peso molecular se utilizó el Blue staining Prestained Marker (Jenna Bioscience). Los geles se visualizaron mediante tinción con azul de Coomassie y se destiñeron con el tampón de destinción (40% MetOH y 10% ácido acético).

Este hecho hizo imposible continuar con la purificación de este clonaje con el plásmido pMALc2x puesto que en la expresión siempre se produjeron estos cuerpos de inclusión.

#### *2.8.1.2 Purificación de la proteína 1SM3\_ScFv a partir del clonaje en pET32*

La muestra proteica se pretendió cortar con la proteasa de PreScission (GE Healthcare Life Sciences), que reconoce un sitio de corte presente en la estructura proteica y libera la cola de histidinas. Se cuantificó la proteína presente en la muestra con Bradford equilibrado con BSA y, se añadió diez veces menos proteasa con respecto a la concentración de proteína. Se incubó la mezcla durante 1-3 días a 4°C y la eficiencia del corte se verificó mediante la electroforesis SDS- PAGE de una muestra de proteína antes y después del corte.

Se utilizó el mismo protocolo que para el clonaje con el plásmido de expresión pMALc2x. En este caso se logró obtener la proteína en la fracción eluida en el gradiente de la columna de níquel. Por lo tanto, se continuó con el protocolo de purificación previsto.

Este paso supuso un problema ya que no se produjo corte, así que se propuso cambiar con un nuevo clonaje el sitio de corte de la proteasa PreScission (PP) al sitio de corte de la proteasa TEV. La proteasa TEV es una proteasa de cisteína altamente específica de la secuencia de Virus del Tabaco (TEV). Es un miembro de las proteasas de tipo quimotripsina. Debido a su alta especificidad de secuencia se utiliza con frecuencia para la escisión controlada de proteínas de fusión in vitro e in vivo. El protocolo del clonaje es el mismo al detallado ya anteriormente en el apartado 2.1.1 con la excepción del uso de los cebadores correspondientes al nuevo sitio de corte (Tabla 2.6).

Se purificó por el mismo método anteriormente descrito el nuevo clonaje con el sitio de corte de la proteasa TEV.

**Tabla 2.6. Cebadores utilizados para cambiar el sitio de corte de PP por el de TEV en el clonaje de 1SM3\_ScFv en pET32a.** El cebador reverso es el mismo que se utilizó como reverso para el clonaje de pPICZαA donde se eliminó la cola de histidinas (Tabla 2.8).

<b>1SM3_EcoRI_T EV_FOR</b>	CGGAATTCGAATTCGGCGGCGAAAACCTGTACTTCCAGGGTAGCGGTAG CCAGGTCCAACCTGCAGGAATCAGGAGG
<b>1SM3_SacII_Sa II_REV</b>	CGGAATTCCTCCGCGGGTCGACTTATCCCAAG

Tras la purificación mediante la columna de níquel, se cambió el tampón de la solución de proteína mediante el uso de la columna HiPrep 26/10 Desalting (GE Healthcare Life Sciences). La proteína se cargó en la columna previamente equilibrada con el tampón TRIS 50 mM pH 8, NaCl 0,5 mM, filtrado por filtro 0,22 μm y desgasificado. La cromatografía se desarrolló a un flujo de 3 ml/min y, la presencia de proteína se visualizó mediante la medida de absorbancia a 280 nm. Tras la elución, se adicionó DTT a una concentración 1 mM para completar las condiciones de corte de la proteasa TEV y se adicionó la proteasa en proporción 1:100 con respecto a la proteína 1SM3\_ScFv para eliminar la cola de histidinas. Se observó un corte parcial de la proteína.

No se pudo recuperar mucha proteína del corte con lo que se abandonó también esta estrategia y se procedió a analizar la construcción en *Pichia pastoris*.

## 2.9 CULTIVOS DE BIOMASA, EXPRESIÓN Y RECOLECCIÓN DE LA PROTEÍNA 1SM3\_ScFv EN P. PASTORIS

### 2.9.1 OBTENCIÓN DE CÉLULAS COMPETENTES DE PICHIA PASTORIS Y TRANSFORMACIÓN POR ELECTROPORACIÓN

Se trabajó con la levadura *P. pastoris* (cepas X33 y SMD1168) como sistema de expresión de proteínas recombinantes ya que *P. pastoris* posee un promotor fuerte y altamente regulable: el promotor de la alcohol oxidasa,

PAOX<sub>1</sub>, inducible por metanol. Además, no secreta una alta cantidad de proteínas endógenas, por consiguiente, las proteínas añadidas son secretadas por el cultivo relativamente puras, lo que facilita su separación (Lin Cereghino *et al.*, 2000).

Para poder introducir los vectores recombinantes en *P. pastoris*, éstos se linealizaron mediante la digestión con la enzima de restricción *Sac*-I. Las condiciones de la reacción se detallan en la tabla 2.7. La reacción se incubó toda la noche a 37°C.

**Tabla 2.7. Linealización mediante la digestión con la enzima de restricción *Sac*-I**

<b>Componente</b>	<b>Concentración inicial</b>	<b>Concentración final</b>	<b>Volumen</b>
<b>ADN</b>	≈ 150 µg/ml	0,1 µg/ml (25 µg totales )	170 µl
<b>Sac-I HF</b>	20000 U/ml	560 U/ml	7 µl
<b>Tampón Cutsmart (10X, New England Biolabs)</b>	10x	1x	25 µl
<b>H<sub>2</sub>O miliQ</b>			48 µl
<b>Volumen final</b>			250 µl

La eficacia del corte se comprobó mediante separación de las muestras en un gel de agarosa al 1% y, una vez linearizados los vectores, la digestión se limpió usando el kit comercial Accuprep Gel Purification (Bioneer).

A continuación, se creció un cultivo de *P. pastoris* en YPD (50 g YPD/ l) a 30 °C y 250 rpm. Después, se tomaron 400 µl del cultivo crecido para inocular 100 ml de YPD y se incubó a 30 °C y 250 rpm hasta alcanzar una densidad

óptica de  $OD_{600}=1-2$ . Las células se recogieron centrifugándolas 5 minutos a 2000 rpm y a 4°C. Se descartó el sobrenadante y se resuspendieron cada  $8 \times 10^8$  células en 8 ml de una solución de 100 mM acetato de litio, 10 mM DTT, 0,6 M sorbitol y 10 mM TRIS-HCl pH 7,5, incubándolas durante 30 minutos a temperatura ambiente. El número de células se calculó de acuerdo a la fórmula  $OD_{600}= 3 \times 10^7$  células. Después, las células se centrifugaron de nuevo 5 minutos a 2000 rpm y 4°C y el pellet se resuspendió en 1,5 ml de 1 M sorbitol. Esta última centrifugación y el lavado con 1 M sorbitol se repitió 3 veces. Finalmente, se resuspendieron en 1 ml de 1 M sorbitol por cada  $10^{10}$  células.

Las células de *P. pastoris* se transformaron por electroporación. Se mezclaron 80 µl de células con aproximadamente 5 µg de plásmido linealizado, se transfirió la mezcla a una cubeta estéril y se incubó en hielo durante 5 minutos. Después se aplicó un pulso eléctrico de 2000 V a las células (electroporador de BioRad) e inmediatamente después se resuspendieron en 1 ml de sorbitol 1 M estéril, mantenido en hielo. Como control negativo se electroporó una muestra de 80 µl de células sin añadirles plásmido.

Las muestras se incubaron a 30°C durante 2 horas sin agitación y después se tomaron 100 µl de cada muestra y se sembraron en placas de YPDS (50 g YPD/ l, 20 g agar/ l, 1 M sorbitol) adicionadas con zeocina (a una concentración de 200 µg/ ml para la cepa SMD1168 y de 100 µg/ml para la cepa X33). Se centrifugó el volumen restante de suspensión celular a 4000 rpm durante 10 minutos y a 4°C, se retiró aproximadamente 700 µl del sobrenadante, se resuspendieron las células y se plaquearon en placas de YPDS con zeocina. Todas las placas se incubaron a 30°C durante 3-4 días hasta que se observó la aparición de colonias.

#### 2.9.2 PRUEBA DE EXPRESIÓN DE COLONIAS POSITIVAS DE LA ELECTROPORACIÓN

Las colonias transformadas se seleccionaron y se inocularon en 1 ml de medio BMGY (1% extracto de levadura, 2% peptona, 100 mM  $KH_2PO_4$  pH6, 1,34% YNB, 1% biotina, 1% glicerol) y se incubaron toda la noche a 30°C, 220 rpm. Después, las células se centrifugaron a 4000 rpm, 10 minutos, 4°C, se descartó el sobrenadante y se resuspendieron en 1 ml de medio BMMY (1%

extracto de levadura, 2% peptona, 100 mM KH<sub>2</sub>PO<sub>4</sub> pH 6, 1,34% YNB, 1% biotina, 1% metanol).

Las células se incubaron a 18°C, 220 rpm y, a las 24 y 48 horas de incubación se les añadió metanol (a una concentración final del 1%). Finalmente, a las 72 horas de incubación, los cultivos se centrifugaron a 4000 rpm, 10 min, 4 °C, y se analizó la expresión proteica presente en el sobrenadante mediante electroforesis SDS-PAGE al 10%.

Gracias a esta prueba se observó que era mayor la expresión en la cepa X33 frente a la SMD1168, por lo que se seleccionó como cepa de trabajo.

Del mismo modo se observó que el primer clonaje en pPICZαA, que contenía colas de His al igual que ocurría para los clonajes de *E. coli* puesto que era la misma construcción, se expresaba a bajos niveles, por lo que para optimizar la expresión se realizó una mutagénesis dirigida, con el mismo protocolo que en pET32a-1SM3 con sitio TEV, para eliminar dicha cola de His ya que limita la expresión en *P. pastoris*.

Los cebadores fueron los siguientes (Tabla 2.8):

**Tabla 2.8. Cebadores utilizados para el nuevo clonaje de pPICZαA para eliminar la cola de histidinas**

<b>1SM3_Xho-I_FOR</b>	CGGAATTCCTCGAGAAGAGAGAAGCAGAAGCACAGGTCCA ACTGCAGGAATCAGGAGG
<b>1SM3_Sac-II_Sal-I_REV</b>	CGGAATCCC GCGGGTCGACTTATCCCAAG

Por último, se realizó el mismo protocolo de expresión con otro medio distinto, para evitarnos los problemas que se presentaban en las purificaciones debido a los compuestos del medio de expresión habitualmente usado. Este medio fue BMM o buffered minimal metanol/tampón mínimo de metanol (100 mM KH<sub>2</sub>PO<sub>4</sub> pH6, 1,34% YNB, 1% biotina, 1% metanol) (Sreekrishna 2010).

### 2.9.3 Cultivo de biomasa, expresión y recolección de la proteína 1SM3\_ScFv

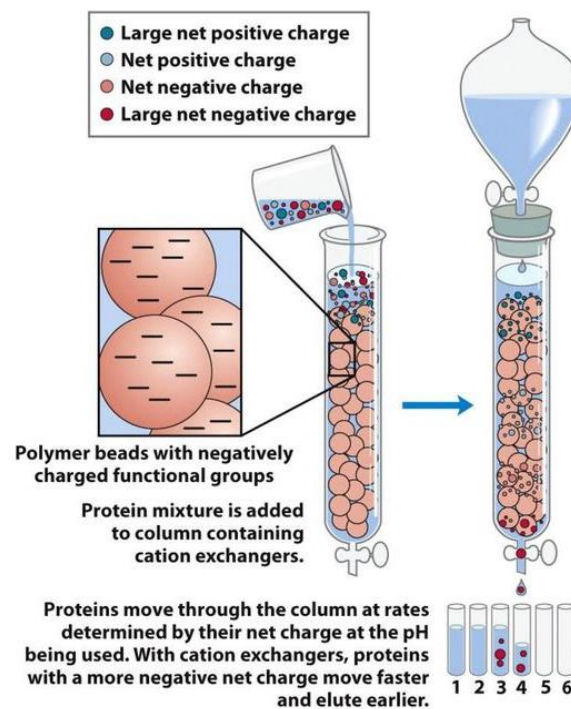
Se creció un cultivo de *P. pastoris* transformada con el plásmido deseado en medio BMGY a 30°C, 250 rpm durante 24 horas. Las células se centrifugaron a 6000 rpm, 15 minutos, 10°C (rotor JA10) y el pellet se resuspendió en un volumen, igual al crecido, de medio BMM en esterilidad. El cultivo se incubó entonces a 18°C, 250 rpm y, a las 24 y 48 horas de incubación se le añadió metanol (a una concentración final del 1 %). Finalmente, a las 72 horas de incubación, el medio de cultivo se centrifugó a 9000 rpm, 30 minutos, 4°C (rotor JA10). Se recogió el sobrenadante y se filtró por un filtro de 0,45 µm y de 0,22 µm. A continuación, se concentró hasta llevarlo a un volumen final de 50 ml mediante el filtro Pellicon XL 10k (Millipore) y, finalmente se dializó en el tampón (25 mM TRIS pH 7,5).

#### 2.9.3.1 Purificación de la proteína 1SM3\_ScFv del clonaje en pPICZαA

Esta construcción, al no poseer colas de histidinas requirió un protocolo diferente a las anteriores. Para su purificación se utilizó la metodología de cromatografía de intercambio iónico, en particular el intercambio aniónico. A través de esta técnica se aprovecha las diferencias en el signo y la magnitud de las cargas eléctricas netas de las proteínas a un pH determinado. La matriz de la columna es un polímero sintético (resina) que tiene unidos grupos cargados; los que tienen unidos grupos aniónicos, se denominan intercambiadores catiónicos, y los que tienen unidos grupos catiónicos se denominan intercambiadores aniónicos. La afinidad de cada proteína por los grupos cargados de la columna está afectada por el pH (que determina el estado de ionización de la molécula) y la concentración de los iones salinos libres competidores presentes en la disolución que las rodea. La separación puede optimizarse cambiando gradualmente el pH y/o la concentración de sal de la fase móvil, de forma que se cree un gradiente de pH o sal. (Lehninger 5<sup>a</sup> Ed. 2009) (Figura 2.24)



La proteína se cargó en la columna aniónica (GE Healthcare Life Sciences) acoplada a un sistema ÄKTA purifier previamente equilibrada con el tampón 25 mM TRIS pH 7,5 filtrado con un filtro de 0,45  $\mu\text{m}$  y, se lavó la columna con el mismo tampón. El tampón de elución fue el mismo salvo con adición de 1 M NaCl. La cromatografía se desarrolló a un flujo de 5 ml/min con un gradiente al 100% de tampón B de 10 min y, la presencia de proteína se visualizó mediante la medida de absorbancia a 280 nm.



**Figura 2.24 Cromatografía de intercambio iónico.** Esquema en el que se muestra como los componentes de una muestra de proteínas interaccionan de distinta forma con la matriz de la columna debido a las cargas ocasionadas por el pH de la disolución. **Fuente:** Lehninger 5<sup>a</sup> Ed

Las fracciones de proteína se concentraron por ultrafiltración en concentradores centricon con un corte de poro de 10 kDa (Amicon Ultra-15 centrifugal Filter, Merck Millipore), hasta alcanzar un volumen de 2,5 ml. No fue necesario un cambio de tampón ya que la muestra se iba a diluir en el propio tampón de la nueva columna. Se cargó la muestra en la columna de exclusión

molecular (HiLoad 16/600 Superdex 75 pg, GE Healthcare Life Sciences) previamente equilibrada con el tampón de la cromatografía (25 mM TRIS pH 7,5, 150 mM NaCl). La cromatografía se desarrolló a un flujo de 2 ml/min y la presencia de proteína se visualizó mediante la medida de absorbancia a 280 nm. Se comprobó la presencia de la proteína de interés mediante la electroforesis SDS-PAGE de las fracciones recogidas.

Se juntaron las fracciones de proteína y se introdujeron en una membrana de diálisis. La membrana se introdujo después en un vaso que contenía 2 l de tampón de diálisis (25 mM TRIS pH 7,5) y, con agitación a 4 °C durante 2 horas.

Se concentró la muestra de proteína de nuevo en los mismos concentradores centrífugos hasta alcanzar el mínimo volumen. Finalmente, se cuantificó la proteína por su absorbancia a 280 nm y su coeficiente de extinción molar y, se conservó a 4°C en hielo.

Para verificar que se trataba de nuestra construcción, se realizó un gel SDS-PAGE acrilamida 10 % el cual sirvió para resolver la banda de nuestra proteína y recortarla, para enviarla a procesar por espectrometría de masas.

Del mismo modo, se realizaron pruebas tanto para asegurar la integridad de nuestra proteína evaluando su estructura secundaria, como el uso de la espectroscopía de dicroísmo circular (DC), y otras técnicas para calcular desde las constantes de afinidad de la proteína 1SM3\_ScFv con los péptidos problema (ITC) como la determinación de su estructura tridimensional con los mismos (cristalografía de rayos X).

## **2.10 TECNICAS BIOFISICAS PARA EL ESTUDIO DE INTERACCIÓN**

### **2.10.1 Calorimetría de Titulación Isotérmica (ITC)**

Como ya hemos explicado en el apartado 2.2.1 de la lectina, para 1SM3 también se realizaron medidas de ITC con el fin de determinar su interacción con los glicopéptidos. El problema de esta técnica es el rango en el que se efectúan. Para analizar la interacción de anticuerpos con sus sustratos, cuya

constante de afinidad es tan baja (del rango nM o incluso pM) hay que realizar otras técnicas para contrastar su veracidad.

Debido a los resultados ambiguos y que la proteína poseía tres aminoácidos aromáticos en su sitio de unión, se planteó la posibilidad de obtener estas afinidades por medio de las técnicas que se detallan a continuación en este apartado.

#### 2.10.2 Interferometria de Biocapa (BLI)

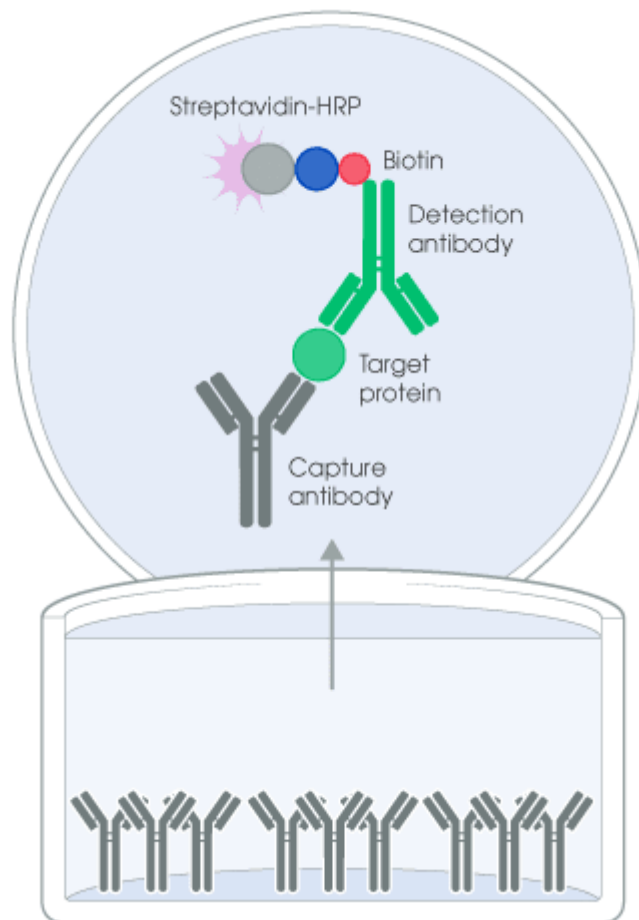
La interferometria de biocapa es una técnica óptica rápida y sencilla para el estudio de las interacciones entre proteínas, péptidos, ligandos de pequeño tamaño, lípidos, etc, a tiempo real. La molécula de interés se inmoviliza en un biosensor de fibra óptica mientras que el ligando se mantienen en solución. Requiere una concentración de muestra muy baja, del orden nanomolar, suponiendo una ventaja frente a otras técnicas como ITC y alternativa a la técnica de SPR. (Kussrow 2012)

#### 2.10.3 Ensayo ELISA

El ensayo inmunoabsorbente ligado a enzimas (ELISA) es un método de captura de antígeno (o anticuerpo) diana en muestras usando un anticuerpo específico (o antígeno), y de detección/cuantificación de molécula diana usando una reacción enzimática con su sustrato.

En ELISA, varios antígenos se usan combinaciones de anticuerpos, siempre incluyendo un antígeno o anticuerpo marcado con una enzima, y la actividad enzimática se mide colorimétricamente (Figura 2.25). La actividad enzimática se mide usando un sustrato que cambia de color cuando la enzima lo modifica. La absorción de luz del producto formado después de la adición del sustrato se mide y se convierte en valores numéricos.

Esta técnica se ha aplicación tras obtención de muestras de suero en el artículo IV para determinar los isotipos que reconoce el anticuerpo.



**Figura 2.25 Esquema del ensayo ELISA.** Se puede observar como dentro de cada pocillo se encuentran anticuerpos unidos a la base, que reaccionaran con la proteína diana y esta unión se analiza con un segundo anticuerpo específico unido a la enzima HRP que producirá fluorescencia en su reacción, permitiendo identificar indirectamente que muestras contienen la proteína diana. **Fuente:** <https://www.abcam.com/kits/elisa-principle>

#### 2.10.4 Espectroscopía de Fluorescencia

La espectroscopia de fluorescencia es un tipo de espectroscopía electromagnética que analiza la fluorescencia de una muestra. Se trata de utilizar un haz de luz, normalmente luz ultravioleta, que excita los electrones en las moléculas de ciertos compuestos y hace que emita luz; típicamente, pero no necesariamente, la luz visible. En proteínas será necesario la presencia de aminoácidos aromáticos para que se produzca la emisión de la luz emitida. Una técnica complementaria es la espectroscopia de absorción.

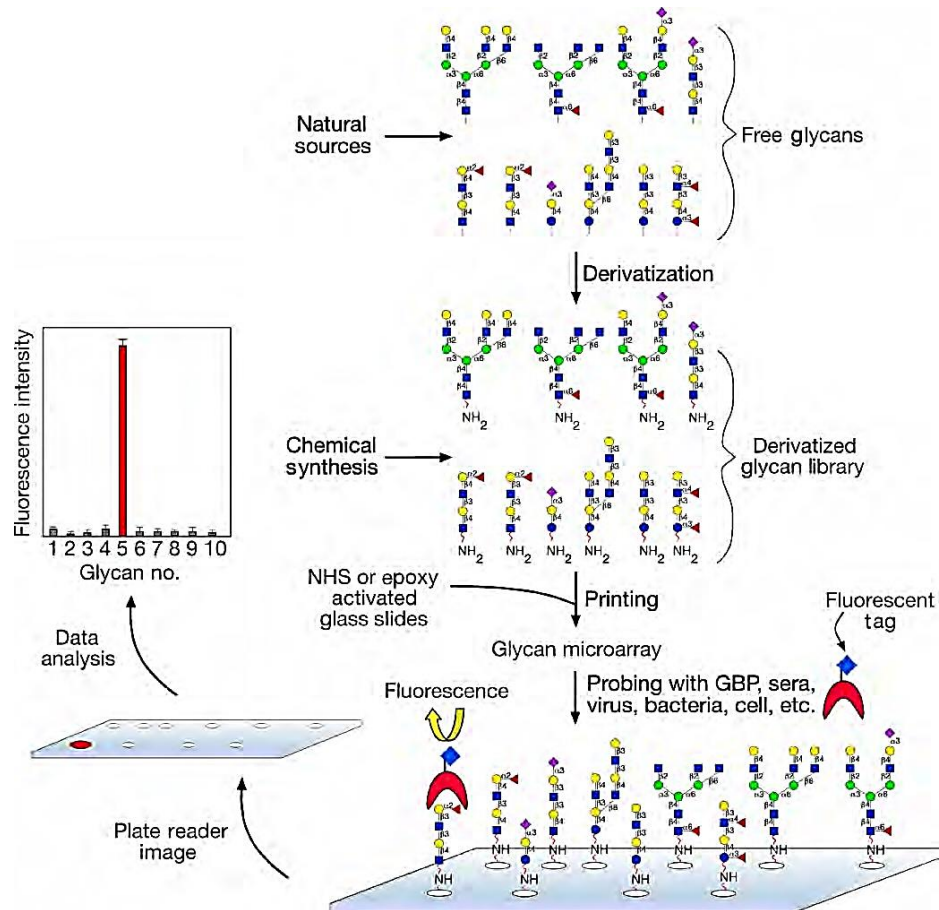
Se realizaron estudios de unión plegamiento térmico en un Espectrofotómetro de fluorescencia Cary Eclipse (Varian) usando una cubeta de cuarzo de longitud de la trayectoria de 1 cm (Hellma SUPRASIL). La temperatura se controló por una unidad Peltier y medida por medio de una sonda de temperatura. Las condiciones de trabajo tras el calibrado fueron las siguientes: 1  $\mu$ M proteína 1SM3\_ScFv en tampón 25 mM TRIS pH 7,5 filtrado por un filtro de 0,45  $\mu$ m.

En primer lugar, se obtuvo el espectro de emisión del blanco en las siguientes condiciones; 25 °C desde 300 hasta 450 nm. La muestra se excitó a 280 nm utilizando un ancho de banda de 5 nm. La velocidad de exploración fue de 120 nm / min y la fluorescencia se recogió por un fotomultiplicador puesta a 700 V. El blanco consistió en 1 ml del tampón al cual se adicionó 0,9  $\mu$ l para obtener la concentración deseada y se procedió a la medida de la fluorescencia para conocer el valor del blanco. A continuación, a esta disolución se adicionó 0,5  $\mu$ l del péptido sucesivamente a concentración conocida para observar las variaciones de fluorescencia en cada medida y así poder analizar la interacción péptido-proteína mediante la pendiente de la curva obtenida al tratar los datos normalizados y el valor de la constante de disociación.

#### 2.10.5 Microarrays

Los “microarrays” (o microchips) de proteínas son un método de detección útil para lisados de células enteras, proteomas fraccionados, glicoproteínas intactas y reacciones antígeno-anticuerpo (Kricka y Master, 2008; Tao, Chen y Zhu, 2007) (Figura 2.26). A diferencia de los oligonucleótidos, las proteínas son ampliamente heterogéneas en tamaño, forma y química. Para maximizar la calidad de los datos obtenidos, es fundamental elegir la superficie con la mejor relación señal / ruido. Sprenger *et al.* desarrollaron superficies fotoactivables recubiertas con una película delgada (portaobjetos de vidrio recubiertos con dextrano) adecuadas para la inmovilización covalente de glicanos, glicoconjugados y lectinas en formatos de microarrays (Angeloni et al., 2005). Los “microarrays” se ha convertido en una herramienta esencial para caracterizar la estructura del glicano y estudiar las interacciones biológicas relacionadas con la glicosilación, mediante el uso de

sondas como un medio para consultar las moléculas glicosiladas marcadas o capturadas en las matrices. Además, el alto rendimiento y la gran reproducibilidad de los “microarrays” se ha destacado por sus amplias aplicaciones para la validación de biomarcadores, donde un gran número de muestras deben analizarse varias veces (Sutandy 2013).



**Figura 2.26.** Ensayo con microchips. Esquema de cómo se realiza la interacción entre los glicanos anclados frente a las muestras a las que son expuestas. **Fuente:** <https://www.creative-proteomics.com/services/glycan-microarray-assay.htm>

## 2.11 TÉCNICAS BIOFÍSICAS DE ANÁLISIS ESTRUCTURAL

### 2.11.1 Cristalización 1SM3\_ScFv y co-cristalización con sus péptidos

La metodología que se usó fue la misma que para la lectina, difusión de vapor con el método de gota sentada. Se usaron diversos cribados comerciales; PEGSalt, JCSG y CLASSIC I (Molecular Dimensions) fueron en los que mejores resultados se obtuvieron.

Se siguió un protocolo idéntico al detallado en el apartado de cristalización de la lectina, con la excepción de usar una concentración de 15 mg/ml para la proteína 1SM3\_ScFv y los péptidos en una concentración de 10 mM.

### 2.11.2 Experimentos RMN

Como ya se ha detallado en el apartado de la lectina, se utilizaron numerosos experimentos de RMN para la determinación estructural de los péptidos y glicopéptidos. Adicionalmente se emplearon nuevos métodos necesarios para resolver cuestiones que con los anteriormente descritos no se podían obtener y son de gran interés.

#### 2.11.2.1 Experimentos 2D ROESY

El experimento ROESY (rotating-frame NOE spectroscopy) permite aplicar lo que los experimentos NOESY no pueden a moléculas grandes. En moléculas pequeñas ambos se comportan de igual forma. (Farmer *et al.* 1969)

### 2.11.3 Dicroísmo Circular (DC)

El dicroísmo circular es una técnica de espectroscopía de absorción electrónica, basada en el cambio de configuración electrónica molecular de un estado fundamental a un estado excitado, debido a la absorción de radiación electromagnética polarizada. La luz polarizada en un plano se puede descomponer en dos componentes polarizados circularmente en sentidos opuestos. Ambas componentes están en fase y son de la misma amplitud. Al pasar por un medio ópticamente activo, cada componente puede interactuar de manera diferente con los centros quirales de las moléculas presentes. Si los índices de refracción del medio son distintos para cada componente, se produce

un desfase entre ellos que provoca la rotación del plano de polarización en un ángulo  $\alpha$ , dándose un proceso de dispersión que se conoce como dispersión óptica rotatoria (DOR). Si los componentes polarizados circularmente son absorbidos en diferentes grados por la sustancia, la luz que pasa a través de la misma estará polarizada elípticamente y se dice que la sustancia tiene dicroísmo circular (DC), magnitud expresada en unidades de elipticidad molecular ( $\theta$ ). Tanto la dispersión óptica rotatoria como el dicroísmo circular varían con la longitud de onda, pudiéndose representar la rotación o elipticidad frente a la misma, obteniéndose espectros de estos fenómenos. Los espectros de dicroísmo circular se obtienen generalmente en las regiones del ultravioleta cercano (250 a 350 nm) y lejano (180 a 250 nm) de la radiación electromagnética.



**Figura 2.27** Espectrómetro de Chirascan termostatizado (Applied Fotofísica). **Fuente:** <https://www.aimil.com/products/chirascan-cd-spectrometer>

Se determinó la integridad de la proteína mediante el estudio de su correcto plegamiento en un espectrómetro de Chirascan termostatizado (Applied Fotofísica). El espectro se registró con un ancho de banda de 1 nm y resolución espectral de 0,5 nm usando una cubeta de cuarzo de longitud de trayectoria de 0,1 cm (Hellma SUPRASIL). El ensayo se realizó en el rango UV lejano (190-260 nm) con un tiempo de respuesta de 6 s. Para los componentes ópticos de seguridad y oxígeno calidad de escaneado se purgó del espectrómetro



a través de un flujo de nitrógeno (2 bar de presión, suministrado 20 minutos antes de que la lámpara se encienda y se mantiene durante todo el experimento). Además, para evitar interrupciones del experimento debido a fallos en la fuente de alimentación, el equipo se conecta a una fuente de alimentación independiente (Sistemas UPS Vision). La medida se llevó a cabo a 25 ° C, y dicha temperatura se controló dentro del soporte de la celda del espectrómetro mediante un controlador de temperatura Peltier acoplado a un baño térmico.

Se utilizó el software Chirscan Pro-Data para establecer las condiciones experimentales y para monitorear la evolución espectros.

El blanco consistió de un tampón 25 mM TRIS pH 7,5 y a continuación se midió 300 µl de disolución *1SM3\_ScFv* en una concentración 20 µM usando el mismo tampón. Se realizó con la intención de verificar un estado correcto de nuestra proteína, que ésta se encontraba en buen estado de plegamiento/conformación y de funcionalidad. Posteriormente también se realizaron medidas de los glicopéptidos utilizados en los artículos.

#### 2.11.4 IRID

La espectroscopía por infrarrojos por inmersión resonante (RIDIR) se ha empleado junto con los cálculos DFT (consultar en técnicas computacionales, 2.12.1) de la teoría funcional de densidad para caracterizar las topologías de enlace de hidrógeno. Los espectros infrarrojos IR de cada especie se pueden registrar sin interferencias entre sí utilizando la espectroscopía RIDIR. Debido a que el grupo OH vibra directamente contra el enlace H, las frecuencias vibratorias y las intensidades IR de los fundamentos de estiramiento OH son funciones sensibles del número, tipo y fuerza de los enlaces H en los que participa cada grupo OH. (Nakamoto 2008; Pretsch 1980)

## 2.12 TECNICAS COMPUTACIONALES

### 2.12.1 Cálculos DFT

La teoría del funcional de la densidad (DFT) es una teoría desarrollada para el cálculo de sistemas de muchos cuerpos, especialmente de grandes números de electrones y átomos. Es una teoría muy importante en el campo de la física molecular, pues permite calcular las interacciones electrónicas y por ello mismo las interacciones entre las moléculas. Esta teoría nació en 1927, propuesta por Thomas y Fermi y extendida posteriormente por Dirac. Pero no fue hasta los años sesenta hasta que la teoría fue reformulada para ser entendida como lo hacemos ahora gracias a los teoremas de Hohenberg y Kohn (Martin 2004). Las ecuaciones de la teoría DFT son mucho más simples de resolver que las ecuaciones de muchos cuerpos de mecánica cuántica u otras aproximaciones, por lo que permiten tratar sistemas más grandes y calcular más propiedades. Por lo general es posible llegar a hacer simulaciones con unos pocos miles de átomos (Sholl 2009).

### 2.12.2 Simulaciones de Dinámica Molecular

Como ya se ha descrito en el apartado 2.6.1 de la lectina, las simulaciones DM combinadas con los datos experimentales obtenidos en ensayos de análisis estructural han permitido mejorar y refinar las estructuras obtenidas de forma experimental (Voordijk 2000).

## 2.13 ENSAYOS EN CÉLULAS Y PRUEBAS EN PACIENTES

Los artículos IV y VI recogen los avances realizados a nivel clínico con el uso de nuestro anticuerpo. 1SM3-ScFv ha sido probado tanto en muestras de suero humano, obtenido del Biobanco-iMM, (Lisbon Academic Medical Center) en Lisboa (artículo IV), como realizando ensayos con líneas celulares de cáncer y, también, en células de cáncer de pacientes con cáncer de mama (artículo VI).

## **ARTICULOS**

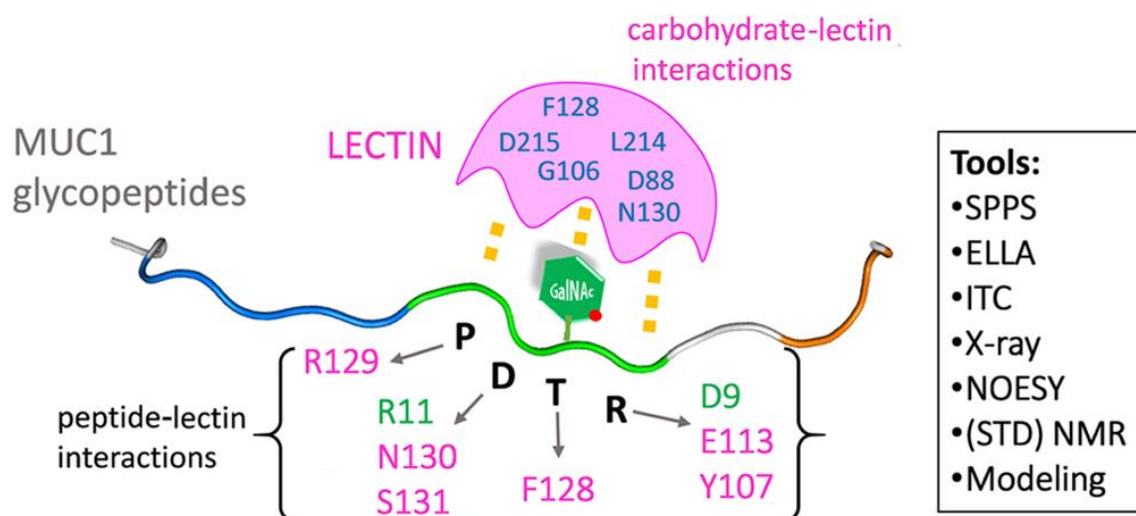
Article I: Detection of Tumor-Associated Glycopeptides by Lectins:  
The Peptide Context Modulates Carbohydrate Recognition

David Madariaga, Nuria Martínez-Sáez, Víctor J. Somovilla, Helena Coelho,<sup>†</sup> Jessika Valero-González,<sup>‡</sup> Jorge Castro-López,<sup>‡</sup> Juan L. Asensio,<sup>§</sup> Jesús Jiménez-Barbero,<sup>||</sup> Jesús H. Busto, Alberto Avenoza, Filipa Marcelo,<sup>†</sup> Ramón Hurtado-Guerrero,<sup>‡</sup> Francisco Corzana,<sup>\*</sup> and Jesús M. Peregrina<sup>\*</sup>

Centro de Investigación en Síntesis Química, Departamento de Química,  
Universidad de La Rioja, E-26006 Logroño, Spain

### 3.4. ABSTRACT

Tn antigen ( $\alpha$ -O-GalNAc-Ser/Thr) is a convenient cancer biomarker that is recognized by antibodies and lectins. This work yields remarkable results for two plant lectins in terms of epitope recognition and reveals that these receptors show higher affinity for Tn antigen when it is incorporated in the Pro-Asp-Thr-Arg (PDTR) peptide region of mucin MUC1. In contrast, a significant affinity loss is observed when Tn antigen is located in the Ala-His-Gly-Val- Thr-Ser-Ala (AHGVTSA) or Ala-Pro-Gly-Ser-Thr-Ala-Pro (APGSTAP) fragments. Our data indicate that the charged residues, Arg and Asp, present in the PDTR sequence establish noteworthy fundamental interactions with the lectin surface as well as fix the conformation of the peptide backbone, favoring the presentation of the sugar moiety toward the lectin. These results may help to better understand glycopeptide – lectin interactions and may contribute to engineer new binding sites, allowing novel glycosensors for Tn antigen detection to be designed.



Lectins are a class of carbohydrate-binding proteins that trigger several important cellular processes.<sup>1</sup> They have been successfully employed to recognize malignant tumors, and a few are indicated for the reduction of treatment-associated side effects as adjuvant agents in chemotherapy and radio-therapy. It has been observed that certain lectins present antitumor activity and can directly kill human cancerous cells.<sup>2,3</sup> These features make lectins a current niche of research that may be adopted as an alternative cancer therapy.<sup>4-6</sup>

Within the context of cancer, Tn antigen ( $\alpha$ -O-GalNAc-Ser/ Thr) is one of the most specific human tumor-associated structures.<sup>7</sup> Expression of this motif occurs early in tumor cells. In fact, there is a direct correlation between carcinoma aggressiveness and density of this antigen. In general, Tn antigen is presented in cancer cells as a part of modified glycoproteins, such as mucins, and in particular mucin MUC1.<sup>8-12</sup> Therefore, MUC1 derivatives have become attractive molecules for the treatment of cancer.<sup>8,13-17</sup> Logically, the premature detection of Tn antigen is essential to properly treat and eradicate tumors. In this regard, a variety of lectins that bind to glycopeptides bearing Tn motifs have been successfully employed as biomarkers<sup>18,19</sup> of this antigen. From a molecular recognition viewpoint, it is well-documented that lectins recognize the sugar moiety as a free form or linked to peptides.<sup>1</sup> Although there is evidence that the natural residue example of this concept in the context of natural MUC1 glycopeptides carrying the Tn antigen. The extracellular domain of MUC1 consists of tandem repeats comprising 20 amino acids of the sequence HGVTSAPDTRPAPGSTAPPA with five possible O-glycosylation sites and three relevant regions of MUC1 glycoproteins: the GVTSA sequence, present in compound 1 (Figure 1), is an effective substrate for GalNAc transferases;<sup>11,23</sup> the PDTR fragment, found in compound 2, is the most immunogenic domain of MUC1 and, consequently, is a well-known epitope recognized by several anti-MUC1 antibodies;<sup>24</sup> and the GSTA region, included in compound 3, which is recognized by different antibodies and represents a potential tool for use in diagnosis and therapeutic applications<sup>25</sup> (Figure 1). To investigate how the peptide sequence influences MUC1 antigen-lectin recognition, we have elaborated a multidisciplinary approach involving (glyco)peptide synthesis,

enzyme-linked lectin assay (ELLA) tests, isothermal titration microcalorimetry (ITC) studies, X-ray analysis, saturation-transfer difference (STD) NMR experiments, and molecular modeling simulations. Fortunately, our efforts yielded noteworthy results in terms of epitope recognition, showing evidence that the residues that flank the Tn moiety play an active role in the molecular recognition process by two plant lectins: soybean agglutinin (SBA) and *Vicia villosa-B-4* agglutinin (VVA).<sup>26–30</sup> According to our data, Tn antigen is (Ser or Thr) bearing the sugar unit does indeed have a clear effect on binding,<sup>20</sup> little is known about the role that the amino acids flanking the glycosylation point play in the recognition process.<sup>21,22</sup> Herein, we study at the structural level a almost undetectable by the lectins when it is presented in the specific peptide sequence AHGVT( $\alpha$ GalNAc)SA. In contrast, the glycosylated APDT( $\alpha$ GalNAc)R sequence is the best epitope for the two selected lectins. Our work shows chemical evidence indicating that this is not merely a matter of chance. As we describe in the present work, the APDTR sequence fixes the bioactive conformation, establishing direct interactions with the protein to properly present the sugar moiety to the corresponding lectin. Indeed, it is shown that the charged residues of this sequence (Asp and Arg amino acids) play a fundamental role in the binding process. The results presented here can help to better understand glycopeptide–lectin interactions and may contribute to engineer new building sites or to design specific mutants of the lectins, allowing the specificity to be tuned and the structure-based design of novel glycosensors that, within the cancer context, are mandatory to early detection of tumor cells.

### 3.5. RESULTS AND DISCUSSION

We have used two model plant lectins as receptors: soybean (SBA) and *Vicia villosa-B-4* (VVA) agglutinins. These lectins recognize GalNAc<sup>26–30</sup> and are fairly stable and commercially available. The selected glycopeptides shown in Figure 1 comprise the three important regions of MUC1 glycoproteins.

The synthesis of compounds 1–3 was carried out following standard solid-phase peptide synthesis (SPPS) procedures (Methods). First, we investigated the binding properties of these derivatives to SBA lectin by using an enzyme-linked lectin assay (ELLA) (Figure 2).

For this purpose, different amounts of glycopeptides 1–3 (from 0 to 1.5 mM) were covalently attached to a maleic anhydride activated surface. Biotinylated SBA lectin was then selected for estimating the binding. Horseradish peroxidase–3,3',5,5'-tetramethylbenzidine (TMB) was chosen as the detection system, which was quantified by measuring the absorbance of the wells at 450 nm (Methods). According to ELLA measurements, SBA lectin showed the maximum affinity for compound 2. In contrast, binding of glycopeptides 3 and 1 was very low and negligible, respectively.

ELLA affinity evaluation was subsequently corroborated by ITC measurements. The results are in-line with those estimated using ELLA (Table 1 and Figure 2). Affinity of SBA lectin for compound 2 was calculated to be around 15-fold higher than that for glycopeptide 3. In addition, binding affinities between compounds 1 and 2 differ by 2 orders of magnitude. Thermodynamic binding parameters (Table 1 and Supporting Information Figure S1) for glycopeptide 2 and SBA highlight a binding process highly favored by enthalpy, with a reduce entropy penalty when compared to that of the corresponding glycopeptide 3 counterpart. These results indicate that more intermolecular interactions should take place between compound 2 and the protein surface of SBA. In addition, conformational behavior, in solution and complexed to SBA, could be distinct when glycopeptides 2 and 3 are compared.

Encouraged by these results, we decided to investigate whether the affinity of these short molecules could be modified by the presence of the rest of the amino acids that comprise the MUC1 tandem repeat domain. Consequently, we synthesized and studied the glycopeptides shown in Scheme 1. The results obtained from the ELLA performed on these glycopeptides were equivalent to those obtained for short glycopeptides 1–3, with derivative m2 being the compound that clearly showed the best binding to SBA lectin (Supporting Information Figure S2). Taking into account that compound m1, with the lower affinity, has its epitope close to the plastic surface and to prevent artifacts due to accessibility issues, we synthesized glycopeptide m1', with the epitope situated opposite the amino acid linked to the plastic surface (Supporting Information Figure S3). The comparable results obtained for m1 and m1' ruled out the existence of accessibility difficulties in the recognition of glycopeptide m1



by SBA lectin. To generalize our results, we also estimated the binding of glycopeptides m1–m3 to *Vicia villosa* agglutinin (VVA) lectin. The results were in agreement with those observed for SBA lectin (Supporting Information Figure S4).

All of these tests disclose that the underlying peptide sequence does play an active role in the molecular recognition process. In fact, our preliminary studies with glycine-rich glycopeptides bearing Tn antigen indicate a clear decrease of affinity with SBA lectin.<sup>20</sup>

To rationalize these experimental findings, and considering the similar results found for the short (1–3) and large glycopeptides (m1–m3), we performed a thorough conformational analysis in solution of glycopeptides 1–3 in both the free and bound states with SBA lectin. First, full assignment of the protons of the three compounds was performed using COSY and HSQC experiments (Supporting Information). Of note, in glycopeptides 2 and 3, a second set of signals (a small percentage) was observed, especially in the NH region. They correspond to the *cis* disposition of the amide bond of proline residues.<sup>31</sup> Next, the proton–proton distances with conformational information deduced from 2D-NOESY experiments (Figure 3) were used as restraints in MD simulations. In all MD simulations, only the set of signals corresponding to the *trans* disposition of the amide bond of Pro was considered. Due to the presumable flexibility of the glycopeptides, we used MD simulations with time-averaged restraints (MD-tar). These simulations have been successfully applied to flexible systems<sup>32–37</sup> and provide a distribution of low-energy conformers able to quantitatively reproduce the NMR data. A good agreement between experimental and theoretically derived distances was obtained for the three glycopeptides (Supporting Information Tables S1–S3).

Remarkably, while glycopeptide 2 is rather rigid in solution, compounds 1 and 3 are flexible (Figure 3). Compound 2 exhibits a main conformation (populated around 71%) characterized by an inverse  $\gamma$ -turn that comprises Pro2 and Asp3 residues. The conformer with an additional  $\gamma$ -turn between Asp3 and Thr4 residues, reported for other MUC1-like glycopeptides in aqueous solution,<sup>24,31,38–40</sup> is present in about 29% of the total trajectory time (Supporting

Information Figure S5). The main conformer of glycopeptide 2 appears to assist the presentation of the GalNAc residue to the target protein. No significant interactions between the side chain of the Asp3 and the peptide backbone or the GalNAc unit were found. Additionally, the electrostatic contact between Arg5 and Asp3 (distance N $\epsilon$ -Arg5/C $\gamma$ -Asp3 < 6.0 Å) appears ca. 30% of the total trajectory time. The geometry of the glycosidic linkage is fairly similar in the three derivatives. Actually, in all cases,  $\psi$  torsional angle takes values around 120°, characteristic of an eclipsed conformation.<sup>41</sup> This geometry is supported by a key NOE cross-peak between the NH group of the threonine residue and the NH of GalNAc (Supporting Information Figure S6).

As a next step, conformation of glycopeptides 1–3 complexed to SBA lectin was studied. We obtained the X-ray structure of the SBA:2 complex (Figure 4 and Table S4 in the Supporting Information), which is the first time that a structure between SBA lectin and a Tn antigen derivative has been solved and described. Of note, 12 complexes form the cell unit with two different binding modes for glycopeptide 2 complexed to SBA lectin (Figure 4, binding modes A and B). The electronic density for Ala1 was not observed in any of the complexes, indicating that this residue is exposed to the solvent and shows high flexibility due to the lack of interactions with the lectin. Although the 3D dispositions of glycopeptide 2 in binding modes A and B slightly differ in the geometry of their glycosidic linkage, the major changes are associated with the peptide backbone conformation, particularly with Thr4. As in solution, the  $\psi$  torsional angle of the glycosidic linkage of glycopeptide 2 adopts the expected eclipsed conformation in the bound state<sup>41</sup> in both binding modes, with values around 120°. Accordingly, the sugar moiety has an almost perpendicular arrangement with respect to the peptide. In contrast, while Thr4 adopts a folded conformation in binding mode A, it displays an extended geometry in mode B. Nevertheless, the overall shape of both conformations of glycopeptide 2 bound to the lectin is rather similar to a root-mean-squared deviation (rmsd) value for the heavy atoms of 0.94 Å (Supporting Information Figure S7). Interestingly, these two conformations for Thr4 of glycopeptide 2 are also found in solution (Figure 3 and Supporting Information Figure S8). Analysis of the X-ray structures reveals that the interaction pattern between glycopeptide 2 in binding

mode A differs from that observed in binding mode B. In mode A, the hydroxyl groups of the GalNAc unit interact with the lectin through different hydrogen bonds. In particular, O<sub>3</sub> and O<sub>4</sub> form hydrogen bonds with the side chain of Asp88. Additionally, O<sub>3</sub> participates in a hydrogen bond with Asn130 and O<sub>4</sub> with Leu214. The GalNAc unit establishes an extra hydrogen bond between O<sub>6</sub> and Asp215. The carbonyl group of the N-acetyl moiety of GalNAc interacts with the NH of Gly106 (Figure 4). This latter interaction may explain the higher affinity of SBA lectin for GalNAc in comparison to that for Gal.<sup>42,43</sup> A CH... $\pi$  contact between the  $\alpha$ -face of the sugar and Phe128 is also established (Figure 4). It is important to point out that the calculated interactions between the sugar and the lectin by Qasba and co-workers are in agreement with those observed in our crystal structure.<sup>44</sup> Regarding the interactions between the peptide and the lectin, the methyl group of Thr4 is involved in a hydrophobic pocket with Phe128. This contact may be responsible for the higher affinity observed for SBA lectin for Tn antigen when the underlying amino acid is a threonine residue.<sup>20</sup> Furthermore, Arg5 participates in a salt bridge with Glu113 and a hydrogen bond with the hydroxyl of Tyr107. On the other hand, the side chain of Asp3 forms a hydrogen bond with the side chains of Asn130 and Ser131 (Figure 4). In contrast, the interactions of glycopeptide 2 with Ser131 and Tyr107 are negligible in binding mode B. Interestingly, in both binding modes, Asp3 participates in one hydrogen bond with the NH group of GalNAc and in a second one with the NH group of Thr4 (Supporting Information Figure S9a). In addition, Asp3 is involved in an electrostatic interaction with the side chain of Arg5. As can be observed in Figure S9, the oxygen of the carboxylic group of the side chain of Asp3 mimics the role of a bridging water molecule suggested by us between the sugar and the peptide moieties in solution for the Tn derivatives<sup>41</sup> (Supporting Information Figure S9b). This aspartic acid residue may hold an effective role by constraining the conformation of both the peptide backbone and the sugar moiety and favoring the appropriate presentation of the sugar to the protein receptor.

With the aim of supporting the active role of Asp3 and Arg5 in the molecular recognition process, we synthesized glycopeptide 4 in which Asp3 and Arg5 were replaced by Ala residues. This novel derivative could not be

recognized by SBA (Supporting Information Figure S10), corroborating the structural evidence that these two amino acids are important for the recognition process.

We also determined the epitope mapping of glycopeptides 1–3 complexed to SBA by (STD) NMR experiments. Figure 5 shows the STD experiments performed on the SBA:2 complex.

The analysis indicates that the main epitope of the glycopeptide is located at the sugar moiety, with strong STD signals for H2, H3, H4, and the N-acetyl group of GalNAc. Additionally, these experiments show a clear interaction between the lectin and the peptide backbone. Hence, the methyl group of Thr4, ring of Pro2, and side chain of Arg5 are in closer contact with the protein. The STD experiments are compatible with the two binding modes observed in the X-ray structure for glycopeptide 2. The weak STD signal detected for the methyl group of Ala1 (<30%) indicates that this residue should be solvent-exposed and explains the poor electronic density found in the X-ray structure for Ala1. This feature is also corroborated by the titration experiments carried out in a solution of SBA lectin with glycopeptide 2. For ligand–protein systems displaying an intermediate exchange rate ( $K_D \pm 1$  mM) in the chemical shift scale, the observed resonance frequencies of the signals do not shift when the concentration of the ligand is varied; however, they experience a significant loss in signal intensity due to line broadening.<sup>45</sup> Thus, the protons that preferentially bind to the protein-binding pocket show a significant decrease of their signal intensities relative to those far from the receptor recognition site. Additionally, qualitatively speaking, our <sup>1</sup>H NMR-monitored titration of glycopeptide 2 with SBA (4:1, 4:2, 4:3, and 1:1 SBA:2 ratio) shows that the signal intensity of CH<sub>3</sub> protons of Ala1 increases slightly faster than that observed for CH<sub>3</sub> protons of Thr4, supporting the idea that the methyl group of Thr4 exhibits a stronger interaction with the protein than that of the methyl group of Ala1. A STD competition binding study between GalNAc and compound 2 toward SBA was also carried out. As expected, both ligands compete for the same binding site, but, interestingly, the affinity of SBA lectin is higher for glycopeptide 2 than for the monosaccharide, which is in good agreement with reported data.<sup>43</sup> In fact, regarding the epitope mapping, the GalNAc unit receives a higher percentage of

saturation from the protein when it is included in the APDTR sequence then it does as a monosaccharide (Supporting Information Figure S11).

A putative 3D model of the SBA:2 complex in aqueous solution was deduced by two independent 100 ns unrestrained MD simulations, using the X-ray structure of the complexes with the two binding modes of glycopeptide 2 as starting coordinates. Logically, this compound was conveniently modified to contain the missing Ala1 residue. The simulations predict that both binding modes are stable in solution (Supporting Information Figures S12 and S13). A super-position of various frames obtained from the 100 ns unrestrained MD simulations, using binding mode A as the initial coordinates, is shown in Figure 6a. The interactions between the sugar and the protein are the same as those in the solid state (Figure 6b), except for Leu214, which prefers to interact with O6, and are also in good agreement with the STD experiments on the SBA:2 complex described above (Figure 5).

The CH $\cdots\pi$  contact between the  $\alpha$ -face of GalNAc and Phe128 is also present. Additionally, the simulations predict the characteristic eclipsed conformation of the glycosidic linkage.

(Supporting Information Figures S12 and S13). In both simulations, the peptide backbone shows more flexibility at the N-terminus region, especially for Ala1. This feature is in good agreement with the lack of electronic density for this residue in the solid state and with the rather weak STD signal observed for Ala1. MD simulations indicate that Pro2 establishes a hydrophobic interaction with the lateral chain of Arg129. This interaction, not observed in the X-ray structure, explains the STD effect found for this Pro residue (Supporting Information Figure S14). The MD simulation starting from binding mode A shows that the hydrogen bond between Asp3 and Asn130 of SBA lectin is also significant, with a population around 46%. An additional hydrogen bond, populated at about 40%, is observed between Asp3 and Ser131 of the protein. The hydrogen bonds Asp3–Thr4 and Asp3–GalNAc (Supporting Information Figure S9a) are only occasionally populated (close to 15% of the total trajectory time). In addition, the salt bridge Arg5–Glu113, observed in the X-ray structure, has a very low population (around 2%). According to our model, Arg5 also establishes

an electrostatic interaction with Asp215 (not observed in the X-ray structure) that is present in about 25% of the total trajectory time (Figure 6c). Therefore, the lateral chain of Arg5 is fairly flexible and can interact with different residues of the protein (Figure 6c). Similarly, the interactions observed in the X-ray structure, when binding mode B is present, are also retained in MD simulation (Supporting Information Figure S15).

Finally, the binding features of glycopeptides 1 and 3 complexed to SBA lectin were also investigated in solution by combining (STD) NMR experiments with unrestrained MD simulations. As expected, the main epitope of these glycopeptides is mostly located at the sugar moiety (Supporting Information Figures S16–S18). STD experiments suggest a slightly different binding mode for the sugar unit in the three glycopeptides. Interestingly, in the SBA:2 complex, the sugar receives more saturation from the protein, supporting it having the strongest interactions with the protein binding site. This experimental finding is also in agreement with the MD simulations performed on glycopeptides 1–3 complexed to SBA (Supporting Information Figure S15). According to (STD) NMR experiments and MD simulations, the contacts between the peptide backbone and protein residues in compounds 1 and 3 are reduced in comparison to those observed in compound 2 bound to SBA lectin. Indeed, the MD simulations indicated that the interactions between glycopeptides 1 and 3 with the surface of SBA are limited to the sugar unit and the glycosylated Thr residue (Supporting Information Figures S17 and S18). This feature may explain the better affinity and the enthalpy values measured by ITC experiments (Table 1) for glycopeptide 2 with SBA.

Additionally, MD simulations carried out on SBA:1 clearly indicate a higher degree of flexibility for compound 1 when complexed to SBA (Supporting Information Figures S19a and S20), which is in agreement with the poor affinity showed by the lectin. Although merely speculative, the presence of a His residue (10% positive charge at pH 7) could work against binding with SBA lectin due to the proximity of this residue to a positive surface potential at the protein (Supporting Information Figure S22). On the other hand, for compound 3 bound to SBA lectin, according to MD simulations, a type II  $\beta$ -turn that comprises the carbonyl group of Ala1 and the NH group of Ser4 may

stabilize the peptide backbone in the bound state. This backbone conformation allows  $H\alpha$  of Ser4 to establish a hydrophobic interaction with Phe128 (Supporting Information Figures S19b and S21). This feature, among others factors, may explain the better affinity of SBA for compound 3 when compared with that for derivative 1.

**Conclusions.** This work provides striking results for the recognition of MUC1-derived peptides by SBA lectin, in terms of epitope selection. Although it is well-known that lectins recognize sugar moieties, we have shown herein that the recognition process is modulated by the peptide sequence close to the glycosylation point. In particular, the studied lectin receptors show higher affinity to the Tn antigen when it is placed at the PDTR peptide region. In contrast, an affinity loss is observed when this antigen is located at the AHGVTSA or APGSTAP fragments. Asp and Arg residues in the PDTR sequence appear to fix the conformation of both the peptide and the sugar moieties to maximize the interactions with the lectins. These results could help to better understand glycopeptide–lectin interactions and could contribute to engineer new binding sites in lectins with higher affinity for MUC1 glycopeptides, allowing the specificity to be tuned and the design of novel glycosensors with enhanced detection of Tn antigen.

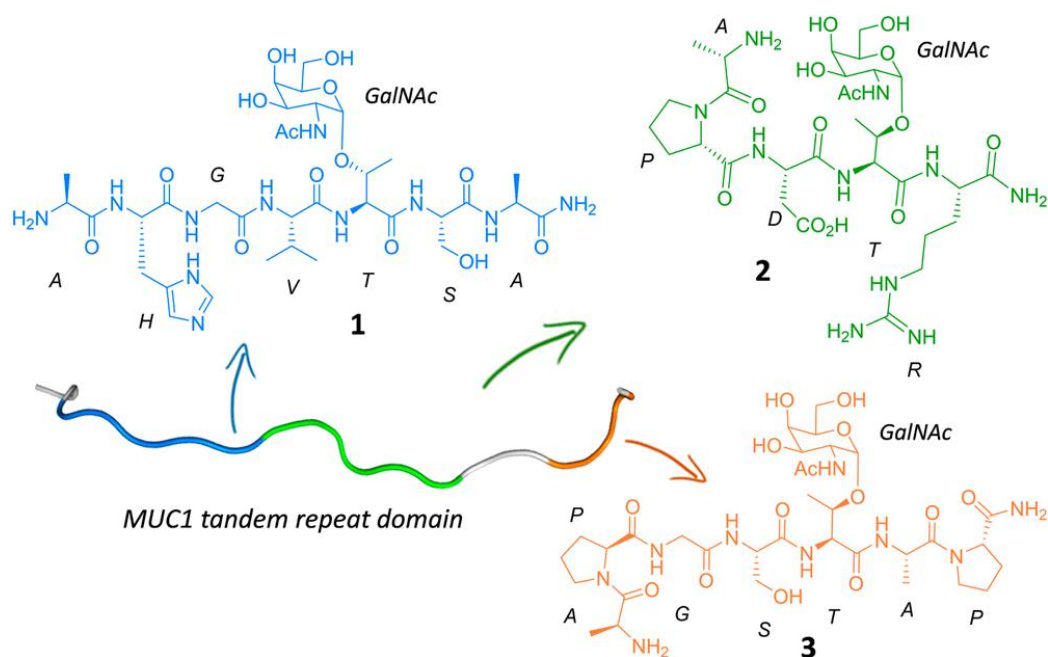


Figure 1. Small glycopeptides synthesized and studied in this work that contain the three relevant epitopes of mucin MUC1.

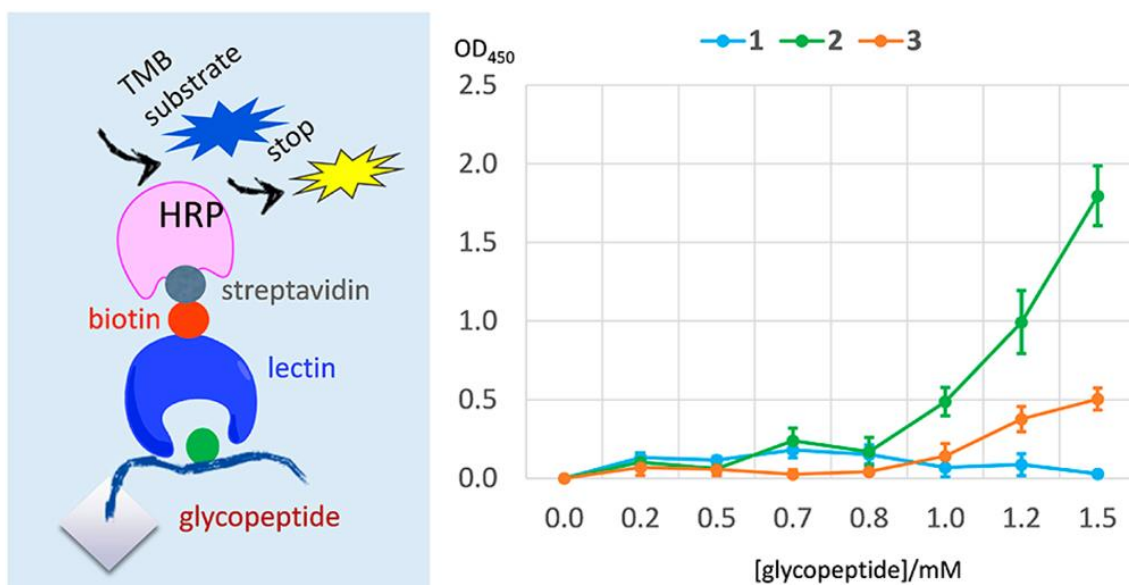
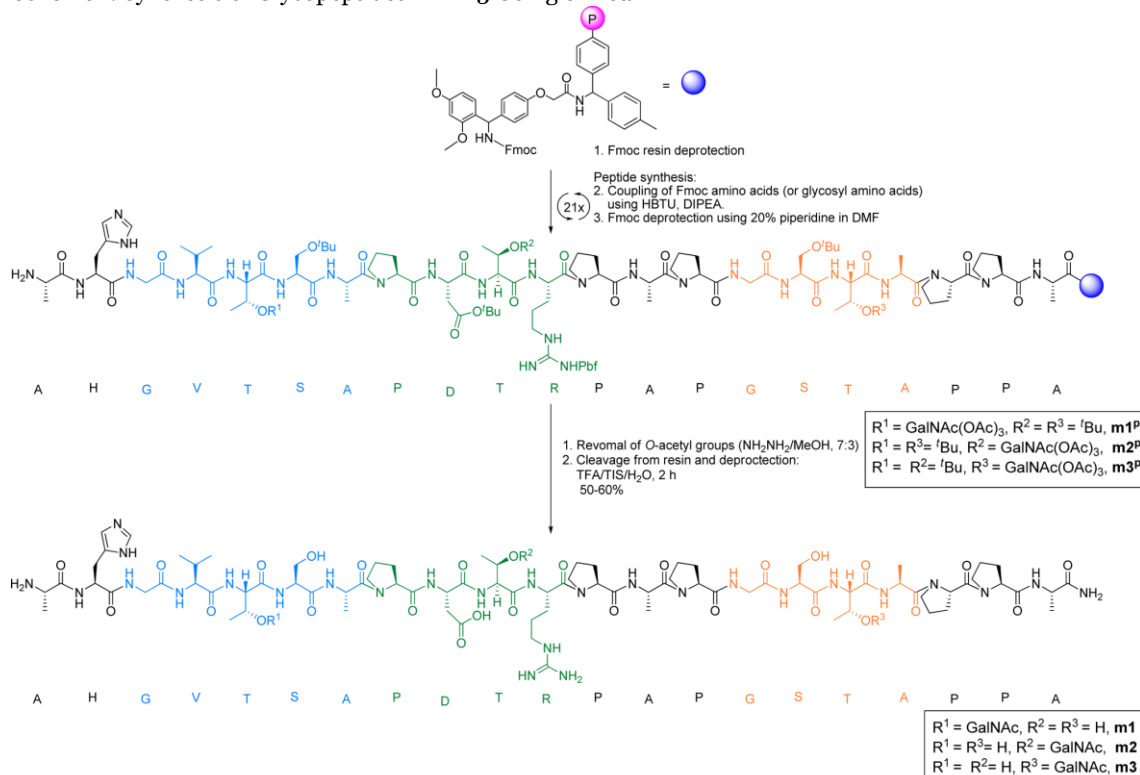


Figure 2. ELLA strategy used in this work together with binding affinity of glycopeptides 1–3 to SBA lectin. Absorbance signals are the average of three replicate wells, and the error bars show the standard deviation of these measurements.

Table 1. Thermodynamic Binding Parameters Obtained by ITC for SBA Lectin and Glycopeptides 1–3 at 25 °C and pH 7.5

	$K_D$ ( $\mu\text{M}$ )	$\Delta G$ ( $\text{kcal mol}^{-1}$ )	$\Delta H$ ( $\text{kcal mol}^{-1}$ )	$T\Delta S$ ( $\text{kcal mol}^{-1}$ )	$n$
1	$\geq 800$				1
2	$6.6 \pm 1.9$	$-7.1 \pm 0.3$	$-20.6 \pm 2.9$	$-13.5 \pm 3.2$	$0.9 \pm 0.3$
3	$100.7 \pm 20.1$	$-5.5 \pm 0.2$	$-8.7 \pm 1.6$	$-3.3 \pm 1.8$	$1.02 \pm 0.2$

Scheme 1. Synthesis of Glycopeptides m1–m3 Using SPPSa



<sup>a</sup>HBTU = O-(1H-benzotriazol-1-yl)-N,N,N',N'-tetramethyluronium hexafluorophosphate. DIPEA = N,N-diisopropylethylamine. TFA = trifluoroacetic acid. TIS = triisopropylsilane.



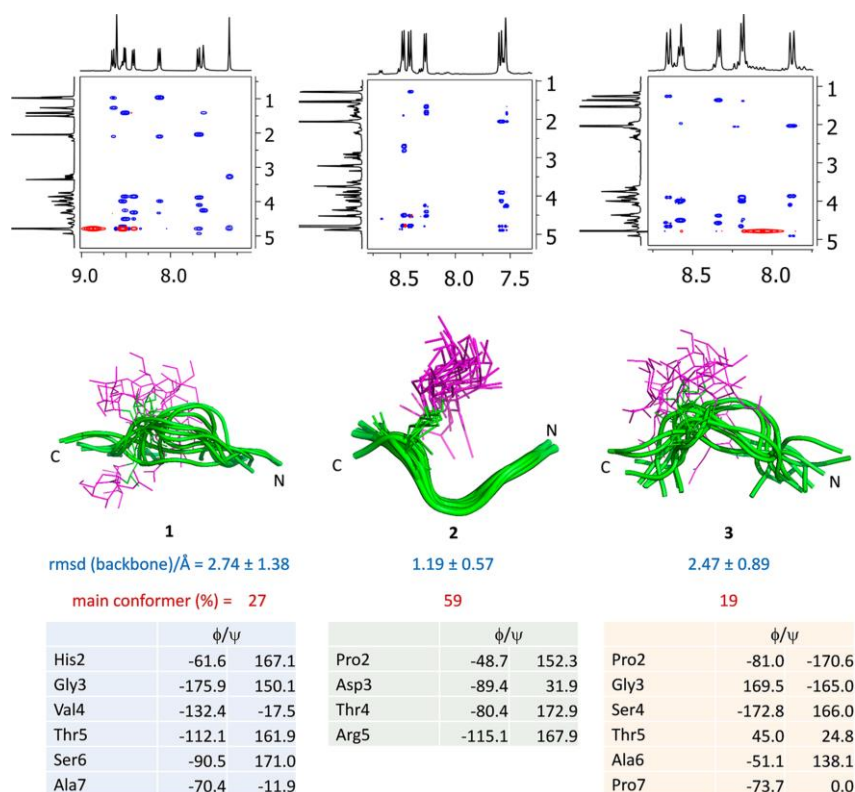


Figure 3. 2D-NOESY ( $\text{H}_2\text{O}/\text{D}_2\text{O}$ , 9:1, pH = 6.5, 293 K) spectra of glycopeptides 1 (left), 2 (middle), and 3 (right), together with the ensembles obtained from 20 ns MD $_{\text{H}_2\text{O}}$ -tar simulations for these derivatives. Diagonal peaks and exchange cross-peaks connecting NH protons and water are negative (red color). The NOE contacts are represented as positive cross-peaks (blue color). The root-mean-squared deviation (rmsd) values for the peptide backbone, together with the population (%) and the peptide geometry of the main conformation in solution, is also shown.

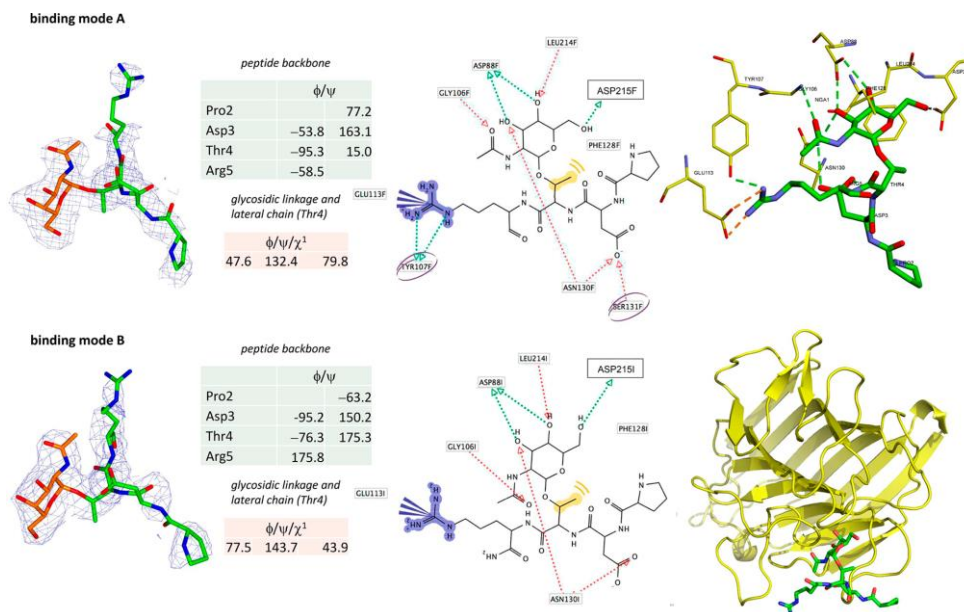


Figure 4. X-ray structure of SBA lectin in complex with compound 2. Two binding modes for glycopeptide 2 were observed. The experimental electron density map obtained for glycopeptide 2 (in both binding modes) is shown, together with the geometry of the peptide backbone and the glycosidic linkage. A schematic representation of the hydrogen-bond network between the SBA lectin and glycopeptide 2 is shown, highlighting the differences between the two binding modes. The hydrophobic contact between the methyl group of Thr and Phe128 is shown in yellow. A  $\text{CH}\cdots\pi$  interaction between GalNAc and Phe128 is also present in the X-ray structure. Two different molecular representations for each binding model are shown on the right.

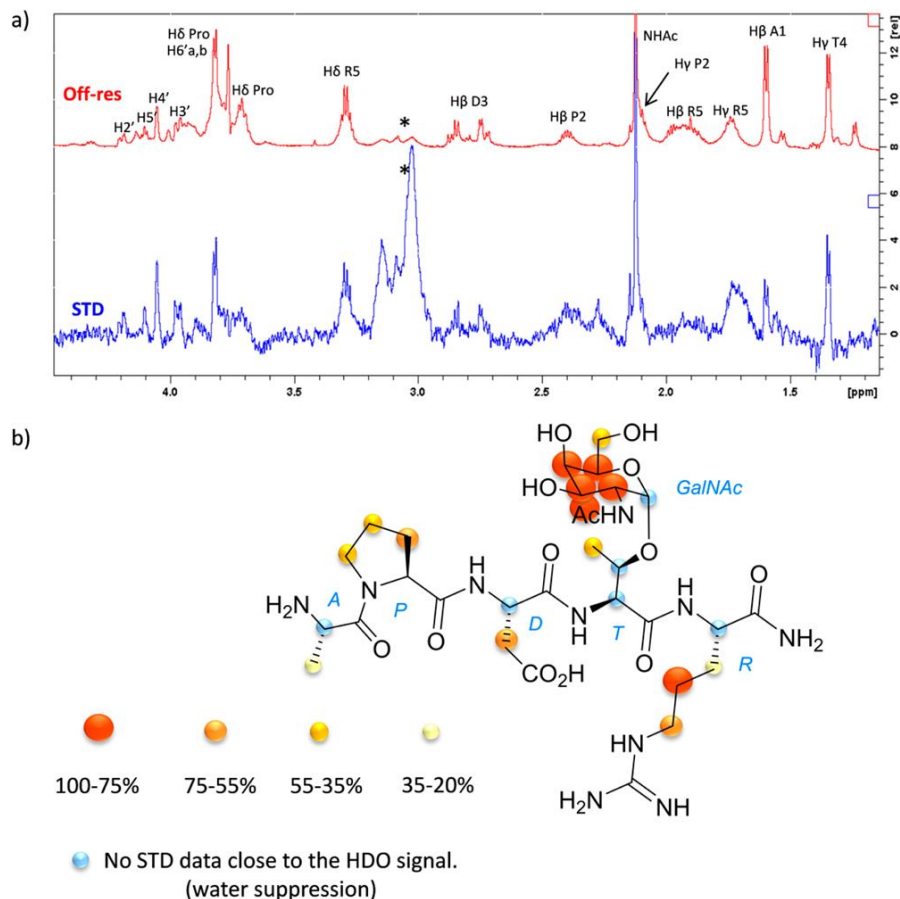


Figure 5. (STD) NMR experiments. (a) STD (blue) and offresonance spectrum (red) of glycopeptide 2 in the presence of SBA. Key proton resonances are marked. (b) Epitope mapping of glycopeptide 2 complexed to SBA lectin. The signal of the anomeric proton of GalNAc and H $\alpha$  of the amino acids could not be analyzed in the STD spectra due to their close distance to the HDO signal.

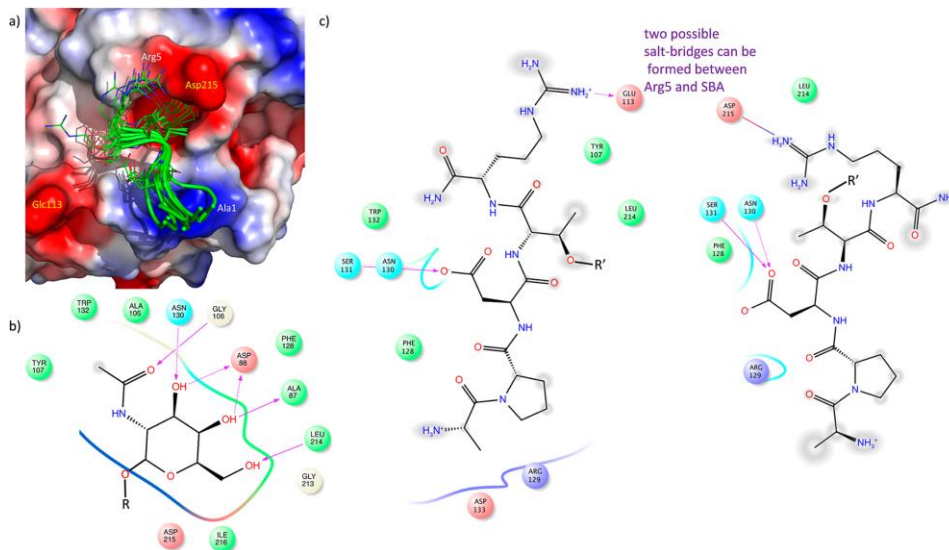


Figure 6. Structure of the SBA:2 complex deduced from MD. (a) Ensembles obtained from the unrestrained 100 ns MD simulations performed on the SBA:2 complex. The electrostatic potential surface of the protein is also represented. Red represents a net negative surface potential, and blue denotes a net positive surface potential. Glycopeptide 2 is shown as a stick model colored according to atom type. (b) Hydrogen bonds between GalNAc moiety and SBA lectin obtained from a representative frame of the unrestrained 100 ns MD simulations carried out on the SBA:2 complex. (c) Peptide interaction diagrams for the SBA:2 complex. MD simulations indicate that Arg5 can have an electrostatic interaction with Asp215 or Glu113.

### 3.6. METHODS

NMR Experiments. COSY, HSQC, and NOESY NMR experiments were performed on a 400 MHz spectrometer at 298 K. Saturation-transfer difference (STD) NMR experiments were recorded on a 600 MHz spectrometer equipped with a triple channel cryoprobe head. The molar ratio was adjusted between 4:1 to 30:1, depending on the ligand–protein complex. STD-NMR experiments were recorded at 298 or 310 K with 2 s irradiation time using pulse sequences with or without water suppression. An excitation sculpting with gradients was employed to suppress the water proton signals.

Glycopeptide Synthesis. All compounds were synthesized by stepwise solid-phase peptide synthesis using the Fmoc strategy. Tn- threonine building block was obtained as described in the literature.<sup>46</sup> This glycosylated amino acid building block was coupled manually, and the other Fmoc amino acids were coupled automatically on a peptide synthesizer. Finally, all compounds were purified by HPLC.

Unrestrained Molecular Dynamics Simulations. All molecular dynamics simulations were carried out on the Finis-Terrae cluster of the Centro de Supercomputación de Galicia (CESGA), Spain. Starting geometries for complexes were generated from the available data deposited in the Protein Data Bank (PDB code: 1SBF) and modified accordingly. A detailed procedure is described in the Supporting Information.

MD Simulations with Time-Averaged Restraints (MD-tar). MD-tar simulations were performed with AMBER 11 (parm99 force field), which was implemented with GLYCAM 06 and ffo3 parameters. Distances derived from NOE cross-peaks were included as time- averaged distance restraints. A  $\langle r^{-6} \rangle^{-1/6}$  average was used for the distances. Final trajectories were run using an exponential decay constant of 2000 ps and a simulation length of 20 ns in explicit TIP3P water molecules.

General Procedure for the Enzyme-Linked Lectin Assay (ELLA). An ELISA plate was coated with 100  $\mu$ L/well of a solution of the corresponding glycopeptide

(0–150 nmol/well) in carbonate/ bicarbonate buffer (0.2 M, pH 9.4) and incubated overnight at 25 °C. The assay procedure was described in our previous work.<sup>20</sup>

Isothermal Titration Calorimetry (ITC). Binding studies were performed at 25 °C in 10 mM phosphate at pH 7.2 by using a titration calorimeter with a reaction cell volume of 300  $\mu$ L.

Crystallization. Crystals were grown by hanging drop diffusion at 18 °C. The drops were prepared by mixing 1  $\mu$ L of protein complex containing 8.5 mg mL<sup>-1</sup> of soybean agglutinin and 10 mM glycopeptide 2 with 1  $\mu$ L of solutions with 20–30% PEG 3350, 0.2 M NaBr, and 0.5  $\mu$ L of 20–30% PEG 400 as additive. Coordinates and structure factors have been deposited in the Worldwide Protein Data Bank (wwPDB; accession code 4D69).

## ASSOCIATED CONTENT

### \*<sub>s</sub>Supporting Information

Additional methods and characterization data.

Figure S1: ITC profiles for glycopeptides 1–3 titration into SBA lectin solutions at 25 °C and pH 7.2.

Figure S2: Binding curves for mucin derivatives m1–m3 with SBA lectin.

Figure S3: Binding curves for mucin derivatives m1 and m1' with SBA lectin.

Figure S4: Binding curves for compounds m1–m3 with VVA lectin.

Figure S5: Main conformations in aqueous solution (free state) for glycopeptides 1–3 obtained from the 20 ns MD-tar in explicit water.

Figure S6: 2D-NOESY spectra of glycopeptides 1–3.

Figure S7: Superposition of the two different conformations found in the X-ray structure of SBA lectin in complex with glycopeptide 2.

Figure S8: Dihedral angles distribution for the backbone of glycopeptide 2 obtained from the 20 ns MD-tar simulations performed in explicit water.

Figure S9: Hydrogen-bond network found in the X-ray structure for Asp3 with sugar and peptide moieties in both binding modes and inter-residue water pocket deduced from the MD-tar simulations on Ac-L-Thr( $\alpha$ -D-GalNAc)-NHMe.

Figure S10: Binding studies of glycopeptides 2 and 4 with SBA lectin using ELLA tests.

Figure S11: <sup>1</sup>H NMR spectra of titration experiments performed with a solution of SBA lectin and increasing amounts of glycopeptide 2 and STD

competition binding experiments between  $\alpha$ - and  $\beta$ -GalNAc vs glycopeptide 2 at 310 K.

Figure S12 and S13: Dihedral angles distribution for the backbone, glycosidic linkage, and side chain of glycopeptide 2 bound to SBA lectin obtained from the unrestrained 100 ns MD-tar simulations performed in explicit water.

Figure S14: Ensembles obtained from the unrestrained 100 ns MD simulations carried out on SBA: glycopeptide 2 complex.

Figure S15: Population of the intramolecular hydrogen bonds found between GalNAc moiety of glycopeptides 1–3 complexed to SBA.

Figure S16: Epitope mapping of glycopeptides 1 and 3 complexed to SBA lectin deduced from STD experiments.

Figure S17:  $^1\text{H}$  NMR 600 MHz spectra for glycopeptide 1 in the presence of SBA at 310 K.

Figure S18:  $^1\text{H}$  NMR 600 MHz spectra for glycopeptide 3 in the presence of SBA at 310 K.

Figure S19: Ensembles obtained from unrestrained 100 ns MD simulations performed on SBA:1 and SBA:3.

Figure S20: Dihedral angles distribution of the peptide backbone obtained from the 100 ns MD simulations in explicit water carried out on SBA:1 complex together with the geometry of the glycosidic linkage and the side chain of Thr5. Figure S21: Dihedral angles distribution of the peptide backbone obtained from the 100 ns MD simulations in explicit water carried out on SBA:3 complex together with the geometry of the glycosidic linkage and the side chain of Thr4.

Figure S22: Representative frame obtained from the unrestrained 100 ns MD simulations on SBA:1.

Table S1: Comparison of the experimental and deduced distances from 20 ns MD<sub>H<sub>2</sub>O</sub>-tar simulations for glycopeptide 1.

Table S2: Comparison of the experimental and deduced distances from 20 ns MD<sub>H<sub>2</sub>O</sub>-tar simulations for glycopeptide 2.

Table S3: Comparison of the experimental and deduced distances from 20 ns MD<sub>H<sub>2</sub>O</sub>-tar simulations for glycopeptide 3.

Table S4: Data collection and refinement statistics for SBA: glycopeptide 2 complex. This material is available free of charge via the Internet at <http://pubs.acs.org>.

#### Accession Codes

Coordinates and structure factors have been deposited in the Worldwide Protein Data Bank (wwPDB) under accession code 4D69.

#### AUTHOR INFORMATION

##### Corresponding Authors

\*(F.C.) E-mail: [francisco.corzana@unirioja.es](mailto:francisco.corzana@unirioja.es).

\*(J.M.P.) E-mail: [jesusmanuel.peregrina@unirioja.es](mailto:jesusmanuel.peregrina@unirioja.es).

##### Present Addresses

+ (H.C. and F.M.) UCIBIO@REQUIMTE, Faculdade Ciências e Tecnologia, Universidade Nova de Lisboa, 2829-516 Caparica, Portugal.

‡ (J.V.-G., J.C.-L., and R.H.-G.) 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, Zaragoza, Spain.

§ (J.L.A.) Instituto de Química Orgánica General, CSIC, Juan de la Cierva 3, 28006 Madrid, Spain.

‡ (J.J.-B.) (a) Structural Biology Unit, CIC bioGUNE, Parque Tecnológico de Bizkaia Building 801A, 48160 Derio, Spain; (b) IKERBASQUE, Basque Foundation for Science, 48011 Bilbao, Spain; (c) Department of Chemical and Physical Biology, Centro de Investigaciones Biológicas, CSIC, Ramiro de Maeztu 9, 28040 Madrid, Spain.

## Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

We thank the Ministerio de Economía y Competitividad (project CTQ2012-36365/FEDER and FPI grant to V.J.S.). N.M.-S. and D.M. thank Universidad de La Rioja for FPI grants. We also thank CESGA for computer support. We thank DLS in Oxford, beamline I03 (MX8035-24). We thank ARAID, MEC (BFU2010-19504 and CTQ2012-32025), DGA (B89), and the FP7 (2007-2013) under BioStruct-X (no. 283570). H.C. and F.M. thank FCT-Portugal for the postdoc research grant (SFRH/BPD/65462/2009) and for the financial support of EXPL/QEQ/MED/0799/2012 and RECI/BBB-BQB/0230/2012.

## REFERENCES

- (1) Arnaud, J., Audfray, A., and Imberty, A. (2013) Binding sugars: from natural lectins to synthetic receptors and engineered neolectins. *Chem. Soc. Rev.* 42, 4798–4813.
- (2) Inamdar, S. R., Savanur, M. A., Eligar, S. M., Chachadi, V. B., Nagre, N. N., Chen, C., Barclays, M., Ingle, A., Mahajan, P., Borges, A., Shastry, P., Kalraiya, R. D., Swamy, B. M., Rhodes, J. M., and Yu, L. G. (2012) The TF-antigen binding lectin from *Sclerotium rolfsii* inhibits growth of human colon cancer cells by inducing apoptosis in vitro and suppresses tumor growth in vivo. *Glycobiology* 22, 1227–1235.
- (3) Varrot, A., Basheer, S. M., and Imberty, A. (2013) Fungal lectins: structure, function and potential applications. *Curr. Opin. Struct. Biol.* 23, 678–685.
- (4) Liu, B., Bian, H.-J., and Bao, J.-K. (2010) Plant lectins: potential antineoplastic drugs from bench to clinic. *Cancer Lett.* 287, 1–12. (5) Fu, L.-L., Zhou, C.-C., Yao, S., Yu, J.-Y.,



Liu, B., and Bao, J.-K. (2011) Plant lectins: targeting programmed cell death pathways as antitumor agents. *Int. J. Biochem. Cell Biol.* 43, 1442–1449.

(6) Poiroux, G., Pitié, M., Culerrier, R., Lafont, E., Ségui, B., Van Damme, E. J. M., Peumans, W. J., Bernadou, J., Levade, T., Rougé, P., Barre, A., and Benoist, H. (2011) Targeting of T/Tn antigens with a plant lectin to kill human leukemia cells by photochemotherapy. *PLoS One* 6, e23315.

(7) Ju, T., Otto, V. I., and Cummings, R. D. (2011) The Tn antigen structural simplicity and biological complexity. *Angew. Chem., Int. Ed.* 50, 1770–1791.

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

(9) Taylor-Papadimitriou, J., Burchell, J., Miles, D. W., and Dalziel, M. (1999) MUC1 and cancer. *Biochim. Biophys. Acta* 1455, 301–313. (10) Hollingsworth, M. A., and Swanson, B. J. (2004) Mucins in cancer: protection and control of the cell surface. *Nat. Rev. Cancer* 4, 45–60.

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

(12) Borgert, A., Heimbürg-Molinaro, J., Song, X., Lasanajak, Y., Ju, T., Liu, M., Thompson, P., Ragupathi, G., Barany, G., Smith, D. F., Cummings, R. D., and Live, D. (2012) Deciphering structural elements of mucin glycoprotein recognition. *ACS Chem. Biol.* 7, 1031–1039.

(13) Gaidzik, N., Westerlind, U., and Kunz, H. (2013) The development of synthetic antitumour vaccines from mucin glycopeptide antigens. *Chem. Soc. Rev.* 42, 4421–4442.

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

(15) Lakshminarayanan, V., Thompson, P., Wolfert, M. A., Buskas, T., Bradley, J. M., Pathangey, L. B., Madsen, C. S., Cohen, P. A., Gendler, S. J., and Boons, G. J. (2012) Immune recognition of tumor-associated mucin MUC1 is achieved by a fully synthetic aberrantly glycosylated MUC1 tripartite vaccine. *Proc. Natl. Acad. Sci. U.S.A.* 109, 261–266.

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

(17) Beatson, R. E., Taylor-Papadimitriou, J., and Burchell, J. M. (2010) MUC1 immunotherapy. *Immunotherapy* 2, 305–327.

(18) Kaszowska, M., Norgren, A. S., Arvidson, P. I., and Sandström, C. (2009) Studies on the interactions between glycosylated  $\beta$ 3-peptides and the lectin *Vicia villosa* by saturation transfer difference NMR spectroscopy. *Carbohydr. Res.* 344, 2577–2580.

(19) Clark, D., and Mao, L. (2012) Cancer biomarker discovery: lectin-based strategies targeting glycoproteins. *Dis. Markers* 33, 1–10.

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

(21) Martin, J. C., Fadda, E., Ito, K., and Woods, R. J. (2013) Defining the structural origin of the substrate sequence independence of O-GlcNAcase using a combination of molecular docking and dynamics simulation. *Glycobiology* 24, 85–96.

(22) Marcelo, F., Garcia-Martin, F., Matsushita, T., Sardinha, J., Coelho, H., Oude-Vrielink, A., Koller, C., André, S., Cabrita, E. J., Gabius, H.-J., Nishimura, S.-I., Jiménez-Barbero, J., and Cañada, F. J.

(2014) Delineating binding modes of Gal/GalNAc and structural elements of the molecular recognition of tumor-associated mucin glycopeptides by the human macrophage galactose-type lectin. *Chem. Eur. J.* 20, 16147–16155.

(23) Wandall, H. H., Hassan, H., Mirgorodskaya, E., Kristensen, A. K., Roepstorff, P., Bennett, E. P., Nielsen, P. A., Hollingsworth, M. A., Burchell, J., Taylor-Papadimitriou, J., and Clausen, H. (1997) Substrate specificities of three members of the human UDP-N-acetyl- $\alpha$ -galactosamine: polypeptide N-acetylgalactosaminyltransferase family, GalNAc-T1, -T2, and -T3. *J. Biol. Chem.* 272, 23503–23514.

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

(25) Huang, Z.-H., Shi, L., Ma, J.-W., Sun, Z.-Y., Cai, H., Chen, Y.-X., Zhao, Y.-F., and Li, Y.-M. (2012) A totally synthetic, self-assembling, adjuvant-free MUC1 glycopeptide vaccine for cancer therapy. *J. Am. Chem. Soc.* 134, 8730–8733.

- (26) Godula, K., and Bertozzi, C. R. (2012) Density variant glycan microarray for evaluating cross-linking of mucin-like glycoconjugates by lectins. *J. Am. Chem. Soc.* 134, 15732–15742.
- (27) Oyelaran, O., Li, Q., Farnsworth, D., and Gildersleeve, J. C. (2009) Microarrays with varying carbohydrate density reveal distinct subpopulations of serum antibodies. *Proteome Res.* 8, 3529–3538.
- (28) Dam, T. K., Gerken, T. A., Cavada, B. S., Nascimento, K. S., Moura, T. R., and Brewer, C. F. (2007) Binding studies of  $\alpha$ -GalNAcspecific lectins to the  $\alpha$ -GalNAc (Tn-antigen) form of porcine submaxillary mucin and its smaller fragments. *J. Biol. Chem.* 282, 28256–28263.
- (29) Puri, K. D., Gopalakrishnan, B., and Surolia, A. (1992) Carbohydrate binding specificity of the Tn-antigen binding lectin from *Vicia villosa* seeds (VVLB4). *FEBS Lett.* 312, 208–212.
- (30) Lotan, R., Siegelman, H. W., Lis, H., and Sharon, N. (1974) Subunit structure of soybean agglutinin. *J. Biol. Chem.* 249, 1219–1224.
- (31) Dziadek, S., Griesinger, C., Kunz, H., and Reinscheid, U. M. (2006) Synthesis and structural model of an  $\alpha$  (2,6)-sialyl-T glycosylated MUC1 eicosapeptide under physiological conditions. *Chem. Eur. J.* 12, 4981–4993.
- (32) Corzana, F., Busto, J. H., Engelsen, S. B., Jiménez-Barbero, J., Asensio, J. L., Peregrina, J. M., and Avenoza, A. (2006) Effect of  $\beta$ -O-glycosylation on L-Ser and L-Thr diamides: a bias toward  $\alpha$ -helical conformations. *Chem. Eur. J.* 12, 7864–7871.
- (33) Jiménez-Moreno, E., Gómez-Pinto, I., Corzana, F., Santana, A. G., Revuelta, J., Bastida, A., Jiménez-Barbero, J., González, C., and Asensio, J. L. (2013) Chemical interrogation of drug/RNA complexes: from chemical reactivity to drug design. *Angew. Chem., Int. Ed.* 52, 3148–3151.
- (34) Muñoz-García, J. C., Corzana, F., de Paz, J. L., Angulo, J., and Nieto, P. M. (2013) Conformations of the iduronate ring in short heparin fragments described by time-averaged distance restrained molecular dynamics. *Glycobiology* 23, 1220–1229.
- (35) Corzana, F., Busto, J. H., Marcelo, F., García de Luis, M., Asensio, J. L., Martín-Santamaría, S., Jiménez-Barbero, J., Avenoza, A., and Peregrina, J. M. (2011) Engineering O-glycosylation points in non-extended peptides: implications for the molecular recognition of short tumor-associated glycopeptides. *Chem. Eur. J.* 17, 3105–3110.

- (36) Hendrickx, P. M. S., Corzana, F., Depraetere, S., Tourwé, D. A., Augustyns, K., and Martins, J. C. (2010) The use of time-averaged 3JHH restrained molecular dynamics (tar-MD) simulations for the conformational analysis of five-membered ring systems: methodology and applications. *J. Comput. Chem.* 31, 561–572.
- (37) Zhang, Z., McCallum, S. A., Xie, J., Nieto, L., Corzana, F., Jiménez-Barbero, J., Chen, M., Liu, J., and Linhardt, R. J. (2008) Solution structures of chemoenzymatically synthesized heparin and its precursors. *J. Am. Chem. Soc.* 130, 12998–13007.
- (38) Matsushita, T., Ohyabu, N., Fujitani, N., Naruchi, K., Shimizu, H., Hinou, H., and Nishimura, S.-I. (2013) Site-specific conformational alteration induced by sialylation of MUC1 tandem repeating glycopeptides at an epitope region for the anti-KL-6 monoclonal antibody. *Biochemistry* 52, 402–414.
- (39) Kinarsky, L., Suryanarayanan, G., Prakash, O., Paulsen, H., Clausen, H., Hanisch, F.-G., Hollingsworth, M. A., and Sherman, S. (2003) Conformational studies on the MUC1 tandem repeat glycopeptides: implication for the enzymatic O-glycosylation of the mucin protein core. *Glycobiology* 13, 929–939.
- (40) Schuman, J., Campbell, A. P., Koganty, R. R., and Longenecker, B. M. (2003) Probing the conformational and dynamical effects of O-glycosylation within the immunodominant region of a MUC1 peptide tumor antigen. *J. Pep. Res.* 61, 91–108.
- (41) Corzana, F., Busto, J. H., Jiménez-Oses, G., García de Luis, M., Asensio, J. L., Jiménez-Barbero, J., Peregrina, J. M., and Avenoza, A. (2007) Serine versus threonine glycosylation: the methyl group causes a drastic alteration on the carbohydrate orientation and on the surrounding water shell. *J. Am. Chem. Soc.* 129, 9458–9467.
- (42) Pereira, M. E., Kabat, E. A., and Sharon, N. (1974) Immunochemical studies on the specificity of soybean agglutinin. *Carbohydr. Res.* 37, 89–102.
- (43) De Boeck, H., Loontjens, F. G., Lis, H., and Sharon, N. (1984) Binding of simple carbohydrates and some N-acetyllactosaminecontaining oligosaccharides to *Erythrina cristagalli* agglutinin as followed with a fluorescent indicator ligand. *Arch. Biochem. Biophys.* 234, 297–304.
- (44) Rao, V. S., Lam, K., and Qasba, P. K. (1998) Three dimensional structure of the soybean agglutinin Gal/GalNAc complexes by homology modeling. *J. Biomol. Struct. Dyn.* 15, 853–860.
- (45) Fernández-Alonso, M. C., Berbis, M. A., Canales, A., Ardá, A., Cañada, F. J., and Jiménez-Barbero, J. (2013) New applications of high-resolution NMR in drug discovery and

development, in *New Applications of NMR in Drug Discovery and Development* (Garrido, L., and Beckmann, N., Eds.) pp 7–42, Royal Society of Chemistry, Cambridge.

(46) Plattner, C., Höfener, M., and Sewald, N. (2011) One-pot azidochlorination of glycals. *Org. Lett.* 13, 545–547.

Article II: Deciphering the Non-Equivalence of Serine and Threonine O-Glycosylation Points: Implications for Molecular Recognition of the Tn Antigen by an anti-MUC1 Antibody\*\*

Nuria Martínez-Sáez, Jorge Castro-López, Jessika Valero-Gonzalez, David Madariaga, Ismael Compañón, Víctor J. Somovilla, Miriam Salvadó, Juan L. Asensio, Jesús Jiménez-Barbero, Alberto Avenoza, Jesús H. Busto, Gonzalo J. L. Bernardes, Jesús M. Peregrina,\* Ramón Hurtado-Guerrero,\* and Francisco Corzana\*

[\*] Dr. N. Martínez-Sáez,<sup>[+]</sup> Dr. D. Madariaga, I. Compañón, Dr. V. J. Somovilla, Dr. A. Avenoza, Dr. J. H. Busto, Dr. J. M. Peregrina, Dr. F. Corzana Departamento de Química, Universidad de La Rioja. Centro de Investigación en Síntesis Química 26006 Logroño (Spain) E-mail: [jesusmanuel.peregrina@unirioja.es](mailto:jesusmanuel.peregrina@unirioja.es); [francisco.corzana@unirioja.es](mailto:francisco.corzana@unirioja.es)

Dr. N. Martínez-Sáez,<sup>[+]</sup> Dr. V. J. Somovilla, M. Salvadó, Dr. G. J. L. Bernardes Department of Chemistry, University of Cambridge. Lensfield Road, Cambridge CB2 1EW (UK)

J. Castro-López,<sup>[+]</sup> J. Valero-González,<sup>[+]</sup> Dr. R. Hurtado-Guerrero Institute of Biocomputation and Physics of Complex Systems (BIFI). University of Zaragoza, BIFI-IQFR (CSIC) Joint Unit. Edificio I+D, 50018 Zaragoza (Spain) and Fundación ARAID Edificio Pignatelli 36, Zaragoza (Spain) E-mail: [rhurtado@bifi.es](mailto:rhurtado@bifi.es)

M. Salvadó Departament de Química Analítica y Química Orgánica, Universitat Rovira i Virgili, C/Marcell. Domingo s/n, 43007 Tarragona (Spain)

Dr. J. L. Asensio Instituto de Química Orgánica General, IQOG-CSIC. Juan de la Cierva 3, 28006 Madrid (Spain)

Dr. J. Jiménez-Barbero Structural Biology Unit, CIC bioGUNE. Parque Tecnológico de Bizkaia Building 801 A, 48160 Derio (Spain) and IKERBASQUE, Basque Foundation for Science 48011 Bilbao (Spain) and Department of Chemical and Physical Biology Centro de Investigaciones Biológicas, CSIC. Ramiro de Maeztu 9, 28040 Madrid (Spain)

Dr. G. J. L. Bernardes Instituto de Medicina Molecular. Faculdade de Medicina da Universidade de Lisboa 1649-028 Lisboa (Portugal)

[+] These authors contributed equally to this work.

[\*\*] We thank the Ministerio de Economía y Competitividad /FEDER (project CTQ2012-36365, CTQ2012-32065, BFU2010-19504, CTQ2013-44367-C2-2-P, UNLR13-4E-1931 and grant I.C.) and DGA (B89) for financial support. N.M.-S. and D.M. thank Universidad de La Rioja for FPI grants. We thank Katherine Stott (Department of Biochemistry, Cambridge University) for technical help with the BLI experiments. G.J.L.B. thanks financial support from the EPSRC. G.J.L.B. is a Royal Society University Research Fellow. M.S. thanks the Generalitat de Catalunya and Universitat Rovira i Virgili for financial support. We thank synchrotron radiation sources DLS(Oxford), and in particular beamlines I04 (experiment number MX8035-26) and I02 (experiment number MX10121-2), respectively. The research leading to these results has also received funding from the FP7 (2007-2013) under BIOSTRUCTX-7687. We also thank CESGA for computer facilities.

### 3.7. ABSTRACT

The structural features of MUC1-like glycopeptides bearing the Tn antigen (a-O-GalNAc-Ser/Thr) in complex with an anti MUC-1 antibody are reported at atomic resolution. For the a-O-GalNAc-Ser derivative, the glycosidic linkage adopts a high-energy conformation, barely populated in the free state. This unusual structure (also observed in an a-SGalNAc- Cys mimic) is stabilized by hydrogen bonds between the peptidic fragment and the sugar. The selection of a particular peptide structure by the antibody is thus propagated to the carbohydrate through carbohydrate/peptide contacts, which force a change in the orientation of the sugar moiety. This seems to be unfeasible in the a-O-GalNAc-Thr glycopeptide owing to the more limited flexibility of the side chain imposed by the methyl group. Our data demonstrate the non-equivalence of Ser and Thr O-glycosylation points in molecular recognition processes. These features provide insight into the occurrence in nature of the APDTRP epitope for anti-MUC1 antibodies.

The Tn antigen (a-O-GalNAc-Ser/Thr) is one of the most specific human tumor-associated structures.[1] This entity, which is also implicated in HIV infection,[2] is expressed in approximately 90 % of carcinomas, and a direct correlation between the aggressiveness of the carcinoma and the occurrence of the antigen has been observed.[3] Consequently, the Tn antigen has found widespread use as biomarker and as a potential therapeutic target against cancer.[1,4–7] Structural analysis of Tn antigen bound to its biological targets is thus of great significance for elucidating the mechanism of recognition, as well as for engineering novel antibodies and biosensors. In general, the Tn antigen is referred to as N- acetylgalactosamine (GalNAc) a-O-linked to serine (Ser) or threonine (Thr), without specifying which of the two amino acids the GalNAc is linked to. However, we and others have observed the existence of subtly different conformational behaviors in solution of the basic Ser- and Thr- containing structures.[8–14]

Herein, we present a detailed analysis of the interaction of these two Tn determinants, as MUC1 glycopeptides, to an anti-MUC1 antibody. MUC1 is a heavily O- glycosylated membrane glycoprotein consisting of tandem repeats of 20 amino acids (AHGVTSAPDTRPAPG- STAPP), with five possible



glycosylation sites.[15,16] This protein is overexpressed and partially glycosylated in cancer cells. Consequently, some peptide fragments that are masked in healthy cells, such as APDTRP and their glycosylated analogues, are now accessible and can interact with the immune system. Although the observed enhancement of antibody affinity has been attributed to conformational changes induced by the glycan in the peptide backbone,[17–20] the molecular basis for this observation remains unclear.

To our knowledge, the only reported crystal structure of a complex between an antibody and a GalNAc-containing glycopeptide is unrelated to mucins.[21] In addition, the X-ray structure of a model anti-MUC1 antibody (SM3) in complex with a mucin is limited to a naked peptide.[22] Moreover, an NMR study on this peptide and its corresponding GalNAc- glycopeptide bound to SM3 led to the hypothesis that the sugar residue fixes the bioactive conformation of the peptide fragment and interacts via the N-acetyl group with the surface of the antibody.[23] However, no detailed information on the intermolecular interactions could be deduced from this ligand-based NMR analysis.

A detailed analysis of the interactions between the SM3 antibody and two synthetic glycopeptides bearing the  $\alpha$ -O- GalNAc-Thr and  $\alpha$ -O-GalNAc-Ser antigens (m1\* and m2\* respectively; Figure 1) is presented herein. These molecules include the tandem repeat sequence of MUC1. The SM3 antibody was selected because its epitope recognition mode is similar to that of other anti-MUC1 antibodies,[24] which expands the scope of these results, and also because of its potential for use in the early diagnosis and treatment of breast cancer.[22]

The influence of the chemical nature of the underlying amino acid, as well as the GalNAcylation, on antibody binding affinity was first evaluated. For this purpose, the related naked peptides (m1 and m2, respectively; Figure 1) were also synthesized and tested. The KD constants for the MUC1 variants with the scFv-SM3 antibody were experimentally determined through bio-layer interferometry (BLI) experiments. The higher affinity (around 3-fold) of SM3 for glycosylated m1\* compared the naked peptide m1 was confirmed by these

tests [24] (Figure 1). Furthermore, the Ser- containing compounds showed significantly lower affinity, thus highlighting the differences between the two Tn antigens. These results were corroborated by ELISA tests (see the Supporting Information).

To explain these results at the atomic level, a scFv-SM3 antibody was produced and purified. High quality crystals of the SM3:1, SM3:1\*, and SM3:2\* complexes were obtained, where 1, 1\* and 2\* are simplified models of m1, m1\* and m2\*, respectively (Figure 2 a, PDB IDs 5a2j, 5a2k, and 5a2i). These compounds include the peptide fragment that represents the minimal epitope recognized by most anti-MUC1 antibodies.[24]

The obtained crystals enabled solution of the structures at high resolution (< 2.0 Å, see the Supporting Information). Crystallographic analysis revealed that the surface groove of the recombinant SM3 antibody nicely fits all of the peptide residues in the three studied complexes (Figure 2 b), independent of the presence of the sugar moiety. Moreover, the overall conformation of the peptide fragment of the different simplified MUC1 variants is nearly identical and is similar to that found in the crystal structure reported for the naked peptide [22] (Figure 2c and Figure 3).

Therefore, the presence of the GalNAc moiety, regardless of the attached amino acid (Ser or Thr), does not significantly affect the conformation of the peptide backbone in the SM3- bound state. In fact, the antibody–antigen pairs display the same pattern of interactions with each of the three substrates (1, 1\* and 2\*) and with the reported peptide [22] (Figure 3). The stabilizing contacts in these complexes involve several hydrogen bonds, some of them mediated by water molecules, as well as several stacking interactions (Figure 3 and the Supporting Information). Pro2 stacks with Trp91L, Trp96L, and Tyr32L, while the side-chains of Asp3 and Arg5 are engaged in hydrophobic contacts with Trp 33H and Tyr 32H, respectively. Additionally, in compounds 1\* and 2\*, the NH group of Ala1 and the carbonyl group of Thr4/Ser4 are involved in hydrogen bonds with Tyr 32L and Gln97H, respectively. The 3D presentation of the side chain of Arg5 in glycopeptide 2\* (Figure 2 c) differs slightly from that found for the other analogues. As a result, there is no hydrogen bond between the side

chain of Arg5 and the carbonyl group of Asn31H. This interaction, among other factors (see below), may be at the cause of the low affinity of SM3 for 2\*.

Interestingly, the major difference between glycopeptides 1\* and 2\* bound to SM3 resides in the geometry of the glycosidic linkage (Figure 4 a). In fact, in the SM3:1\* complex, this linkage adopts the expected exo-anomeric/syn conformation, with  $\phi$  and  $\psi$  values of around 638 and 918, respectively.[9] This conformation is similar to that found for a non-related MUC1 glycopeptide bound to 237-mAb[21] (Figure 4 b) and that exhibited for 1\* bound to a model lectin (Soybean agglutinin).[25] This geometry allows the formation of an intermolecular hydrogen bond between the hydroxymethyl group of GalNAc and the side chain of Tyr32L on the SM3 antibody. Moreover, the N-acetyl group of the sugar stacks with the aromatic ring of Trp 33H, thus providing the impetus for the observed selectivity of SM3 for GalNAc-containing antigens. This interaction is compatible with the previous solution NMR data.[23] By contrast, the GalNAc unit of compound 2\* establishes only weak water-mediated hydrogen bonds with the antibody (see the Supporting Information).

The glycosidic linkage shows a high-energy conformation, with a  $\psi$  value of around 978 (Figure 4 a, c). This conformation is barely populated for the a-O-GalNAc-Ser motif in solution. In fact, the conformational analysis performed on glycopeptides 1\* and 2\* in the free state in water by means of NMR experiments and Molecular Dynamics (MD) simulations with time-averaged restraints[9] indicates that compound 2\* displays this conformation with a population approximately 20 % (see the Supporting Information). As previously observed by us, [8,9] while in compound 1\*, the glycosidic linkage adopts mainly the typical eclipsed conformation, in variant 2\* it prefers to adopt the alternate conformation. According to the X-ray structure, the unusual conformation of the glycosidic linkage in compound 2\* bound to the antibody is stabilized by two intramolecular hydrogen bonds with the peptide chain (Figure 5). The GalNAc endocyclic oxygen O5 engages in a hydrogen bond with the NH group of the attached Ser and the O6 interacts with the side chain of Asp3. To our knowledge, this geometry of the glycosidic linkage has not been previously observed for protein-bound Tn-containing peptides. It is important to note that the quality of the SM3:2\* structure was lower and that the B-factor for 2\* was

higher compared to the other structures, which may indicate some degree of flexibility.

MD simulations performed for the SM3:2\* complex corroborated this flexibility in the bound state. The X-ray conformation was retained only for the first 4 ns of the trajectory (see the Supporting Information). Then, the glycosidic linkage exhibits the typical parallel orientation found for this compound in the free state.[8] To reinforce the significance of the b-methyl group of the threonine residue for the conformation of the glycosidic linkage, and to corroborate the exceptional 3D conformation of glycopeptide 2\* in the complex with the antibody, we solved the structure of the cysteine analogue in complex with scFv-SM3 (compound 3\* in Figure 2, PDB ID: 5a2L). Notably, the GalNAc unit and the peptide backbone adopt an almost identical spatial conformation in glycopeptides 2\* and 3\*, with the main difference being the conformation of the side chain of Arg5 (Figure 5). Therefore, the selection of a particular peptide structure by the antibody is somehow propagated to the carbohydrate through specific carbohydrate/peptide contacts, thereby forcing a drastic change in the orientation of the sugar moiety. This situation is not possible in the Thr-containing derivative owing to the limited conformational freedom of its side chain imposed by the methyl group (Figure 5). As a result, the GalNAc unit adopts a completely different presentation when linked to the threonine residue, with most of the hydroxy groups exposed and able to interact with the corresponding partners of the immune system.

In conclusion, we have uncovered, at the atomic level, the reasons why Ser- and Thr-linked glycopeptides bind differently to SM3. We have provided experimental evidences for distinct presentations of the Tn-carrying serine (α-GalNAc-Ser) and threonine (α-GalNAc-Thr) antigens when bound to SM3. The reasons for the observed limited enhancement in SM3 affinity in 1\* versus 1 can be attributed to the weak hydrogen bond between O6 and Tyr32L, and a hydrophobic contact between the methyl group of the GalNAc unit and Trp33H. These findings emphasize the differences between these two Tn antigens in the context of recognition by anti-MUC1 antibodies and may have important implications for the design of novel antibodies and biosensors. In addition, our

findings may provide insight into the occurrence in nature of the APDTRP epitope for anti-MUC1 antibodies.

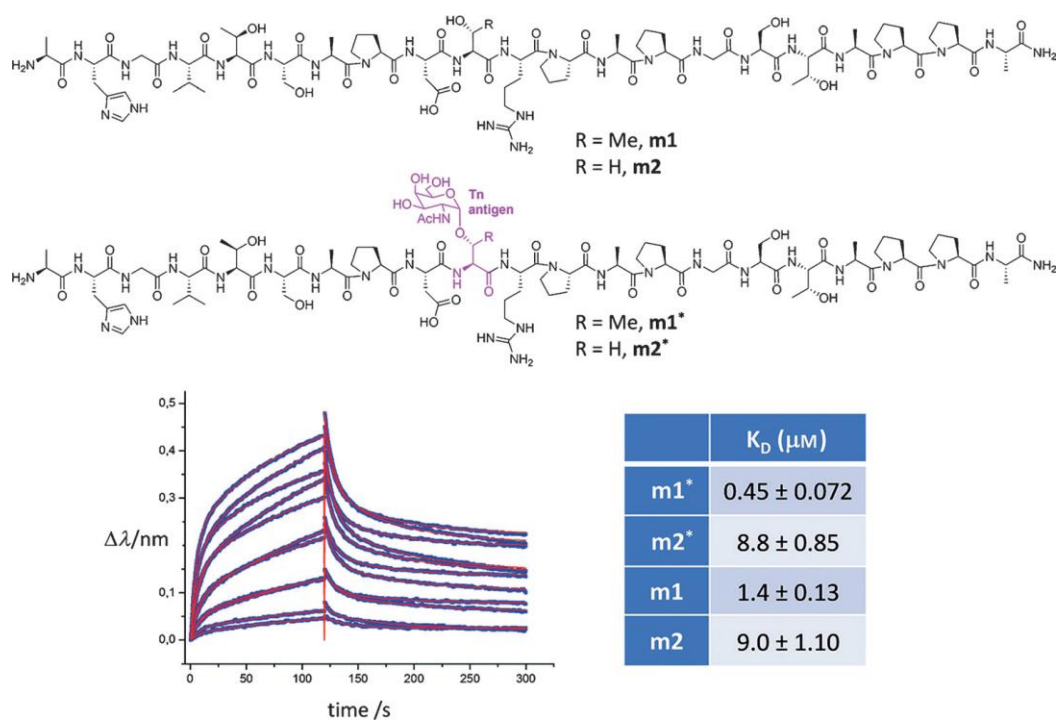


Figure 1. The MUC1-like peptides and glycopeptides studied in this work (upper panel). Bio-layer interferometry (BLI) curves and fit obtained for glycopeptide m1\* and scFv-SM3, together with the KD constants derived from BLI experiments for all of the MUC1-related compounds (lower panel).

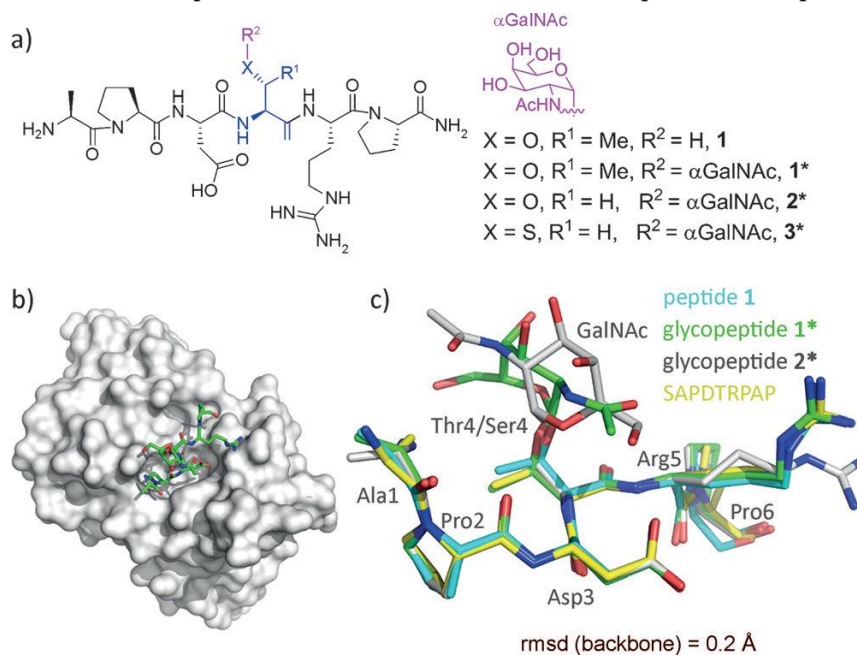


Figure 2. a) Simplified MUC1 variants used for crystallization. b) Surface representation of SM3 in complex with 1\*. The antigen is shown as a stick model with carbon atoms in green. c) Superposition of the peptide backbone of compounds 1, 1\*, 2\*, and the SAPDTRPAP peptide<sub>[22]</sub> in complex with SM3.

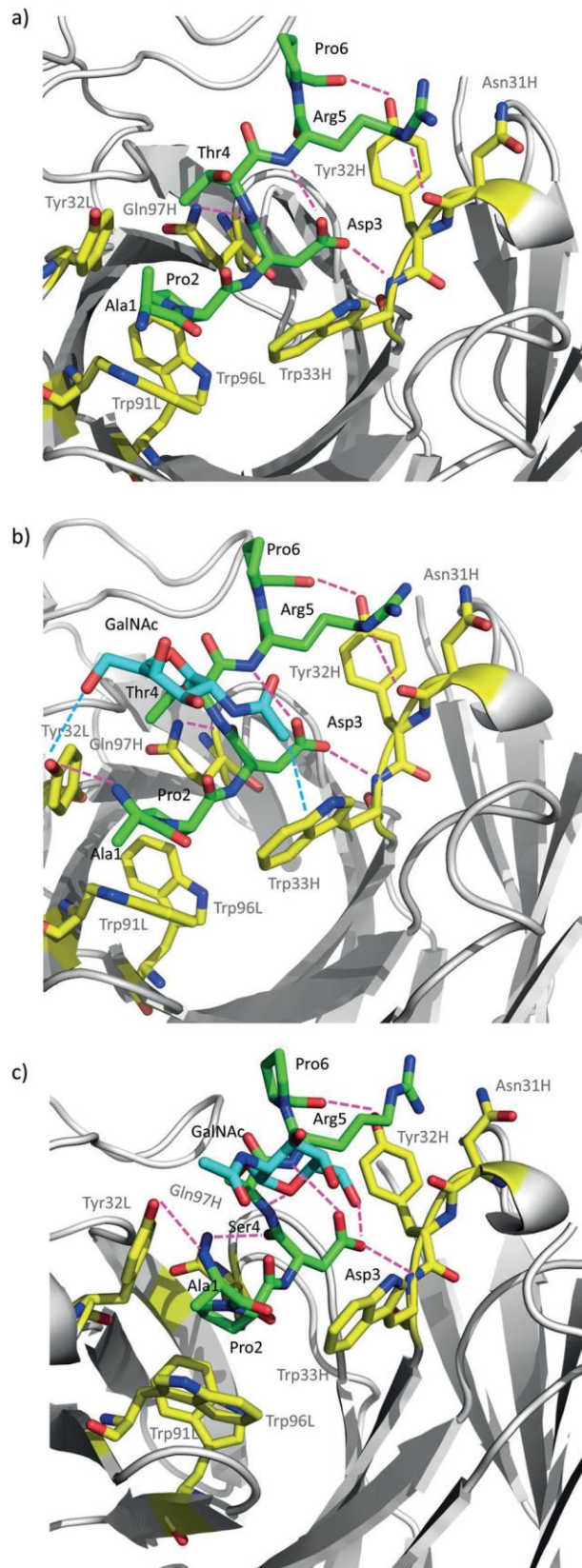


Figure 3. Key binding interactions of peptide 1 (a), glycopeptide 1\* (b), and glycopeptide 2\* (c) with SM3 mAb, as observed in the X-ray crystal structures. Peptide and glycopeptide carbon atoms are shown in green. GalNAc carbon atoms are shown in cyan. Carbon atoms of key residues of SM3 are colored yellow. Blue dashed lines indicate interactions between GalNAc and SM3 surface, and pink dashed lines indicate hydrogen bonds between peptide backbones and SM3 antibody.

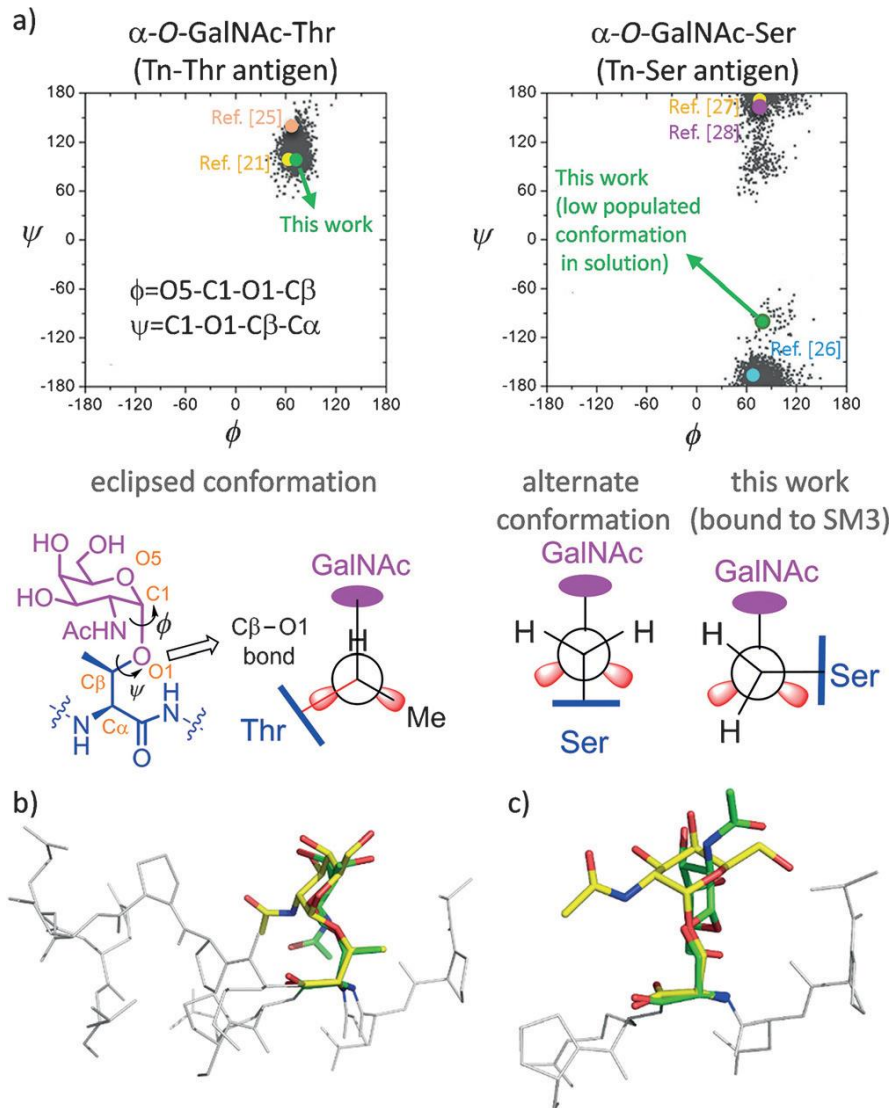


Figure 4. a) Distribution of  $\phi/\psi$  torsional angles found for  $\alpha$ -O-GalNAc-Ser and  $\alpha$ -O-GalNAc-Thr in water,<sup>[8, 9]</sup> together with the geometries found in X-ray structures for these determinants when bound to some biological targets.<sup>[21, 25–28]</sup> Newman projections of Cb–O1 bond are shown. b) Superposition of the Tn antigen moiety  $\alpha$ -O-GalNAc-Thr bound to SM3 (in green) and 237-mAb (in yellow; mAb=monoclonal antibody).<sup>[21]</sup> c) Superposition of the Tn antigen moiety  $\alpha$ -O-GalNAc-Ser bound to SM3 mAb (in green) and to HPA lectin (in yellow).<sup>[27]</sup>

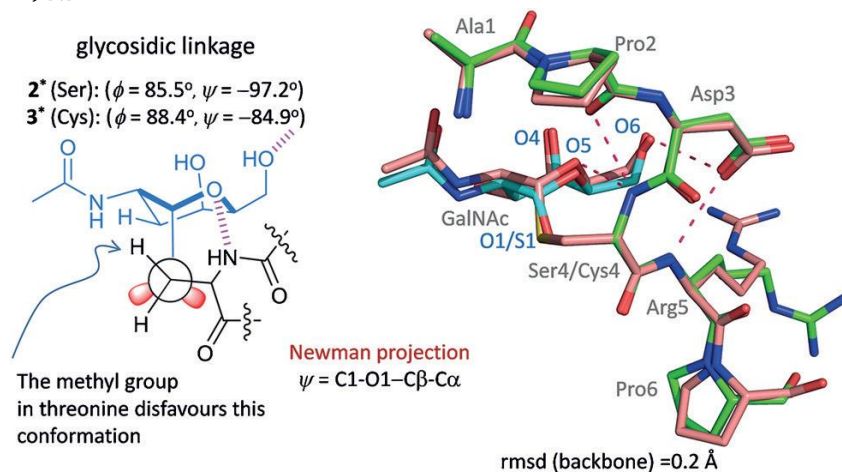


Figure 5. Conformation of glycopeptides 2\* (in green and cyan) and 3\* (in brown) in complex with scFv-SM3 antibody, together with the geometry of the glycosidic linkage and the hydrogen bonds established with the peptide fragment.

### Article III: Design of $\alpha$ -S-Neoglycopeptides Derived from MUC1 with a Flexible and Solvent-Exposed Sugar Moiety

Víctor Rojas-Ocáriz,<sup>†,‡</sup> Ismael Compañón,<sup>†,‡</sup> Carlos Aydillo,<sup>†,§</sup> Jorge Castro-López,<sup>||</sup> Jesús Jiménez-Barbero,<sup>#,∇,°</sup> Ramón Hurtado-Guerrero,<sup>||,⊥</sup> Alberto Avenoz,<sup>†</sup> María M. Zurbano,<sup>†</sup> Jesús M. Peregrina,<sup>\*,†</sup> Jesús H. Busto,<sup>\*,†</sup> and Francisco Corzana<sup>\*,†</sup>

<sup>†</sup>Departamento de Química, Centro de Investigación en Síntesis Química, Universidad de La Rioja, Madre de Dios 53, 26006 Logroño, Spain.

<sup>||</sup>BIFI, University of Zaragoza, BIFI-IQFR (CSIC) Joint Unit, Mariano Esquillor s/n, Campus Rio Ebro, Edificio I+D, Zaragoza, Spain.

<sup>⊥</sup>Fundación ARAID, 50018 Zaragoza, Spain.

<sup>#</sup>Structural Biology Unit, CIC bioGUNE, Parque Tecnológico de Bizkaia Building 801 A, 48160 Derio, Spain.

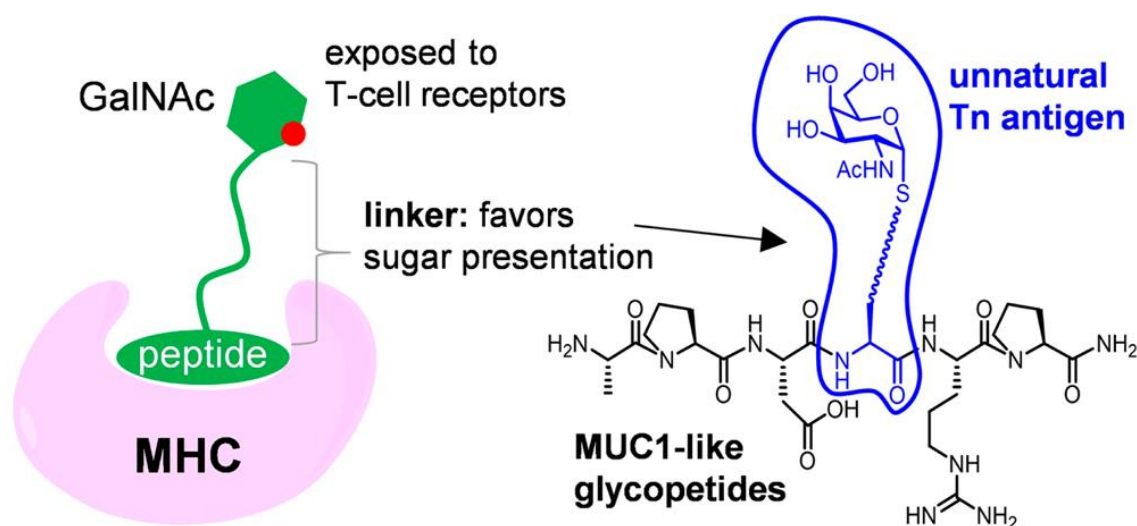
<sup>∇</sup>IKERBASQUE, Basque Foundation for Science, 48011 Bilbao, Spain.

<sup>°</sup>Department of Chemical and Physical Biology, Centro de Investigaciones Biológicas, CSIC Ramiro de Maeztu 9, 28040 Madrid, Spain



### 3.8. 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.



### 3.9. 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.<sup>1-4</sup> 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 ( $\alpha$ -O-GalNAc-Ser/ Thr),<sup>5</sup> are exposed to the immune system and can be recognized by different antibodies.<sup>6</sup> For this reason, MUC1 derivatives are attracting great interest as a potential tool in developing therapeutic vaccines for the treatment of cancer.<sup>4,7,8</sup> However, to date, none of these vaccines have succeeded in clinical trials.<sup>9</sup> On one hand, natural TACAs are tolerated by the immune system.<sup>10</sup> To overcome this problem, chemical modifications of these antigens, generating non-natural determinants, have been proposed.<sup>7,11-20</sup> On the other hand, most glycan antigens are B-cell epitopes and produce only short-lived and low-affinity antibodies.<sup>4</sup> 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.<sup>21</sup> The crystal structure of an MHC-I molecule with a glycopeptide containing a linker between the peptide and the sugar<sup>21e</sup> 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).

In an effort to combine these two concepts, we present herein a set of unnatural Tn antigens characterized by the incorporation of a flexible and variable linker in their structures, placed between a peptide fragment and the GalNAc moiety (Figure 1c). These new sulfa-Tn derivatives have been prepared as Fmoc-protected glycosyl amino acids, ready to use in solidphase peptide synthesis (SPPS). They have been included in the Ala-Pro-Asp-Xaa( $\alpha$ -S-

GalNAc)-Arg-Pro peptide sequence, which constitutes the key immunogenic epitope of MUC1<sub>6</sub> where Xaa is a threonine. In particular, in this work, we synthesized and studied six neoglycopeptides, where Xaa is a cysteine, homocysteine, homohomocysteine, O-(mercaptopropyl) serine, S-(mercaptopropyl)cysteine, or O-(mercaptopropyl)threonine residue. The main conformations of these molecules in solution have been evaluated. The linker plays a key role in the conformational behavior of these molecules, from the global 3D structure to the presentation of the glycan part of these compounds. Indeed, the linker blocks the direct contact between the sugar moiety and the backbone, promotes a helix-like conformation for the glycosylated residue, provides extra flexibility to the molecule, and might favor the presentation of the sugar unit for molecular recognition events. These novel compounds are indeed mimics of MUC1 antigens as demonstrated by the analysis of the X-ray diffraction data obtained for one of these unnatural glycopeptides complexed to an anti-MUC1 monoclonal antibody. The global behavior of this complex has also been analyzed by MD simulations.

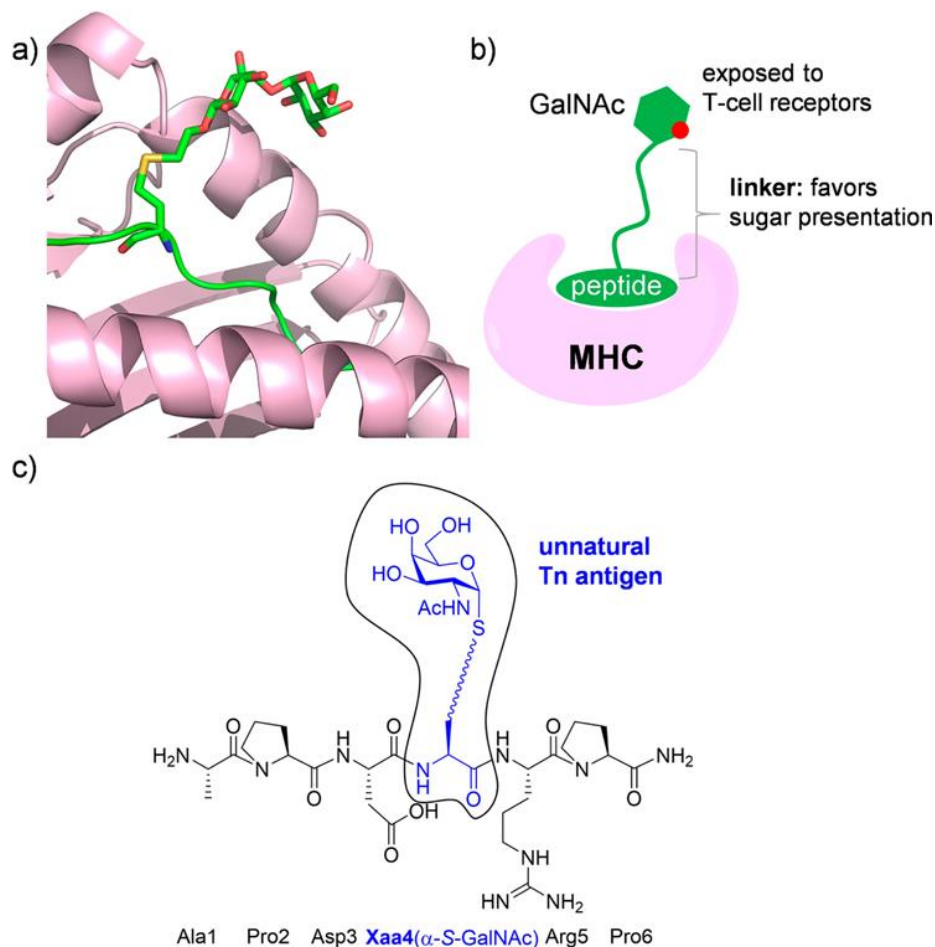


Figure 1. (a) Crystal structure of a MHC-I molecule in complex with a glycopeptide containing a linker between the peptide backbone and the sugar moiety (PDB id 1KBG, ref 21e). (b) Design of novel glycopeptides containing a linker that favors the presentation of the sugar moiety and facilitates the interaction with T-cell receptors. (c) General structure of the neoglycopeptides synthesized and studied in this work. The peptide sequence PDTRP constitutes the most immunogenic domain of mucin MUC1.6

### 3.10. RESULTS AND DISCUSSION

From a synthetic perspective, the existence of S-glycopeptides with important biological properties has stimulated the development of a variety of glycosylation methods. These methods generally use protected carbohydrates as electrophiles and Cys derivatives as nucleophiles. Alternatively, dehydroalanine (Dha)<sup>22–24</sup> containing peptides and nucleophilic thiocarbohydrate derivatives<sup>25</sup> have been used in 1,4-conjugate additions with poor results in diastereoselectivity.<sup>26</sup> Recently, we reported a new chiral Dha derivative,<sup>27</sup> which has been used as acceptor in stereoselective sulfa-Michael additions of appropriately protected thiocarbohydrates. This procedure has been applied to the synthesis of S-(2-acetamido-2-deoxy- $\alpha$ -Dgalactopyranosyl)- L-cysteine,

which could be considered as a mimic of the Tn antigen.<sup>28</sup> A method based on the alkene hydrothiolation reaction,<sup>29</sup> which is called thiolene coupling (TEC), has also been developed. The thermally or photochemically induced TEC versions occur through a radical mechanism in anti-Markovnikov regioselective fashion.<sup>29</sup> These methods have been applied, in general, to the synthesis of  $\beta$ -S-glycosyl amino acids and  $\beta$ -S-glycopeptides.<sup>29</sup> Nevertheless, the  $\alpha$ -S-glycosidic bond formation in these motifs,<sup>29c</sup> especially the  $\alpha$ -S-GalNAc moiety of the Tn antigen,<sup>30,31</sup> has received little attention. In that context, the synthesis of  $\alpha$ -thio-linked glycosyl amino acids and glycopeptides by an anomeric thiol group S-alkylation with  $\beta$ -iodo(or bromo)alanine has been described.<sup>30a</sup> Later, this methodology was modified to a “two-step one-pot” reaction, generating  $\alpha$ -thio-linked glycosyl amino acids by condensation of a  $\beta$ -bromoalanine derivative with the corresponding thiolate salt.<sup>31</sup>

We report herein new stereocontrolled entries to adequately protected S-glycosyl amino acids, mimics of the Tn antigen, as building blocks for solid-phase peptide synthesis of S-linked glycopeptides, exploiting the reactivity of the sulfhydryl group of tri-O-acetyl-2-acetamido-2-deoxy-1-thio- $\alpha$ -D-galactopyranose (1, abbreviated as GalNAc- $\alpha$ -SH) in two types of reactions. The first one comprises the hydrothiolation reaction of double bonds included in the side chains of amino acid derivatives. This reaction has been extensively used in glycochemistry.<sup>32</sup> Moreover, the mild reaction conditions, atom economy, and regioselectivity of this process satisfy essential requirements of the click concept.<sup>33</sup> The second type of reaction assayed involves the anomeric thiol group S-alkylation with bromo- amino acid derivatives.

To know the viability of this project, we started with the synthesis of the shortest glycopeptide (S-glycosyl amino acid ended in amide groups) that includes the largest distance between carbohydrate and peptide backbone moieties. We selected GalNAc- $\alpha$ -SH (1) as a glycosyl donor, whose synthesis was made following the procedure described in the literature<sup>34a</sup> but properly modified to carry out in a gram scale (Supporting Information). GalNAc- $\alpha$ -SH (1) was subjected to an alkene hydrothiolation reaction with O-allyl serine derivative 2b-Ser, which was prepared according to the procedure described<sup>34b</sup> from commercially available N-Boc-serine derivative 2a-Ser. The

hydrothiolation reaction was carried out using azobis- (isobutyronitrile) (AIBN) as a radical initiator. The best conditions involved the use of 1.7 equiv of thiol 1, 1.2 equiv of AIBN, and 1.0 equiv of alkene 2b-Ser to afford compound 3 in a 56% yield (Scheme 1). This derivative was transformed into the required diamide 6. Deprotection of the Boc group, followed by acetylation of the amino group, gave compound 4. This compound was subjected to hydrogenolysis of benzyl ester, and the corresponding carboxylic acid was transformed into methyl amide. Diamide derivative 5 was deacetylated to give target compound 6 (Scheme 1).

Encouraged by this result, we synthesized the corresponding Fmoc-protected 5-glycosyl amino acids 10-Ser/Cys/Thr, ready to use in SPPS and in a gram scale. To this aim, it was necessary to prepare the (O or S)-allyl N-Fmoc-serine/cysteine/ threonine tert-butyl esters 9-Ser/Cys/Thr, which had not been described to date. Their synthesis started from commercially available N-Boc-amino acids 2a-Ser/Cys/Thr, whose Boc groups were removed, giving amino acids 7-Ser/ Cys/Thr. The amino groups of these compounds were then protected as Fmoc carbamates, affording compounds 8-Ser/ Cys/Thr. Esterifications of acids 8-Ser/Cys/Thr with tert-butyl alcohol gave the required allylic derivatives 9-Ser/Cys/Thr (Scheme 2).

As a next step, we assayed hydrothiolation reactions between glycosyl-thiol 1 and alkenes 9-Ser/Cys/Thr, following the conditions above described for compound 2b-Ser, obtaining good yields of the required building blocks 10-Ser and 10-Thr (70% and 71%, respectively). However, compound 10-Cys was obtained only in a 20% yield (Scheme 3). To increase these yields, the photochemical version of the hydrothiolation reaction<sup>29a,b</sup> was tested because it is considered as a lightinduced click reaction,<sup>35</sup> with applications to diverse fields of study of biomolecular systems. We used a photochemical reactor equipped with 16 UVA (350 nm) lamps of 8 W. The reaction was initiated with 2,2-dimethoxy-2-phenylacetophenone (DPAP), and after testing several conditions, the best yields involved the use of 1.0 equiv of alkene, 1.2 equiv of compound 1, and 0.4 equiv of DPAP in dichloromethane (1 mL). In this way, building blocks 10-Ser/Cys/Thr were obtained in a 53%, 30%, and 62%, respectively (Scheme 3). It is well known that this type of reaction involving O- or S-allyl derivatives

works with moderate to low yields.<sup>36</sup> We also tested a recent method with 9-Ser/Cys/Thr, involving the use of  $\text{BEt}_3$  as an initiator and catechol as a coreagent.<sup>37</sup> Although only a slight improvement of the yield was achieved, this method provided an improved way to purify the desired compounds (Scheme 3).

Compound 12, which contains a three-methylene bridge between S- $\alpha$ -GalNAc and amino acid moieties, was synthesized following the same hydrothiolation protocol, tested again by both thermal and photochemical ways, starting from  $\alpha$ -allylglycine 11. This protected compound was obtained according to the protocol previously published,<sup>38</sup> and it was reacted with glycosyl donor 1, giving a 70% and 50% yield of desired building block 12 in thermal and photochemical ways, respectively (Scheme 3). We also tested the repairing way using catechol/ $\text{BEt}_3$ , achieving compound 12 in a better yield than the photochemical way (70%).

To access GalNAc-S-glycosylated building blocks 14 and 16, L-homocysteine (hCys) and L-cysteine (Cys) derivatives, respectively, we used a different methodology based on the anomeric thiol group S-alkylation with bromo-amino acid derivatives (Scheme 4). This methodology has been abovecommented<sup>30a,31</sup> as well as their difficulties to carry out the nucleophilic substitution in a basic medium in the presence of Fmoc protecting groups. We followed a recently published mild synthetic strategy<sup>39</sup> that simply requires activated molecular sieves to selectively promote S-alkylation. Therefore, activated molecular sieves (4 Å) were employed as a base to promote Salkylation of GalNAc- $\alpha$ -SH 1 with appropriate protected  $\gamma$ -bromohomoalanine<sup>40</sup> 13 and  $\beta$ -bromoalanine<sup>30,41</sup> 15 using dry DMF as a solvent (Scheme 4).

Although the reactions proceeded with moderate yields, the procedure was easy to perform and fast and provided similar overall yields when it was compared with other methods published recently.<sup>28</sup> These known bromo derivatives 13 and 15 were obtained from L-homoserine and L-serine, respectively, following the procedure described in the literature, that involved

N-Fmoc protection, tert-butyl ester formation, and bromination under standard conditions.<sup>30</sup>

These six mimics of the Tn antigen (10-Ser/Cys/Thr, 12, 14, and 16) are adequate building blocks for SPPS of glycopeptides. Consequently, they were used to synthesize different  $\alpha$ -S-neoglycopeptides shown in Scheme 5. The automated SPPS protocol with a Rink amide MBHA resin and Fmoc and side chain protected amino acids.  $\alpha$ -S-Glycosylamino acid building blocks 10-Ser/Cys/Thr, 12, 14, and 16 were manually coupled to the peptide sequence to improve the yield of this step. After removal of the acetyl groups of the carbohydrate with a mixture of  $\text{NH}_2\text{NH}_2/\text{MeOH}$ ,  $\alpha$ -S-neoglycopeptides were released from the resin using trifluoroacetic acid (TFA). Under these conditions, all acidlabile side-chain protecting groups were also removed. Purification by preparative HPLC and subsequent lyophilization gave the target  $\alpha$ -S-neoglycopeptides 17–22 in good overall yields.

The conformational behavior of the synthesized neoglycopeptides in solution was then deduced by using NMR. In particular, NOE data, supported by molecular dynamics (MD) simulations with time-averaged restraints (MD-tar), were employed.<sup>42,43</sup> In a first step, full assignment of the protons in all derivatives was carried out using standard COSY and HSQC experiments. 2D-NOESY experiments in  $\text{H}_2\text{O}/\text{D}_2\text{O}$  (9/1) (25 °C, pH = 5.5) were then performed (Figure 2 and Supporting Information). Distances involving NH protons were semiquantitatively deduced by integrating the volume of the corresponding cross-peaks. This experimental data was used as restraints in MD-tar simulations (Figure 2 and Supporting Information).

The structure ensembles obtained through the MD-tar simulations indicate that the neoglycopeptides are rather flexible in solution (Figure 3 and Supporting Information).

In general, the studied neoglycopeptides exhibit a random coil conformation for the peptide backbone. This result was corroborated by circular dichroism (CD) spectroscopic studies (Supporting Information). The CD spectra



of these derivatives differed from that observed for the natural APDT( $\alpha$ -OGalNAc) RP glycopeptide, which adopts mainly a PPII conformation for the peptide backbone in water.<sup>44</sup> (Supporting Information). As a point of interest, the contacts between the sugar and the peptide fragment are almost insignificant. As a consequence, the glycosylated residue exhibits a helix-like conformation in all studied compounds (Figure 4a for compound 22 and Supporting Information for the other compounds). This feature is corroborated by the medium-size NOE intensities observed for the  $\text{NH}_{\text{Xaa4}}-\text{H}\alpha_{\text{Xaa4}}$  and  $\text{NH}_{\text{Arg5}}-\text{H}\alpha_{\text{Xaa4}}$  cross-peaks, where Xaa is the glycosylated residue.<sup>45</sup> Conversely, it has been shown that GalNAc-glycosylation of the natural Ser or Thr residues forces these amino acids to display extended conformations,<sup>46–52</sup> due to the presence of stabilizing contacts between the sugar and the peptide moieties.<sup>46a,50</sup>

Of note, compounds 19–22, which are bearing a longer linker, adopt a low population (<25%) of a type I  $\beta$ -turn-like, stabilized by a hydrogen bond between Pro2 (CO) and Arg5 (NH), which has been reported for natural MUC1-like derivatives<sup>52</sup> (Figure 4b).

Concerning the S-glycosidic linkage, as stated by us for other derivatives,<sup>34</sup> it is rather flexible in all compounds, with two main conformations, characterized by  $\psi$  ca.  $180^\circ$  or  $60^\circ$ , corresponding to the local minima calculated for methyl 4-thio- $\alpha$ -maltoside.<sup>53</sup> The studied glycopeptides display a conformation centered at  $\phi/\psi \approx 60^\circ/-60^\circ$ , also close to that observed in the crystal structure of a mucin-like glycopeptide incorporating GalNAc-Cys<sub>44</sub> (Figure 4c and Supporting Information).

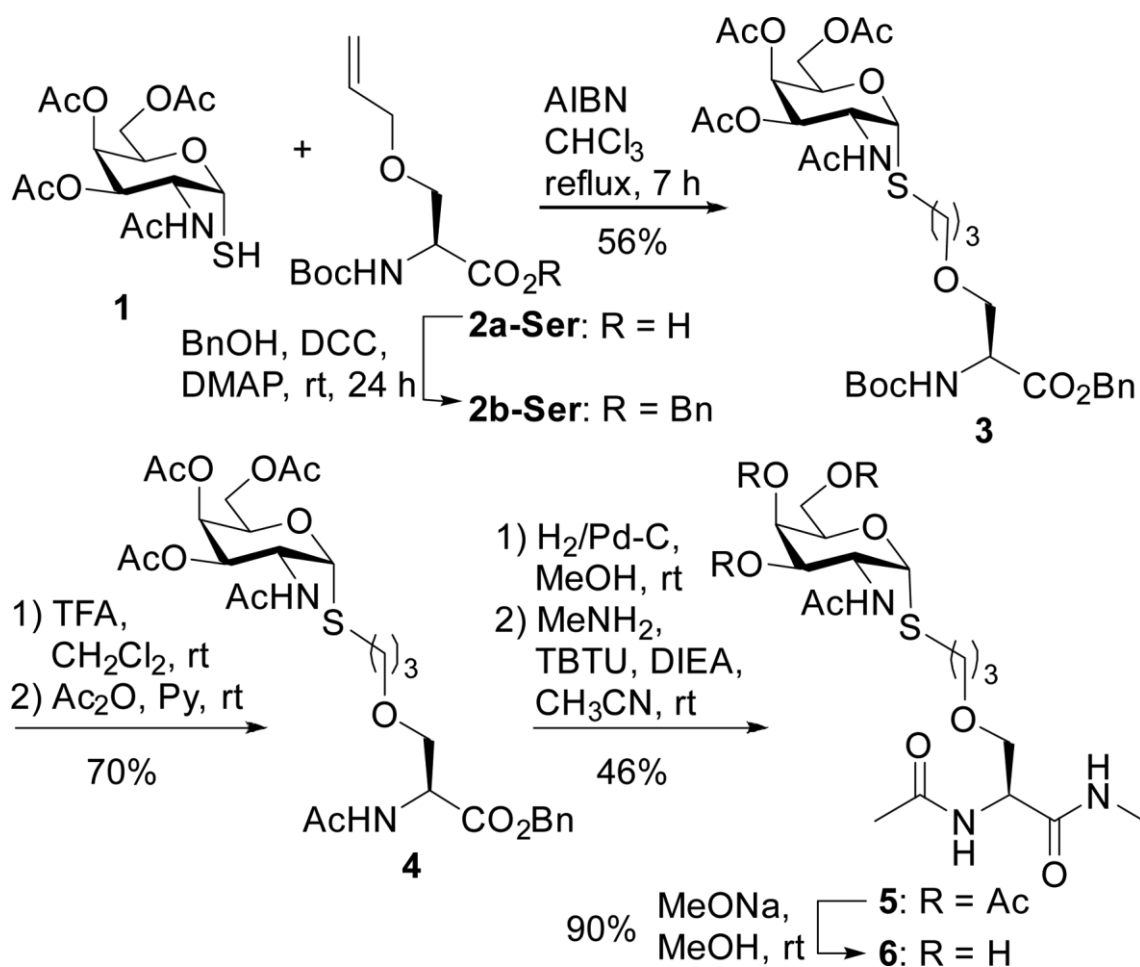
Regarding the lateral chain, derivatives carrying a homocysteine, homohomocysteine or serine (compounds 18, 19, and 20, respectively) present a flexible dihedral angle  $\chi_1$ , with two main values (Supporting Information). This result is similar to that found for other glycopeptides bearing a serine residue.<sup>49</sup> For the other derivatives,  $\chi_1$  is fixed around  $60^\circ$ . Glycopeptides 21 and 22 present a similar behavior for the side chain, being derivative 22, with a threonine amino acid, the one with the more rigid linker in solution (Supporting

Information). The result is in agreement with our previous studies on Thr-derived glycopeptides.<sup>50</sup> In general and according to our simulations, the linker of all glycopeptides is located almost perpendicularly to the peptide backbone (Figure 3 and Supporting Information).

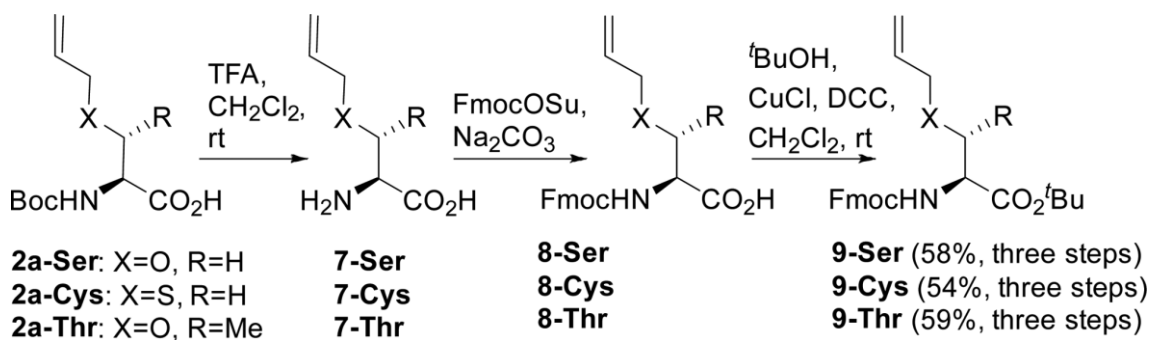
The viability of these novel compounds as mimics of MUC1 antigen was validated by the analysis of the crystal structure of one of these unnatural glycopeptides, compound 22, complexed to the anti-MUC1 monoclonal antibody SM3 produced in our lab.<sup>44</sup> The analysis of the X-ray structure (PDB id 5fxc, Figure 5a and Supporting Information) revealed that the hydrogen bonds as well as the hydrophobic contacts present in the natural glycopeptide APDT(GalNAc)RP bound to SM3<sup>44</sup> are also conserved in the scFv-SM3:22 complex. In detail, Pro2 stacks with Trp91L, Trp96L, and Tyr32L. The side chains of Asp3 and Arg5 are engaged in hydrophobic contacts with Trp33H and Tyr32H, respectively. NH group of Ala1 and carbonyl group of Thr4 are involved in hydrogen bonds with Tyr32L and Gln97H, respectively. A hydrogen bond between the side chain of Arg5 and the carbonyl group of Asn31H was also observed.

The conformation of the peptide fragment of 22 bound to the mAb is almost identical to those observed for natural glycopeptides in complex with this antibody, indicating the minor impact that the linker has on the peptide conformation in the bound state (Figure 5b). Unfortunately, the structure of the linker and the GalNAc could not be solved due to the lack of electron density. This is presumably due to the high degree of flexibility of the linker. Therefore, molecular dynamics (MD) simulations (100 ns) were performed. All interactions between glycopeptide 22 and the mAb observed in the X-ray structure are also observed through MD simulations. As deduced from these calculations, the linker is rather flexible and the GalNAc residue is exposed to the solvent (Figure 5c).

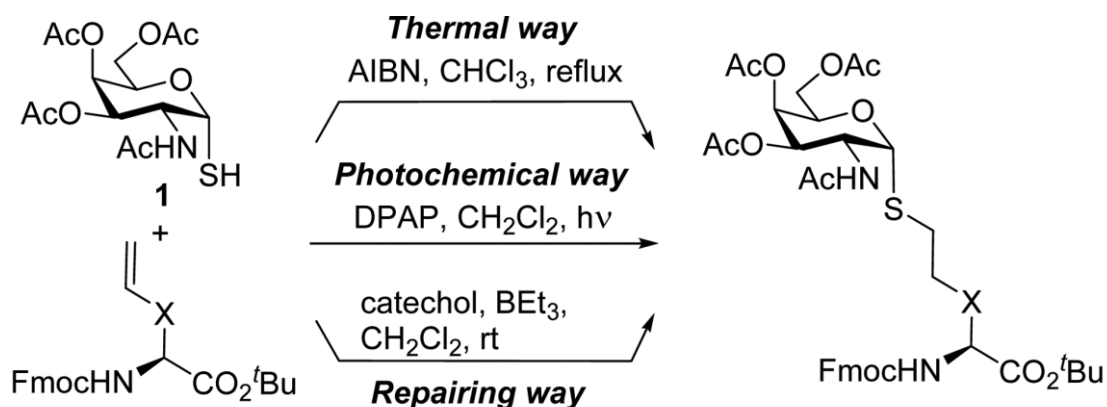
Scheme 1. Synthesis of S-Glycosyl Amino Acid 6 as a Diamide



Scheme 2. Synthesis of N-Fmoc-Ser/Cys/Thr(O- or S-allyl)-OtBu Derivatives 9-Ser/Cys/Thr



Scheme 3. Synthesis of S-Glycosyl Amino Acid Building Blocks 10-Ser/Cys/Thr and 12 via Hydrothiolation



**9-Ser** (X = -CH<sub>2</sub>OCH<sub>2</sub>-)

**9-Cys** (X = -CH<sub>2</sub>SCH<sub>2</sub>-)

**9-Thr** (X = -CH(CH<sub>3</sub>)OCH<sub>2</sub>-)

**11** (X = -CH<sub>2</sub>-)

**10-Ser** (TW: 70%, PW: 53%, RW: 70%)

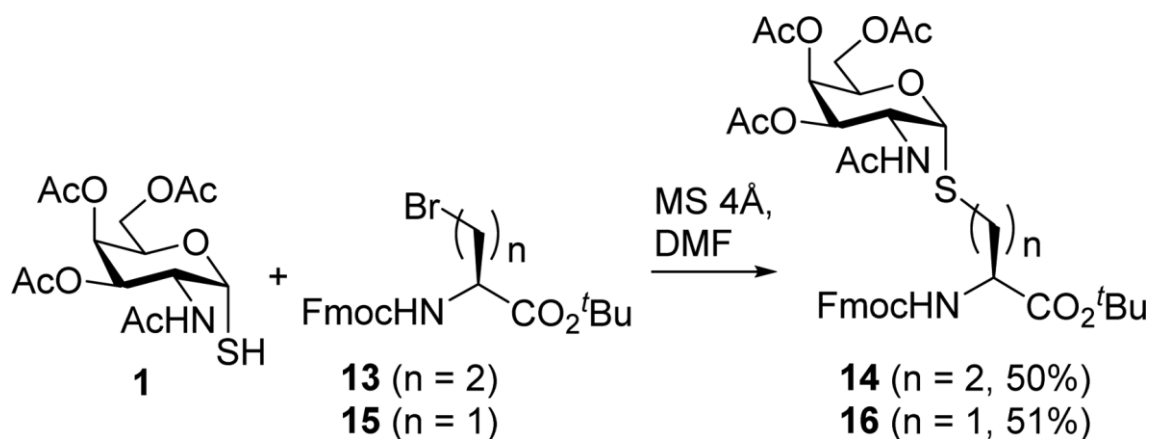
**10-Cys** (TW: 20%, PW: 30%, RW: 36%)

**10-Thr** (TW: 71%, PW: 62%, RW: 73%)

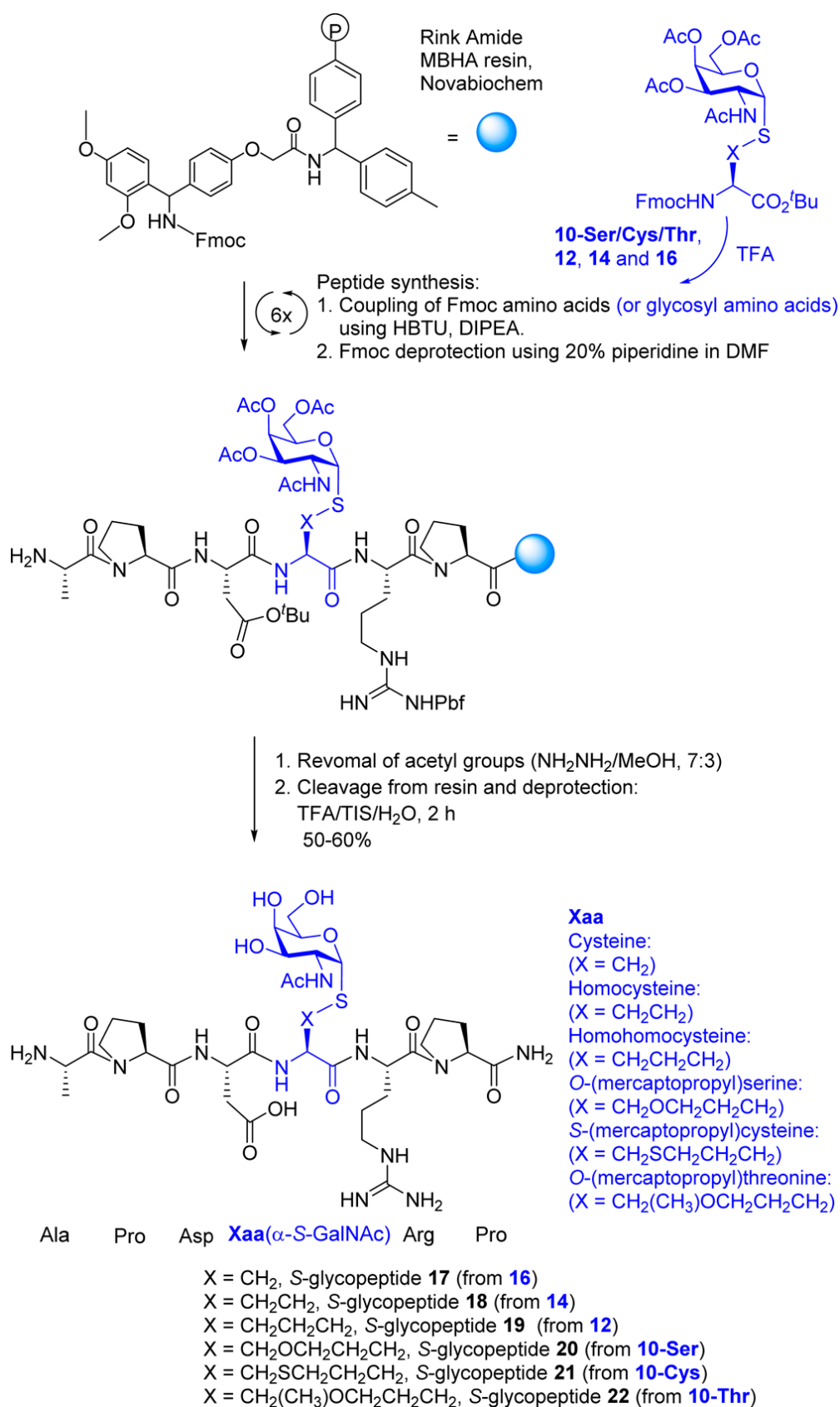
**12** (TW: 70%, PW: 50%, RW: 70%)

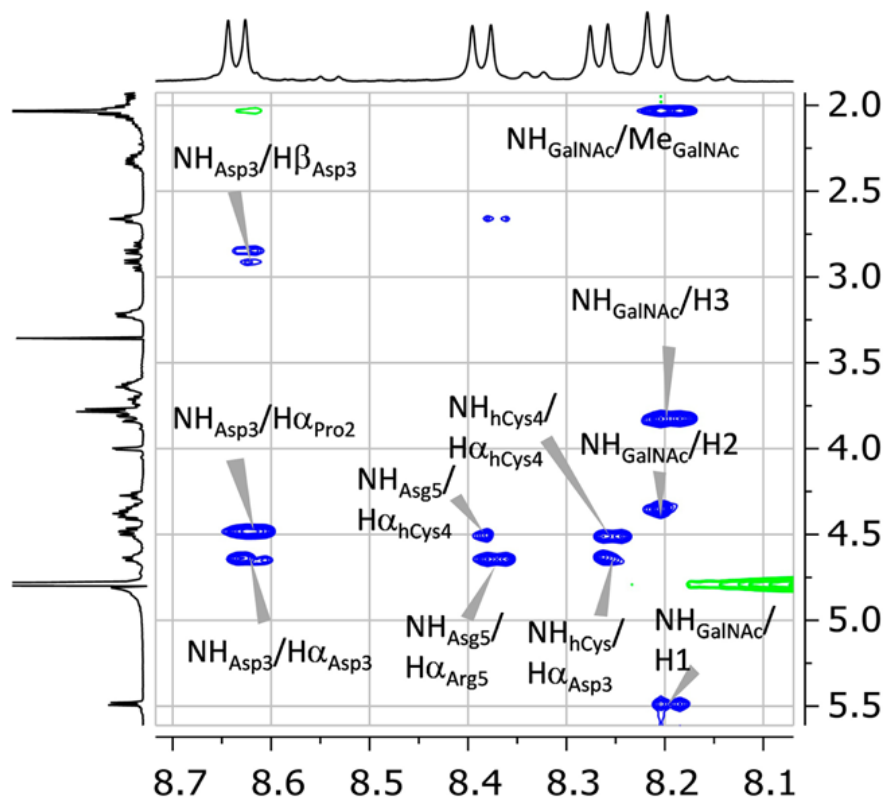
<sup>a</sup>TW, PW, and RW stand for thermal way, photochemical way, and repairing way, respectively.

Scheme 4. Synthesis of S-Glycosyl Amino Acid Building Blocks 14 and 16 via Nucleophilic Substitution



Scheme 5.  $\alpha$ -S-Neoglycopeptides Synthesized in This Work





Distance	Exp. <sup>[a]</sup>	MD-tar <sup>[a]</sup>
$\text{NH}_{\text{Asp3}} - \text{H}\alpha_{\text{Pro2}}$	2.6	2.4
$\text{NH}_{\text{Asp3}} - \text{H}\alpha_{\text{Asp3}}$	3.0	2.9
$\text{NH}_{\text{hCys}} - \text{H}\alpha_{\text{Asp3}}$	3.0	2.6
$\text{NH}_{\text{hCys}} - \text{H}\alpha_{\text{hCys}}$	3.0	2.9
$\text{NH}_{\text{Arg5}} - \text{H}\alpha_{\text{hhSer4}}$	3.0	3.0
$\text{NH}_{\text{Arg5}} - \text{H}\alpha_{\text{Arg5}}$	3.1	2.9

<sup>[a]</sup> Distances are given in Å.

Figure 2. Section of the 500 ms 2D NOESY spectrum (400 MHz) in H<sub>2</sub>O/D<sub>2</sub>O (9:1) at 25 °C of derivative 18, showing the amide region. Diagonal peaks and exchange cross-peaks connecting NH protons and water are negative (green). NOE contacts are represented as positive cross-peaks (blue). Comparison of the experimental and MD-tar-derived distances for this glycopeptide is also shown.

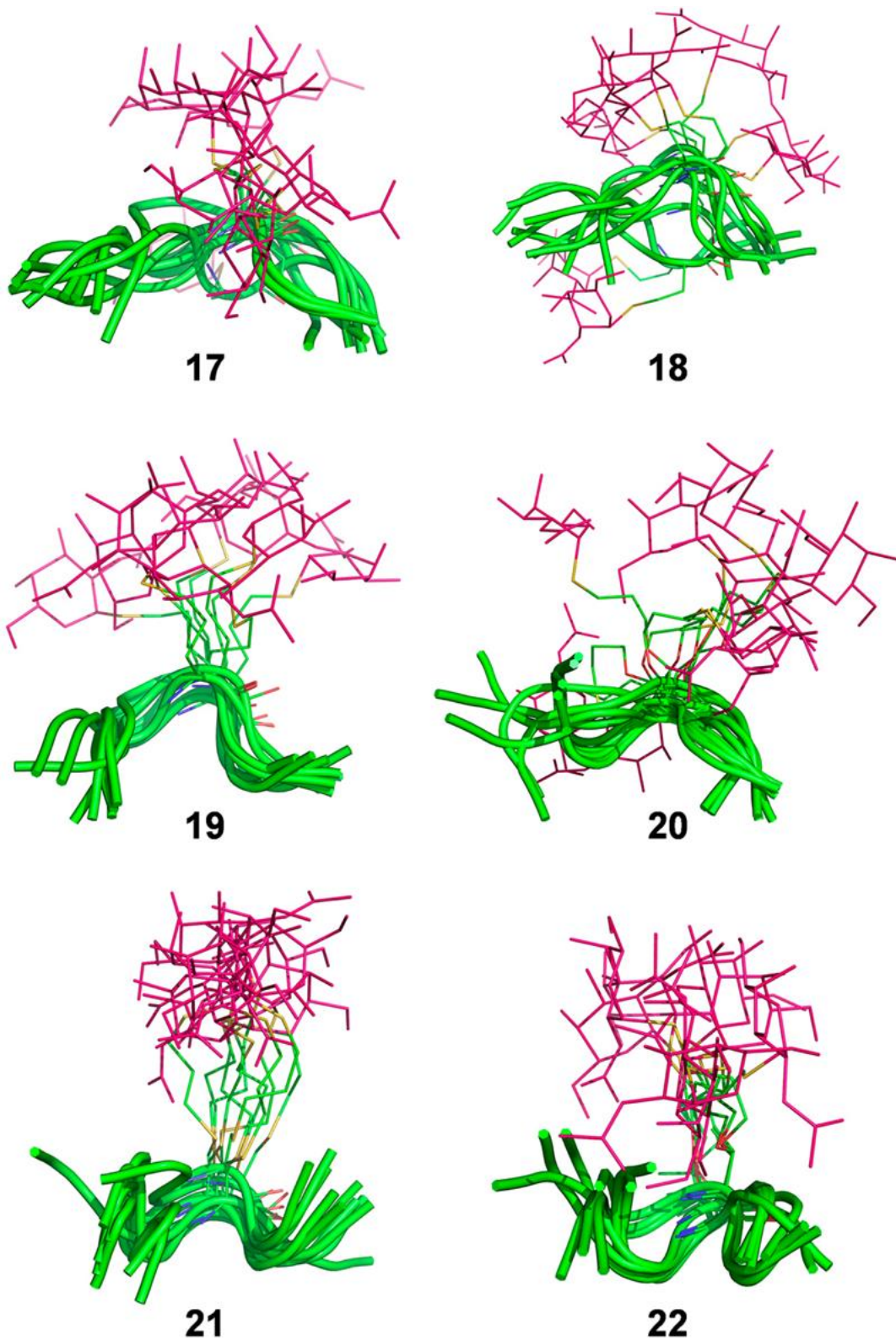


Figure 3. Ensembles obtained through 20 ns MD-tar simulations for neoglycopeptides 17–22. The backbone is shown in green and the GalNAc in purple.

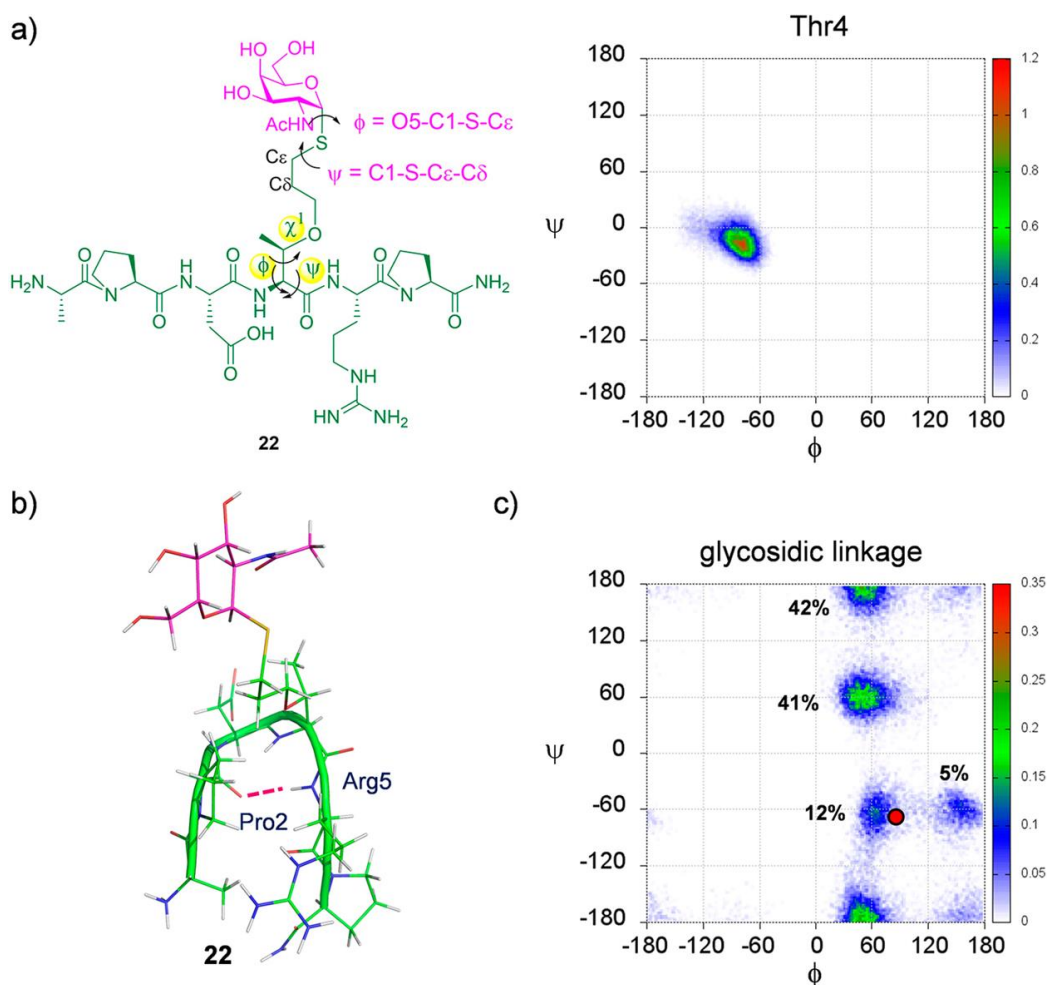


Figure 4. (a) Definition of the glycosidic linkage in this work, together with  $\phi/\psi$  distribution obtained through 20 ns MD-tar simulations for amino acid Thr4 in compound 22. (b) Type I  $\beta$ -turn conformer found in water for glycopeptides 19–22. (c)  $\phi/\psi$  distribution obtained through 20 ns MDtar simulations for the glycosidic linkage in compound 22. The conformation found in the crystal structure for the glycosidic linkage of APDC(GalNAc)RP in complex to SM3 antibody (PDB id 5A2L, ref 44) is represented as a red circle.

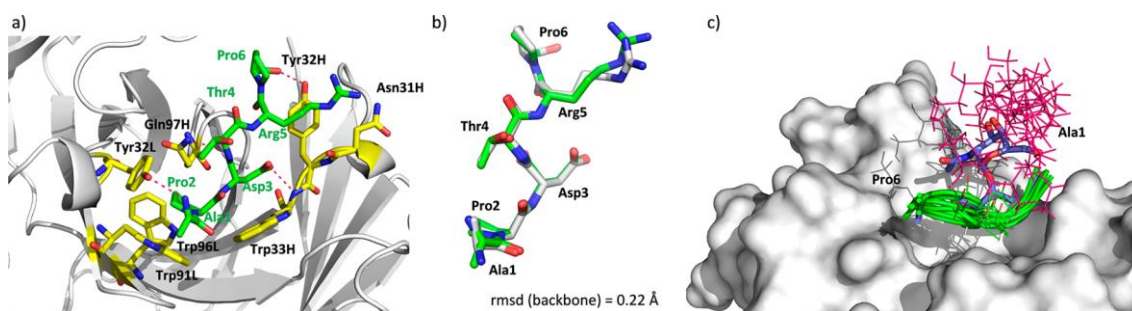


Figure 5. (a) Crystal structure of glycopeptide 22 in complex with scFv-SM3, showing the stabilizing interactions antigen–antibody. Carbon atoms of glycopeptide 22 are in green. Carbon atoms of key residues of SM3 are colored in yellow. Interactions between the peptidic fragment and SM3 surface are in pink. (b) Superposition of the peptide backbone of glycopeptide APDT(GalNAc)RP and glycopeptide 22 in complex with scFv-SM3, showing the root-mean-square deviation (rmsd) value for the peptide backbone. (c) Ensembles obtained from the unrestrained 100 ns MD simulations performed on complex scFv-SM3:22. The antibody is shown as a gray surface, and the peptide backbone of glycopeptide 22 is shown in green. Carbon atoms of GalNAc unit are in pink. The conformation of glycopeptide APDT(GalNAc)RP found in the X-ray structure is also shown in purple-blue (PDB id 5A2K, ref 44).



### 3.11. CONCLUSIONS

Several sulfa-Tn antigen derivatives with variable lengths between the glycan and the peptide backbone have been synthesized using two types of reactions with GalNAc(OAc)<sub>3</sub>- $\alpha$ -SH. The first one comprised a thiol-ene coupling of alkenyl- $\alpha$ -amino acids, and the second one involved an anomeric thiol group S-alkylation of bromo- $\alpha$ -amino acids. The presence of a linker between the peptide backbone and the sugar precludes the intramolecular interactions between these two moieties, potentially favoring the presentation of the sugar for molecular recognition events. This property, together with the potential lack of immunosuppression suffered by the natural derivatives, makes the studied compounds promising candidates for the design of novel therapeutic vaccines to treat tumors.

## EXPERIMENTAL SECTION

Reagents and General Procedures. Commercial reagents were used without further purification. Analytical thin layer chromatography (TLC) was performed on Macherey-Nagel precoated aluminum sheets with a 0.20 mm thickness of silica gel 60 with fluorescent indicator UV<sub>254</sub>. TLC plates were visualized with UV light and by staining with phosphomolybdic acid (PMA) solution (5 g of PMA in 100 mL of absolute ethanol) or sulfuric acid–ethanol solution. Column chromatography was performed on silica gel (230–400 mesh). <sup>1</sup>H and <sup>13</sup>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 experiments. All NMR chemical shifts ( $\delta$ ) were recorded in parts per million (ppm) and coupling constants (J) were reported in Hertz (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.

General Procedure for the Thiol–Ene Coupling by the Thermal Way. To a solution of alkene–amino acid derivative (1.0 equiv) and tri-O-acetyl-2-acetamido-2-deoxy-1-thio- $\alpha$ -D-galactopyranose 1 (1.7 equiv) in CHCl<sub>3</sub> (5 mL),  $\alpha$ ,  $\alpha'$ -azoisobutyronitrile (AIBN or VAZO 64, Molekula 78-67-1) (0.6 equiv) was added and the reaction mixture was then heated at reflux. After stirring for 3 h, another portion of AIBN (0.6 equiv) was added and the mixture was stirred for 4 h more. The reaction was concentrated, and the residue was purified by silica gel column chromatography to give the corresponding 5-galactosaminyl amino acid derivative.

General Procedure for the Thiol–Ene Coupling by the Photochemical Way. The reaction was carried out in a glass vial (diameter: 2 cm), sealed with a natural rubber septum, and located 2.5 cm away from the UVA lamp apparatus (LUZCHEM-LZC-4 V, 16 UVA lamps of 350 nm and 8 W). To a solution of

alkene–amino acid derivative (1.0 equiv) and tri-*O*-acetyl-2-acetamido-2-deoxy-1-thio- $\alpha$ -D-galactopyranose 1 (1.2 equiv) in CH<sub>2</sub>Cl<sub>2</sub> (2.5 mL), 2,2-dimethoxy-2-phenylacetophenone (DPAP) (0.2 equiv) was added. After stirring for 3 h, another portion of DPAP (0.2 equiv) was added and the mixture was stirred for 4 h more. The reaction was concentrated, and the residue was purified by a silica gel column chromatography to give the corresponding 5-galactosaminyl amino acid derivative.

General Procedure for the Thiol–Ene Coupling by the Repairing Way. To a solution of tri-*O*-acetyl-2-acetamido-2-deoxy-1-thio- $\alpha$ -D-galactopyranose 1 (2.0 equiv), catechol (1.2 equiv), and alkene–amino acid derivative (0.75 M in CH<sub>2</sub>Cl<sub>2</sub>, 1.0 equiv), BEt<sub>3</sub> (1.0 M in hexane, 1.2 equiv) was added. The reaction was stirred at room temperature, and BEt<sub>3</sub> was regularly added until complete consumption of the alkenyl–amino acid derivative (0.1 equiv every 45min). The reaction was then concentrated, and the residue was purified by silica gel column chromatography to give the corresponding 5-galactosaminyl amino acid derivative.

Solid-Phase (Glyco)Peptide Synthesis (SPPS). All glycopeptides were synthesized by a stepwise solid-phase peptide synthesis using the Fmoc strategy on Rink Amide MBHA resin (128 mg, 0.1 mmol, loading of the resin 0.78 mmol/g). The glycosylated amino acid building block (2.0 equiv) was 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)<sub>3</sub>GalNAc moiety were removed in a mixture of NH<sub>2</sub>NH<sub>2</sub>/MeOH (7:3). The glycopeptides were then released from the resin, and all acid-sensitive side-chain protecting groups were simultaneously removed using TFA 95%, TIS 2.5%, and H<sub>2</sub>O 2.5%, followed by precipitation with diethyl ether. Finally, all glycopeptides were purified by HPLC on a Waters Delta Prep 4000 reverse-phase HPLC and Waters 2987 Dual Absorbance Detector using a Phenomenex Luna C18(2) column (10  $\mu$ , 250 mm  $\times$  21.2 mm), 2% (v/v) CH<sub>3</sub>CN in H<sub>2</sub>O (containing 0.1% v/v TFA) gradient to 13% CH<sub>3</sub>CN (t =

27 min) and then to 60% CH<sub>3</sub>CN (t = 33 min), 10 mL/min, and registered at  $\lambda$  = 212 nm.

3,4,6-Tri-O-acetyl-2-acetamido-2-deoxy-1-thio- $\alpha$ -D-galactopyranose (1). A stirred suspension of a mixture of  $\alpha$ ,  $\beta$ -anomers of 1,3,4,6-tetra-O-acetyl-2-acetamido-2-deoxy-D-galactopyranose (20.5 g, 52.65 mmol) and Lawesson's reagent (18.1 g, 44.7 mmol) in (1:1) toluene/1,2-dichloroethane solution (200 mL) was heated at 80 °C for 12 h. The reaction was then evaporated and the crude used directly in the next reaction. This crude was dissolved in 200 mL of methanol and stirred at 0 °C. TFA (10 mL) and water (10 mL) were then added, and the reaction was allowed to warm to room temperature and stirred for 4 h. The reaction was concentrated, and the residue was purified by column chromatography using EtOAc/CH<sub>2</sub>Cl<sub>2</sub> (8:2) as eluent to give compound 1 (9.55 g, 50%) as a white foam. Spectroscopic data are consistent with the literature (refs 26b and 34a).

N-Boc-O-(S-((2-acetamido-2-deoxy)-(3,4,6-tri-O-acetyl)- $\alpha$ -D-galactopyranosyl) mercaptopropyl)-L-serine Benzyl Ester (3). A solution of N-Boc-L-Ser(O-allyl)-OBn (2b-Ser) (90 mg, 0.27 mmol), 3,4,6-tri-O-acetyl-2-acetamido-2-deoxy-1-thio- $\alpha$ -D-galactopyranose (1) (166 mg, 0.46 mmol), and AIBN (27 mg, 0.16 mmol) in CHCl<sub>3</sub> (3 mL) was heated at reflux. After stirring for 3 h, another portion of AIBN (27 mg, 0.16 mmol) was added and the reaction was stirred for 4 h more. The reaction mixture was cooled, concentrated, and then chromatographed on silica gel using EtOAc/CH<sub>2</sub>Cl<sub>2</sub> (8:2) as eluent to give compound 3 (104 mg, 56%) as a white foam.  $[\alpha]_{D 20}$  (c = 1.00, CHCl<sub>3</sub>): +71.4. HRMS (ESI+) (m/z) 699.2791 [M + H]<sup>+</sup>; calcd C<sub>32</sub>H<sub>47</sub>N<sub>2</sub>O<sub>13</sub>S<sup>+</sup> 699.2793. <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>)  $\delta$  (ppm): 7.39–7.29 (m, 5H, arom), 5.58 (d, J = 8.6 Hz, 1H, NH sugar), 5.45 (d, J = 5.1 Hz, 1H, H1), 5.41–5.34 (m, 2H, NH, H4), 5.25 (d, J = 12.4 Hz, 1H, CH<sub>2</sub>Bn), 5.13 (d, J = 12.4 Hz, 1H, CH<sub>2</sub>Bn), 5.03 (dd, J = 11.6, 2.7 Hz, 1H, H3), 4.81–4.71 (m, 1H, H2), 4.51 (t, J = 6.4 Hz, 1H, H5), 4.49–4.43 (m, 1H, H $\alpha$ ), 4.16–4.03 (m, 2H, 2H6), 3.84 (dd, J = 9.4, 2.6 Hz, 1H, H $\beta$ ), 3.64 (dd, J = 9.4, 3.2 Hz, 1H, H $\beta$ ), 3.51–3.35 (m, 2H, 2H $\delta$ ), 2.67–2.47 (m, 2H, 2H $\zeta$ ), 2.15 (s,

3H, COCH<sub>3</sub>), 2.03–1.95 (m, 9H, COCH<sub>3</sub>, COCH<sub>3</sub>, NHCOCH<sub>3</sub>), 1.87–1.71 (m, 2H, 2H ε ), 1.44 (s, 9H, C(CH<sub>3</sub>)<sub>3</sub>). <sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>) δ (ppm): 171.1, 170.7, 170.5, 170.4, 170.2, 155.6 (CO), 135.6, 128.7, 128.5, 128.3 (arom), 85.3 (C1), 80.2 (C(CH<sub>3</sub>)<sub>3</sub>), 71.0 (C β ), 69.7 (C δ ), 68.6 (C3), 67.5 (C4, C5), 67.2 (CH<sub>2Bn</sub>), 62.0 (C6), 54.3 (C α ), 48.5 (C2), 29.8 (C ε ), 28.5 (C(CH<sub>3</sub>)<sub>3</sub>), 28.0 (C ζ ), 23.5 (NHCOCH<sub>3</sub>), 20.8 (3 COCH<sub>3</sub>).

N-Acetyl-O-(S-((2-acetamido-2-deoxy)-(3,4,6-tri-O-acetyl)- $\alpha$ -D-galactopyranosyl) mercaptopropyl)-L-serine Benzyl Ester (4). Trifluoroacetic acid (378  $\mu$  L, 4.91 mmol) was added over a solution of compound 3 (104 mg, 0.15 mmol) in CH<sub>2</sub>Cl<sub>2</sub> (8 mL), the resulting solution was stirred for 2 h at room temperature, and the solvent was removed under reduced pressure. The residue was dissolved in Et<sub>2</sub>O (20 mL) and then evaporated. This operation was repeated several times to obtain the corresponding compound, which was used directly in the next reaction. Acetic anhydride (2.0 mL) and pyridine (4 mL) were added to the previous compound, and the solution was stirred for 12 h at room temperature. The solvent was removed under reduced pressure, and the residue was chromatographed on silica gel using CH<sub>2</sub>Cl<sub>2</sub>/MeOH (9:1) as eluent to give compound 4 (67 mg, 70%) as a white foam. [ $\alpha$ ]<sub>D</sub><sup>20</sup> (c = 1.01, CHCl<sub>3</sub>): +81.9. HRMS (ESI+) (m/z) 641.2382 [M + H]<sup>+</sup>; calcd C<sub>29</sub>H<sub>41</sub>N<sub>2</sub>O<sub>12</sub>S<sup>+</sup> 641.2375. <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ (ppm): 7.43–7.22 (m, 5H, arom), 6.42 (d, J = 8.0 Hz, 1H, NH), 5.57 (d, J = 8.7 Hz, 1H, NH sugar), 5.48 (d, J = 5.2 Hz, 1H, H1), 5.38 (s, 1H, H4), 5.25 (d, J = 12.3 Hz, 1H, CH<sub>2Bn</sub>), 5.16 (d, J = 12.3 Hz, 1H, CH<sub>2Bn</sub>), 5.04 (dd, J = 11.7, 2.5 Hz, 1H, H3), 4.83–4.71 (m, 2H, H α , H2), 4.51 (t, J = 6.4 Hz, 1H, H5), 4.17–4.05 (m, 2H, 2H6), 3.86 (dd, J = 9.3, 1.5 Hz, 1H, H β ), 3.69–3.63 (m, 1H, H β ), 3.52–3.46 (m, 1H, H δ ), 3.43–3.35 (m, 1H, H δ ), 2.66–2.58 (m, 1H, H ζ ), 2.55–2.48 (m, 1H, H ζ ), 2.15 (s, 3H, COCH<sub>3</sub>), 2.08–1.95 (m, 12H, COCH<sub>3</sub>, COCH<sub>3</sub>, NHCOCH<sub>3</sub>, NHCOCH<sub>3</sub>), 1.86–1.72 (m, 2H, 2H ε ). <sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>) δ (ppm): 171.1, 170.5, 170.1, 170.2, 170.3 (6 CO), 135.5, 128.7, 128.5, 128.2 (arom), 85.1 (C1), 70.7 (C β ), 69.4 (C δ ), 68.5 (C3), 67.4 (C4, C5), 67.3 (CH<sub>2Bn</sub>), 61.8 (C6), 52.8 (C α ), 48.5 (C2), 29.7 (C ε ), 27.6 (C ζ ), 23.3 (NHCOCH<sub>3</sub>), 23.2 (NHCOCH<sub>3</sub>), 20.8, 20.8 (3 COCH<sub>3</sub>).

N-Acetyl-O-(S-((2-acetamido-2-deoxy)-(3,4,6-tri-O-acetyl)- $\alpha$ -D-galactopyranosyl) mercaptopropyl)-L-serine Methyl Amide (5). A solution of compound 4 (67 mg, 0.10 mmol) in MeOH (10 mL) was treated with hydrogen under atmospheric pressure at room temperature using Pd/C 10% as a catalyst (67 mg). This reaction mixture was stirred for 24 h. The solution was then filtered over Celite and evaporated, and this crude was used directly in the next reaction. It was dissolved in dry CH<sub>3</sub>CN (20 mL) under an inert atmosphere. Methylamine hydrochloride (21 mg, 0.31 mmol), TBTU (44 mg, 0.14 mmol), and DIEA (91  $\mu$ L, 0.52 mmol) were then added, and the reaction was stirred at room temperature for 24 h. After evaporation, the residue was purified by a silica gel column chromatography, eluting with CH<sub>2</sub>Cl<sub>2</sub>/MeOH (9:1), to give compound 5 (27 mg, 46%) as a yellow oil.  $[\alpha]_{D 20}$  (c = 1.00, CHCl<sub>3</sub>): +101.4. HRMS (ESI+) (m/z) 564.2234 [M + H]<sup>+</sup>; calcd C<sub>23</sub>H<sub>38</sub>N<sub>3</sub>O<sub>11</sub>S<sup>+</sup> 564.2222. <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>)  $\delta$  (ppm): 6.71–6.60 (m, 2H, NHCOCH<sub>3</sub>, NHCH<sub>3</sub>), 6.00 (t, J = 8.4 Hz, 1H, NH sugar), 5.53 (dd, J = 8.0, 5.4 Hz, 1H, H1), 5.38–5.36 (m, 1H, H4), 5.04 (dt, J = 11.8, 3.1 Hz, 1H, H3), 4.75–4.66 (m, 1H, H2), 4.55–4.47 (m, 2H, H5, H $\alpha$ ), 4.14–4.07 (m, 2H, 2H6), 3.77 (ddd, J = 14.4, 9.4, 4.6 Hz, 1H, H $\beta$ ), 3.61–3.43 (m, 3H, H $\beta$ , 2H $\delta$ ), 2.81 (d, J = 4.8 Hz, 3H, NHCH<sub>3</sub>), 2.73–2.58 (m, 2H, 2H $\zeta$ ), 2.14 (s, 3H, COCH<sub>3</sub>), 2.06–1.95 (m, 12H, COCH<sub>3</sub>, COCH<sub>3</sub>, NHCOCH<sub>3</sub>, NHCOCH<sub>3</sub>), 1.92–1.77 (m, 2H, 2H $\epsilon$ ). <sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>)  $\delta$  (ppm) (Duplication of signals is observed, probably due to the coexistence of various conformers, since when the spectra were registered at higher temperature some of them collapse in one signal): 171.2, 170.6, 170.6, 170.5, 170.3, 170.3 (CO), 85.4, 85.4 (C1), 70.1, 70.1 (C $\beta$ ), 69.3, 69.2 (C $\delta$ ), 68.5 (C3), 67.6, 67.5, 67.4, 67.4 (C4, C5), 61.9, 61.9 (C6), 52.8, 52.8 (C $\alpha$ ), 48.8, 48.7 (C2), 29.7, 29.6 (C $\epsilon$ ), 28.0, 27.9 (C $\zeta$ ), 26.5 (NHCH<sub>3</sub>), 23.4, 23.4 (NHCOCH<sub>3</sub>), 23.3, 23.3 (NHCOCH<sub>3</sub>), 20.9, 20.8, 20.8 (3 COCH<sub>3</sub>).

N-Acetyl-O-(S-((2-acetamido-2-deoxy)- $\alpha$ -D-galactopyranosyl)-mercaptopropyl)-L-serine Methyl Amide (6). To a solution of compound 5 (27 mg, 0.05 mmol) in MeOH (2 mL), MeONa (0.5 M, 150  $\mu$ L) was added and the reaction was stirred at room temperature for 2 h. Then Dowex 50W-X8 (40 mg)

was added to the reaction, and it was stirred for 10 min. After that the mixture was filtered and evaporated obtaining compound 6 in a 90% yield. In order to perform ELLA test, the residue was dissolved in H<sub>2</sub>O (0.5 mL) and compound 6 was purified by semipreparative HPLC analysis using the following procedure: a 0.5 mL injection loop was conducted on a Waters Delta Prep 4000 reverse phase HPLC and Waters 2987 Dual Absorbance Detector using a Phenomenex Luna C18(2) column (10  $\mu$ , 250 mm  $\times$  21.2 mm), 5% (v/v) CH<sub>3</sub>CN in H<sub>2</sub>O (containing 0.1% v/v TFA) gradient to 30% CH<sub>3</sub>CN (t = 30 min), 10 mL/min,  $\lambda$  = 212 nm to furnish 6 (10 mg, 50%, t<sub>R</sub> 25.3 min) as a white foam after lyophilization.  $[\alpha]_{D_{20}}$  (c = 1.00, CHCl<sub>3</sub>) +85.0. HRMS (ESI+) (m/z) 438.1911 [M + H]<sup>+</sup>; calcd C<sub>17</sub>H<sub>32</sub>N<sub>3</sub>O<sub>8</sub>S<sup>+</sup> 438.1905. <sup>1</sup>H NMR (400 MHz, H<sub>2</sub>O/D<sub>2</sub>O 9:1)  $\delta$  (ppm) amide protons: 8.20 (d, J = 5.5 Hz, 1H, NHCOCH<sub>3</sub>), 8.13 (d, J = 7.6 Hz, 1H, NH sugar), 7.91 (s, 1H, NHCH<sub>3</sub>). <sup>1</sup>H NMR (400 MHz, D<sub>2</sub>O)  $\delta$  (ppm): 5.50 (d, J = 5.5 Hz, 1H, H1), 4.46 (t, J = 4.9 Hz, 1H, H $\alpha$ ), 4.35 (dd, J = 11.4, 5.5 Hz, 1H, H2), 4.28 (t, J = 6.1 Hz, 1H, H5), 4.00 (d, J = 2.4 Hz, 1H, H4), 3.84 (dd, J = 11.4, 3.1 Hz, 1H, H3), 3.80–3.71 (m, 4H, 2H $\beta$ , 2H $\delta$ ), 3.69–3.55 (m, 2H, 2H $\delta$ ), 2.75 (s, 3H, NHCH<sub>3</sub>), 2.72–2.61 (m, 2H, 2H $\zeta$ ), 2.07 (s, 3H, NHCOCH<sub>3</sub>), 2.05 (s, 3H, COCH<sub>3</sub>), 1.94–1.81 (m, 2H, 2H $\epsilon$ ). <sup>13</sup>C NMR (100 MHz, D<sub>2</sub>O)  $\delta$  (ppm): 174.6, 174.6, 172.2 (CO), 83.9 (C1), 71.7 (C5), 69.2 (C $\beta$ , C $\delta$ ), 68.6 (C4), 67.8 (C3), 61.3 (C6), 54.0 (C $\alpha$ ), 50.3 (C2), 28.5 (C $\epsilon$ ), 26.8 (C $\zeta$ ), 26.1 (NHCH<sub>3</sub>), 22.0 (NHCOCH<sub>3</sub>), 21.9 (NHCOCH<sub>3</sub>).

N-Fmoc-O-(allyl)-L-serine tert-Butyl Ester (9-Ser). To a solution of compound 2a-Ser (1.02 g, 4.16 mmol) in CH<sub>2</sub>Cl<sub>2</sub> (5 mL), trifluoroacetic acid (5 mL) was added. The resulting solution was stirred for 2 h at room temperature, and the solvent was removed under reduced pressure. The residue was dissolved in Et<sub>2</sub>O (10 mL) and then evaporated. This operation was repeated several times to obtain compound 7-Ser, which was used directly in the next reaction. A solution of acetone/H<sub>2</sub>O (1:1) (30 mL) was added to the previous compound, and it was then treated with Na<sub>2</sub>CO<sub>3</sub> (882 mg, 8.32 mmol) and FmocOSu (1.4 g, 4.16 mmol). The mixture was stirred for 18 h at room temperature and concentrated. The crude mixture was diluted with H<sub>2</sub>O and extracted with Et<sub>2</sub>O.

The aqueous layer was then acidified to pH 3–4 with a 10% KHSO<sub>4</sub> solution. This aqueous solution was extracted with CH<sub>2</sub>Cl<sub>2</sub> (3 × 50 mL). The combined organic layers were washed with brine, dried over Na<sub>2</sub>SO<sub>4</sub>, filtered, and concentrated to give compound 8-Ser (1.50 g, 98%) as a white solid, which was used directly in the next reaction. Butyl alcohol (1.6 mL, 16.73 mmol) was added over DCC (2.72 g, 13.07 mmol) and CuCl (40 mg, 0.41 mmol) under inert atmosphere and light protected, and the resulting mixture was stirred for 5 days, and dry CH<sub>2</sub>Cl<sub>2</sub> (6 mL) was then added. After that compound 8-Ser (1.50 g, 4.08 mmol) dissolved in dry CH<sub>2</sub>Cl<sub>2</sub> (6 mL) was added over the previous solution, and the mixture was stirred for 4 h. The dicyclohexylurea byproduct was removed by filtration over Celite, and the solution was washed with saturated NaHCO<sub>3</sub> (2 × 50 mL). The organic layer was dried, filtered, and concentrated, and the residue was purified by silica gel column chromatography using hexane/EtOAc (8:2) as eluent to give compound 9-Ser (1.0 g, 58%) as a colorless oil. [α]<sub>D</sub><sup>20</sup> (c = 1.00, CHCl<sub>3</sub>): +16.5. HRMS (ESI+) (m/z) 446.1943 [M + Na]<sub>+</sub>; calcd C<sub>25</sub>H<sub>29</sub>NNaO<sub>5</sub> + 446.1938. <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ (ppm): 7.80–7.73 (m, 2H, arom), 7.65–7.59 (m, 2H, arom), 7.44–7.36 (m, 2H, arom), 7.34–7.28 (m, 2H, arom), 5.86 (m, 1H, H ε), 5.67 (d, J = 8.3 Hz, 1H, NH), 5.27 (d, J = 17.3 Hz, 1H, H ζ trans), 5.20 (d, J = 10.4 Hz, 1H, H ζ cis), 4.47–4.33 (m, 3H, H α, CH<sub>2</sub>Fmoc), 4.24 (t, J = 7.2 Hz, 1H, CH<sub>Fmoc</sub>), 4.04 (dd, J = 12.8, 5.3 Hz, 1H, H δ), 3.96 (dd, J = 12.8, 5.4 Hz, 1H, H δ), 3.86 (dd, J = 9.4, 2.5 Hz, 1H, H β), 3.69 (dd, J = 9.3, 2.2 Hz, 1H, H β), 1.49 (s, 9H, C(CH<sub>3</sub>)<sub>3</sub>). <sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>) δ (ppm): 169.5 (CO), 156.2 (CO), 144.1, 144.0, 141.4 (arom), 134.2 (C ε), 127.8, 127.2, 125.3, 120.1 (arom), 117.5 (C ζ), 82.4 (C(CH<sub>3</sub>)<sub>3</sub>), 72.4 (C δ), 70.4 (C β), 67.3 (CH<sub>2</sub>Fmoc), 55.1 (C α), 47.3 (CH<sub>Fmoc</sub>), 28.2 (C(CH<sub>3</sub>)<sub>3</sub>).

N-Fmoc-S-(allyl)-L-cysteine tert-Butyl Ester (9-Cys). Following the same procedure above described for compound 9-Ser and starting from 2a-Cys (0.76 g, 2.91 mmol), compound 9-Cys (679 mg) was obtained as a white solid with a 54% overall yield. Mp: 61–63 °C. [α]<sub>D</sub><sup>20</sup> (c = 1.00, CHCl<sub>3</sub>): –6.0. HRMS (ESI+) (m/z) 440.1896 [M + H]<sub>+</sub>; calcd C<sub>25</sub>H<sub>30</sub>NO<sub>4</sub>S<sub>+</sub> 440.1890. <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ (ppm): 7.78–7.76 (m, 2H, arom), 7.64–7.60 (m, 2H, arom), 7.43–7.38



(m, 2H, arom), 7.35–7.29 (m, 2H, arom), 5.82–5.71 (m, 1H, H  $\epsilon$ ), 5.62 (d, J = 7.7 Hz, 1H, NH), 5.16–5.08 (m, 2H, 2H  $\zeta$ ), 4.50 (m, 1H, H  $\alpha$ ), 4.40 (d, J = 7.1 Hz, 1H, CH<sub>2</sub>Fmoc), 4.25 (t, J = 7.1 Hz, 1H, CH<sub>Fmoc</sub>), 3.18–3.15 (m, 2H, 2H  $\delta$ ), 2.97 (dd, J = 13.8, 4.5 Hz, 1H, H  $\beta$ ), 2.87 (dd, J = 13.8, 5.6 Hz, 1H, H  $\beta$ ), 1.50 (s, 9H, C(CH<sub>3</sub>)<sub>3</sub>). <sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>)  $\delta$  (ppm): 169.9 (CO), 155.8 (CO), 144.0, 143.9, 141.4 (arom), 133.9 (C  $\epsilon$ ), 127.8, 127.2, 125.2, 120.1 (arom), 117.9 (C  $\zeta$ ), 82.9 (C(CH<sub>3</sub>)<sub>3</sub>), 67.2 (CH<sub>2</sub>Fmoc), 54.2 (C  $\alpha$ ), 47.2 (CH<sub>Fmoc</sub>), 35.5 (C  $\delta$ ), 33.3 (C  $\beta$ ), 28.1 (C(CH<sub>3</sub>)<sub>3</sub>).

N-Fmoc-O-(allyl)-L-threonine tert-Butyl Ester (9-Thr). Following the same procedure above described for compound 9-Ser and starting from 2a-Thr (2.24 g, 8.64 mmol), compound 9-Thr (2.03 g) was obtained as a white solid with a 59% overall yield. Mp: 58–60 °C. [ $\alpha$ ]<sub>D</sub><sup>20</sup> (c = 1.00, CHCl<sub>3</sub>): –0.3. HRMS (ESI+) (m/z) 438.2292 [M + H]<sup>+</sup>; calcd C<sub>26</sub>H<sub>32</sub>NO<sub>5</sub> + 438.2275. <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>)  $\delta$  (ppm): 7.78–7.74 (m, 2H, arom), 7.68–7.63 (m, 2H, arom), 7.43–7.38 (m, 2H, arom), 7.35–7.30 (m, 2H, arom), 5.94–5.82 (m, 1H, H  $\epsilon$ ), 5.60 (d, J = 9.6 Hz, 1H, NH), 5.29 (d, J = 17.0 Hz, 1H, H  $\zeta$  trans), 5.19 (d, J = 10.4 Hz, 1H, H  $\zeta$  cis), 4.48–4.38 (m, 2H, CH<sub>2</sub>Fmoc), 4.34 (dd, J = 9.6, 1.3 Hz, 1H, H  $\alpha$ ), 4.26 (t, J = 7.2 Hz, 1H, CH<sub>Fmoc</sub>), 4.14–4.06 (m, 2H, H  $\beta$ , H  $\delta$ ), 3.91 (dd, J = 12.6, 5.5 Hz, 1H, H  $\delta$ ), 1.51 (s, 9H, C(CH<sub>3</sub>)<sub>3</sub>), 1.25 (d, J = 6.2 Hz, 3H, CH<sub>3</sub>  $\beta$ ). <sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>)  $\delta$  (ppm): 169.6 (CO), 156.7 (CO), 144.0, 143.8, 141.2 (arom), 134.5 (C  $\epsilon$ ), 127.6, 127.0, 127.0, 125.1, 125.1, 119.9 (arom), 116.9 (C  $\zeta$ ), 81.8 (C(CH<sub>3</sub>)<sub>3</sub>), 74.7 (C  $\beta$ ), 70.0 (C  $\delta$ ), 67.1 (CH<sub>2</sub>Fmoc), 59.1 (C  $\alpha$ ), 47.1 (CH<sub>Fmoc</sub>), 28.0 (C(CH<sub>3</sub>)<sub>3</sub>), 16.2 (CH<sub>3</sub>  $\beta$ ).

N-Fmoc-O-(S-((2-acetamido-2-deoxy)-(3,4,6-tri-O-acetyl)- $\alpha$ -D-galactopyranosyl) mercaptopropyl)-L-serine tert-Butyl Ester (10-Ser). Thermal Way. A solution of 9-Ser (195 mg, 0.46 mmol), compound 1 (284 mg, 0.78 mmol), and AIBN (46 mg, 0.28 mmol) in CHCl<sub>3</sub> (5 mL) was heated at reflux. After stirring for 3 h, another portion of AIBN (46 mg, 0.28 mmol) was added and the reaction was stirred for 4 h more. The reaction mixture was cooled, concentrated, and then chromatographed on silica gel using EtOAc/CH<sub>2</sub>Cl<sub>2</sub> (8:2)

as eluent to give protected glycosyl amino acid 10-Ser (253 mg, 70%) as a white solid.

**Photochemical Way.** A solution of compound 9-Ser (232 mg, 0.55 mmol), compound 1 (240 mg, 0.66 mmol), and DPAP (30 mg, 0.11 mmol) in CH<sub>2</sub>Cl<sub>2</sub> (2.5 mL) was irradiated at  $\lambda_{\max}$  365 nm. After stirring for 3 h, another portion of DPAP (30 mg, 0.11 mmol) was added and the reaction was stirred for 4 h more. The reaction mixture was then concentrated, and the residue was chromatographed on silica gel using EtOAc/CH<sub>2</sub>Cl<sub>2</sub> (8:2) as eluent to give protected glycosyl amino acid 10-Ser (227 mg, 53%) as a white solid.

**Repairing Way.** A solution of compound 9-Ser (205 mg, 0.48 mmol), compound 1 (352 mg, 0.97 mmol), catechol (64 mg, 0.58 mmol), and BEt<sub>3</sub> (84  $\mu$ L, 0.58 mmol) in CH<sub>2</sub>Cl<sub>2</sub> (0.65 mL) was stirred at room temperature. After stirring for 1 h, another portion of BEt<sub>3</sub> (7  $\mu$ L) was added (until complete consumption of compound 9-Ser). The reaction mixture was then concentrated, and the residue was chromatographed on silica gel using EtOAc/CH<sub>2</sub>Cl<sub>2</sub> (8:2) as eluent to give protected glycosyl amino acid 10-Ser (265 mg, 70%) as a white solid.

**Physical data:** mp 63–65 °C.  $[\alpha]_{\text{D}}^{20}$  (c = 1.00, CHCl<sub>3</sub>): +95.5. HRMS (ESI+) (m/z) 787.3119 [M + H]<sup>+</sup>; calcd C<sub>39</sub>H<sub>51</sub>N<sub>2</sub>O<sub>13</sub>S<sup>+</sup> 787.3106. <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>)  $\delta$  (ppm): 7.78–7.74 (m, 2H, arom), 7.65–7.59 (m, 2H, arom), 7.42–7.37 (m, 2H, arom), 7.34–7.29 (m, 2H, arom), 5.70–5.60 (m, 2H, NHAc, NHFmoc), 5.49 (d, J = 5.2 Hz, 1H, H1), 5.38 (s, 1H, H4), 5.04 (dd, J = 11.6, 2.6 Hz, 1H, H3), 4.82–4.72 (m, 1H, H2), 4.53 (t, J = 6.3 Hz, 1H, H5), 4.46–4.33 (m, 3H, CH<sub>2</sub>Fmoc, H $\alpha$ ), 4.25 (t, J = 7.1 Hz, 1H, CH<sub>Fmoc</sub>), 4.17–4.03 (m, 2H, 2H6), 3.82 (dd, J = 9.4, 2.8 Hz, 1H, H $\beta$ ), 3.67 (dd, J = 9.3, 2.3 Hz, 1H, H $\beta$ ), 3.60–3.43 (m, 2H, 2H $\delta$ ), 2.75–2.57 (m, 2H, 2H $\zeta$ ), 2.14 (s, 3H, COCH<sub>3</sub>), 2.06–1.95 (m, 9H, 2 COCH<sub>3</sub>, NHCOCH<sub>3</sub>), 1.91–1.83 (m, 2H, 2H $\epsilon$ ), 1.48 (s, 9H, C(CH<sub>3</sub>)<sub>3</sub>). <sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>)  $\delta$  (ppm): 171.1, 170.4, 170.4, 170.2, 169.4, 156.1 (6 CO), 144.0, 144.0, 141.4, 127.8, 127.2, 125.3, 120.0 (arom), 85.2 (C1), 82.4 (C(CH<sub>3</sub>)<sub>3</sub>), 71.2 (C $\beta$ ), 69.7 (C $\delta$ ), 68.6 (C3), 67.5, 67.4 (C4, C5), 67.2 (CH<sub>2</sub>Fmoc), 61.9 (C6),

55.0 (C  $\alpha$ ), 48.5 (C2), 47.3 (CH<sub>Fmoc</sub>), 29.8 (C  $\epsilon$ ), 28.1 (C(CH<sub>3</sub>)<sub>3</sub>), 28.0 (C  $\zeta$ ), 23.4 (NHCOCH<sub>3</sub>), 20.8, 20.8 (3 COCH<sub>3</sub>).

N-Fmoc-S-(S-((2-acetamido-2-deoxy)-(3,4,6-tri-O-acetyl)- $\alpha$ -D-galactopyranosyl) mercaptopropyl)-L-cysteine tert-Butyl Ester (10-Cys). Thermal Way. Following the same procedure above described for compound 10-Ser and starting from 9-Cys (191 mg, 0.43 mmol), compound 10-Cys (70 mg, 20%) was obtained as a white solid.

Photochemical Way. Following the same procedure above described for compound 10-Ser and starting from 9-Cys (203 mg, 0.46 mmol), compound 10-Cys (111 mg, 30%) was obtained as a white solid. Repairing Way. Following the same procedure above described for compound 10-Ser and starting from 9-Cys (310 mg, 0.71 mmol), compound 10-Cys (201 mg, 36%) was obtained as a white solid. Physical data: mp 69–71 °C.  $[\alpha]_{D_{20}}$  (c = 1.00, CHCl<sub>3</sub>): +82.9. HRMS (ESI+) (m/z) 803.2878 [M + H]<sub>+</sub>; calcd C<sub>39</sub>H<sub>51</sub>N<sub>2</sub>O<sub>12</sub>S<sub>2</sub> + 803.2878. <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>)  $\delta$  (ppm): 7.77–7.74 (m, 2H, arom), 7.62–7.58 (m, 2H, arom), 7.41–7.36 (m, 2H, arom), 7.33–7.28 (m, 2H, arom), 5.69 (d, J = 8.0 Hz, 2H, NHAc, NHFmoc), 5.49 (d, J = 5.3 Hz, 1H, H1), 5.36 (d, J = 2.4 Hz, 1H, H4), 5.02 (dd, J = 11.7, 3.1 Hz, 1H, H3), 4.79–4.71 (m, 1H, H2), 4.54–4.44 (m, 2H, H5, H  $\alpha$ ), 4.43–4.34 (m, 2H, CH<sub>2</sub>Fmoc), 4.22 (t, J = 7.0 Hz, 1H, CH<sub>Fmoc</sub>), 4.15–4.03 (m, 2H, 2H6), 2.99 (dd, J = 13.7, 4.8 Hz, 1H, H  $\beta$ ), 2.92 (dd, J = 13.7, 5.1 Hz, 1H, H  $\beta$ ), 2.75–2.13 (m, 4H, 2H  $\zeta$ , 2H  $\delta$ ), 2.13 (s, 3H, COCH<sub>3</sub>), 2.02–1.93 (m, 9H, 2 COCH<sub>3</sub>, NHCOCH<sub>3</sub>), 1.91–1.82 (m, 2H, 2H  $\epsilon$ ), 1.48 (s, 9H, C(CH<sub>3</sub>)<sub>3</sub>). <sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>)  $\delta$  (ppm): 171.0, 170.4, 170.3, 170.2, 169.8, 155.8 (6 CO), 143.9, 143.9, 141.4, 127.8, 127.2, 125.2, 120.1 (arom), 85.2 (C1), 83.0 (C(CH<sub>3</sub>)<sub>3</sub>), 68.6 (C3), 67.5 (C5), 67.4 (C4), 67.2 (CH<sub>2</sub>Fmoc), 61.9 (C6), 54.4 (C  $\alpha$ ), 48.4 (C2), 47.2 (CH<sub>Fmoc</sub>), 34.8 (C  $\beta$ ), 31.7 (C  $\delta$ ), 29.9 (C  $\zeta$ ), 29.3 (C  $\epsilon$ ), 28.1 (C(CH<sub>3</sub>)<sub>3</sub>), 23.4 (NHCOCH<sub>3</sub>), 20.8, 20.8 (3 COCH<sub>3</sub>).

The MS spectrum of the crude reaction mixture, before being purified by column chromatography, showed signals corresponding to the expected product 10-Cys and side products named 24, 26, 27, 28, and 29 that probably are formed

by radical couplings as shown in the scheme shown in the Supporting Information. This feature may explain the low yield obtained in this reaction, following the three methodologies, when it is compared with the formation of 10-Ser, 10-Thr, or 12.

N-Fmoc-O-(S-((2-acetamido-2-deoxy)-(3,4,6-tri-O-acetyl)- $\alpha$ -D-galactopyranosyl) mercaptopropyl)-L-threonine tert-Butyl Ester (10-Thr). Thermal Way. Following the same procedure above described for compound 10-Ser and starting from 9-Thr (206 mg, 0.47 mmol), compound 10-Thr (268 mg, 71%) was obtained as a white solid.

Photochemical Way. Following the same procedure above described for compound 10-Ser and starting from 9-Thr (160 mg, 0.37 mmol), compound 10-Thr (181 mg, 62%) was obtained as a white solid.

Repairing Way. Following the same procedure above described for compound 10-Ser and starting from 9-Thr (242 mg, 0.55 mmol), compound 10-Thr (322 mg, 73%) was obtained as a white solid. Physical data: mp 76–78 °C. [ $\alpha$ ]<sub>D20</sub>(c = 1.00, CHCl<sub>3</sub>): +78.3.

HRMS (ESI+) (m/z) 801.3271 [M + H]<sup>+</sup>; calcd C<sub>40</sub>H<sub>53</sub>N<sub>2</sub>O<sub>13</sub>S+801.3263. <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>)  $\delta$  (ppm): 7.76–7.73 (m, 2H, arom), 7.64–7.59 (m, 2H, arom), 7.40–7.35 (m, 2H, arom), 7.33–7.27 (m, 2H, arom), 5.75 (d, J = 8.7 Hz, 1H, NHAc), 5.54–5.46 (m, 2H, H1, NHFmoc), 5.37 (s, 1H, H4), 5.04 (dd, J = 11.7, 2.7 Hz, 1H, H3), 4.71–4.82 (m, 1H, H2), 4.53 (t, J = 6.4 Hz, 1H, H5), 4.43–4.34 (m, 2H, CH<sub>2</sub>Fmoc), 4.20–4.19 (m, 2H, H $\alpha$ , CH<sub>Fmoc</sub>), 4.14–4.04 (m, 2H, 2H6), 3.98 (d, J = 6.1 Hz, 1H, H $\beta$ ), 3.62–3.54 (m, 1H, H $\delta$ ), 3.41–3.30 (m, 1H, H $\delta$ ), 2.74–2.55 (m, 2H, 2H $\zeta$ ), 2.12 (s, 3H, COCH<sub>3</sub>), 2.02–1.95 (m, 9H, 3 COCH<sub>3</sub>, NHCOCH<sub>3</sub>), 1.89–1.78 (m, 2H, 2H $\epsilon$ ), 1.46 (s, 9H, C(CH<sub>3</sub>)<sub>3</sub>), 1.18 (d, J = 6.2 Hz, 3H, 3H $\gamma$ ). <sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>)  $\delta$  (ppm): 171.0, 170.4, 170.3, 170.2, 169.8, 156.8 (6 CO), 144.0, 143.9, 141.4, 127.8, 127.1, 125.2, 120.0 (arom), 85.1 (C1), 82.1 (C(CH<sub>3</sub>)<sub>3</sub>), 75.6 (C $\beta$ ), 68.6 (C3), 67.4 (C4, C5, C $\delta$ ), 67.2 (CH<sub>2</sub>Fmoc),

61.9 (C6), 59.1 (C $\alpha$ ), 48.4 (C2), 47.3 (CH<sub>Fmoc</sub>), 30.2 (C $\epsilon$ ), 28.2 (C(CH<sub>3</sub>)<sub>3</sub>), 28.0 (C $\zeta$ ), 23.3 (NHCOCH<sub>3</sub>), 20.8, 20.8 (3 COCH<sub>3</sub>), 16.4 (C $\gamma$ ).

N-Fmoc-S-((2-acetamido-2-deoxy)-(3,4,6-tri-O-acetyl)- $\alpha$ -D-galactopyranosyl)-L-homohomocysteine tert-Butyl Ester (12). Thermal Way. A solution of N-Fmoc-L-allylglycine tert-butyl ester 11 (100 mg, 0.25 mmol), compound 1 (157 mg, 0.43 mmol), and AIBN (25 mg, 0.15 mmol) in CHCl<sub>3</sub> (2.5 mL) was heated at reflux. After stirring for 3 h, another portion of AIBN (25 mg, 0.15 mmol) was added and the reaction was stirred for 4 h more. The reaction mixture was cooled, concentrated, and then chromatographed on silica gel using EtOAc/CH<sub>2</sub>Cl<sub>2</sub> (8:2) as eluent to give protected glycosyl amino acid 12 (140 mg, 70%) as a white solid.

Photochemical Way. A solution of compound 11 (142 mg, 0.36 mmol), compound 1 (156 mg, 0.43 mmol), and DPAP (19 mg, 0.07 mmol) in CH<sub>2</sub>Cl<sub>2</sub> (2.5 mL) was irradiated at  $\lambda_{\max}$  365 nm. After stirring for 3 h, another portion of DPAP (19 mg, 0.07 mmol) was added and the reaction was stirred for an additional time (4 h). The reaction mixture was concentrated, and the residue was chromatographed on silica gel using EtOAc/CH<sub>2</sub>Cl<sub>2</sub> (8:2) as eluent to give protected glycosyl amino acid 12 (137 mg, 50%) as a white solid.

Repairing Way. A solution of compound 11 (214 mg, 0.54 mmol), compound 1 (395 mg, 1.09 mmol), catechol (72 mg, 0.65 mmol), and BEt<sub>3</sub> (95  $\mu$  L, 0.65 mmol) in CH<sub>2</sub>Cl<sub>2</sub> (0.72 mL) was stirred at room temperature. After stirring for 1 h, another portion of BEt<sub>3</sub> (8  $\mu$  L, 0.05 mmol) was added (until complete consumption of compound 11).

The reaction mixture was concentrated, and the residue chromatographed on silica gel using EtOAc/CH<sub>2</sub>Cl<sub>2</sub> (7:3) as eluent to give protected glycosyl amino acid 12 (288 mg, 70%) as a white solid. Physical data: mp 59–61 °C. [ $\alpha$ ]<sub>D20</sub> (c = 1.00, CHCl<sub>3</sub>): + 85.5.

HRMS (ESI+) (m/z) 757.3012 [M + H]<sub>+</sub>; calcd C<sub>38</sub>H<sub>49</sub>N<sub>2</sub>O<sub>12</sub>S<sub>+</sub> 757.3001. <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>)  $\delta$  (ppm): 7.76–7.71 (m, 2H, arom), 7.61–7.56 (m, 2H,

arom), 7.40–7.30 (m, 2H, arom), 7.33–7.27 (m, 2H, arom), 5.80 (d,  $J = 8.7$  Hz, 1H, NHAc), 5.54–5.48 (m, 2H, NHFmoc, H1), 5.35 (s, 1H, H4), 5.03 (dd,  $J = 11.6, 2.3$  Hz, 1H, H3), 4.79–4.71 (m, 1H, H2), 4.51 (t,  $J = 6.4$  Hz, 1H, H5), 4.37 (d,  $J = 6.9$  Hz, 2H, CH<sub>2</sub>Fmoc), 4.29–4.18 (m, 2H, CH<sub>Fmoc</sub>, H $\alpha$ ), 4.13–4.03 (m, 2H, H6), 2.66–2.54 (m, 2H, 2H  $\delta$ ), 2.12 (s, 3H, COCH<sub>3</sub>), 2.01–1.85 (m, 10H, 2 COCH<sub>3</sub>, NHCOCH<sub>3</sub>, H $\beta$ ), 1.78–1.59 (m, 3H, H $\beta$ , 2H  $\gamma$ ), 1.46 (s, 9H, C(CH<sub>3</sub>)<sub>3</sub>). <sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>)  $\delta$  (ppm): 171.4, 171.2, 170.5, 170.4, 170.2, 156.1 (6 CO), 144.0, 141.4, 127.9, 127.2, 125.2, 120.1 (arom), 85.3 (C1), 82.6 (C(CH<sub>3</sub>)<sub>3</sub>), 68.7 (C3), 67.4, 67.4 (C4, C5), 67.1 (CH<sub>2</sub>Fmoc), 61.8 (C6), 53.6 (C $\alpha$ ), 48.5 (C2), 47.3 (CH<sub>Fmoc</sub>), 32.2 (C $\beta$ ), 30.9 (C $\delta$ ), 28.7 (C(CH<sub>3</sub>)<sub>3</sub>), 25.7 (C $\gamma$ ), 23.5 (NHCOCH<sub>3</sub>), 20.9, 20.8 (3 COCH<sub>3</sub>).

N-Fmoc-S-((2-acetamido-2-deoxy)-(3,4,6-tri-O-acetyl)- $\alpha$ -D-galactopyranosyl)-L-homocysteine tert-butyl ester (14). Under an inert atmosphere, a solution of compound 1 (1.21 g, 3.36 mmol) in dry DMF (6 mL) was transferred into a Schlenk, which contains 4 Å molecular sieves, previously activated at 280 °C for 4 h under vacuum. After stirring for 5 min, a solution of compound 13 (500 mg, 1.12 mmol) in dry DMF (12 mL) was dropped into the Schlenk. The reaction was stirred for 24 h at room temperature and filtered over Celite (to discard the molecular sieves), and the solvent was concentrated. The residue was purified by gradient column chromatography on silica gel using the following set of proportions: EtOAc/hexane 1:1, 6:4, 7:3) to give protected glycosyl amino acid 14 (416 mg, 50%) as a white solid. Mp: 54–56 °C. [ $\alpha$ ]<sub>D</sub><sup>20</sup> (c = 1.00, CHCl<sub>3</sub>): +68.2. HRMS (ESI+) (m/z) 743.2844 [M + H]<sup>+</sup>; calcd C<sub>37</sub>H<sub>47</sub>N<sub>2</sub>O<sub>12</sub>S<sup>+</sup> 743.2846. <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>)  $\delta$  (ppm): 7.77–7.75 (m, 2H, arom), 7.60–7.58 (m, 2H, arom), 7.42–7.38 (m, 2H, arom), 7.33–7.30 (m, 2H, arom), 5.56 (d,  $J = 8.7$  Hz, 1H, NHAc), 5.48 (d,  $J = 5.0$  Hz, 1H, H1), 5.43 (d,  $J = 7.9$  Hz, 1H, NHFmoc), 5.39–5.38 (m, 1H, H4), 5.06–5.02 (m, 1H, H3), 4.82–4.75 (m, 1H, H2), 4.56–4.53 (m, 1H, H5), 4.40 (d,  $J = 6.7$ , 2H, CH<sub>2</sub>Fmoc), 4.35–4.31 (m, 1H, H $\alpha$ ), 4.22 (t,  $J = 6.2$  Hz, 1H, CH<sub>Fmoc</sub>), 4.17–4.04 (m, 2H, 2H6), 2.74–2.55 (m, 2H, 2H  $\gamma$ ), 2.16–1.97 (m, 14H, 3 COCH<sub>3</sub>, NHCOCH<sub>3</sub>, 2H  $\beta$ ), 1.48 (s, 9H, C(CH<sub>3</sub>)<sub>3</sub>). <sup>13</sup>C

NMR (100 MHz, CDCl<sub>3</sub>)  $\delta$  (ppm): 171.0, 170.4, 170.3, 170.2, 167.8, 155.9 (6 CO), 143.9, 143.8, 141.3, 127.8, 127.1, 125.0, 120.0 (arom), 85.7 (C1), 82.9 (C(CH<sub>3</sub>)<sub>3</sub>), 68.5 (C3), 67.5 (C5), 67.3 (C4), 67.0 (CH<sub>2</sub>Fmoc), 61.7 (C6), 53.6 (C $\alpha$ ), 48.3 (C2), 47.2 (CH<sub>Fmoc</sub>), 33.5 (C $\beta$ ), 28.0 (C(CH<sub>3</sub>)<sub>3</sub>), 27.5 (C $\gamma$ ), 23.3 (NHCOCH<sub>3</sub>), 20.7, 20.7, 20.7 (3 COCH<sub>3</sub>).

N-Fmoc-S-((2-acetamido-2-deoxy)-(3,4,6-tri-O-acetyl)- $\alpha$ -D-galactopyranosyl)-L-cysteine tert-Butyl Ester (16). Under an inert atmosphere, a solution of compound 1 (1.01 g, 2.784 mmol) in dry DMF (6 mL) was transferred into a Schlenk, which contains 4 Å molecular sieves, previously activated at 280 °C for 4 h under vacuum. After stirring for 5 min, a solution of compound 15 (400 mg, 0.928 mmol) in dry DMF (12 mL) was dropped into the Schlenk. The reaction was stirred for 24 h at room temperature and filtered over Celite (to discard the molecular sieves), and the solvent was concentrated. The residue was purified by gradient column chromatography on silica gel using the following set of proportions: EtOAc/hexane (1:1, 6:4, 7:3) to give protected glycosyl amino acid 16 (330 mg, 51%) as a white solid. Mp: 59–61 °C. [ $\alpha$ ]<sub>D20</sub> (c = 1.00, CHCl<sub>3</sub>): +72.8. HRMS (ESI+) (m/z) 729.2686 [M + H]<sup>+</sup>; calcd C<sub>36</sub>H<sub>45</sub>N<sub>2</sub>O<sub>12</sub>S<sup>+</sup> 729.2688. <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>)  $\delta$  (ppm): 7.77–7.75 (m, 2H, arom), 7.64–7.62 (m, 2H, arom), 7.42–7.29 (m, 4H, arom), 6.10 (d, J = 8.5 Hz, 1H, NHAc), 5.65 (d, J = 8.9 Hz, 1H, NHFmoc), 5.38–5.37 (m, 2H, H1, H4), 4.98–4.94 (m, 1H, H3), 4.85–4.78 (m, 1H, H2), 4.59–4.55 (m, 1H, H $\alpha$ ), 4.48–4.38 (m, 3H, H5, CH<sub>2</sub>Fmoc), 4.23–4.19 (m, 2H, H6, CH<sub>Fmoc</sub>), 4.02–3.98 (m, 1H, H6), 3.29–3.20 (m, 1H, H $\beta$ ), 3.05–2.98 (m, 1H, H $\beta$ ), 2.18 (s, 3H, COCH<sub>3</sub>), 2.01 (s, 3H, NHCOCH<sub>3</sub>), 1.99 (s, 3H, COCH<sub>3</sub>), 1.97 (s, 3H, COCH<sub>3</sub>), 1.47 (s, 9H, C(CH<sub>3</sub>)<sub>3</sub>). <sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>)  $\delta$  (ppm): 171.0, 170.5, 170.2, 170.0, 168.8, 155.8 (6 CO), 143.9, 143.8, 141.4, 127.7, 127.1, 125.1, 120.1 (arom), 87.3 (C1), 82.9 (C(CH<sub>3</sub>)<sub>3</sub>), 68.3 (C3), 68.2 (C5), 67.4 (C4), 66.9 (CH<sub>2</sub>Fmoc), 62.0 (C6), 54.5 (C $\alpha$ ), 48.4 (C2), 47.2 (CH<sub>Fmoc</sub>), 36.2 (C $\beta$ ), 28.0 (C(CH<sub>3</sub>)<sub>3</sub>), 23.3 (NHCOCH<sub>3</sub>), 20.7, 20.7, 20.6 (3 COCH<sub>3</sub>).

Ala-Pro-Asp-Cys( $\alpha$ -S-GalNAc)-Arg-Pro-NH<sub>2</sub> (17). Following SPPS methodology with Fmoc-Pro-OH (337 mg, 1 mmol), Fmoc-Arg(Pbf)-OH (649 mg, 1 mmol), Fmoc-Cys( $\alpha$ -S-(AcO)<sub>3</sub>GalNAc)-OH (137 mg, 0.2 mmol), Fmoc-Asp(O<sub>i</sub>Bu)-OH (412 mg, 1 mmol), Fmoc-Pro-OH (337 mg, 1 mmol), and Fmoc-Ala-OH (311 mg, 1 mmol), peptide 17 was obtained with a 58% yield after purification by reversed phase HPLC and lyophilization (*t<sub>R</sub>* 20.7 min). HRMS (ESI+) (*m/z*) 860.3936 [M + H]<sup>+</sup>; calcd C<sub>34</sub>H<sub>58</sub>N<sub>11</sub>O<sub>13</sub>S<sup>+</sup> 860.3938. <sup>1</sup>H NMR (400 MHz, H<sub>2</sub>O/D<sub>2</sub>O 9:1)  $\delta$  (ppm) amide protons: 8.67 (t, *J* = 6.9 Hz, 1H, NHAsp), 8.40 (d, *J* = 7.5 Hz, 1H, NHCys), 8.36 (d, *J* = 7.4 Hz, 1H, NHArg), 8.20 (d, *J* = 8.2 Hz, 1H, NHCOCH<sub>3</sub>), 7.70 (s, 1H, CONH<sub>2</sub>), 7.20 (t, *J* = 5.3 Hz, 1H, NH  $\epsilon$  Arg), 7.00 (s, 1H, CONH<sub>2</sub>), 6.67 (br s, 2H, NH<sub>2</sub>  $\eta$  Arg). <sup>1</sup>H NMR (400 MHz, D<sub>2</sub>O)  $\delta$  (ppm): 5.58 (d, *J* = 5.4 Hz, 1H, H<sub>1</sub>), 4.73 (t, *J* = 6.7 Hz, 1H, H  $\alpha$  Asp), 4.68 (d, *J* = 3.9 Hz, 1H, H  $\alpha$  Arg), 4.65–4.57 (m, 1H, H  $\alpha$  Cys), 4.54–4.46 (m, 1H, H  $\alpha$  Pro), 4.44–4.35 (m, 3H, H<sub>2</sub>, H  $\alpha$  Pro, H  $\alpha$  Ala), 4.27–4.17 (m, 1H, H<sub>5</sub>), 4.00 (d, *J* = 2.9 Hz, 1H, H<sub>4</sub>), 3.85–3.59 (m, 7H, 2H  $\delta$  Pro, 2H  $\delta$  Pro, 2H<sub>6</sub>, H<sub>3</sub>), 3.30–3.18 (m, 2H, 2H  $\delta$  Arg), 3.16–2.83 (m, 4H, 2H  $\beta$  Asp, 2H  $\beta$  Cys), 2.45–2.25 (m, 2H, H  $\beta$  Pro, H  $\beta$  Pro), 2.14–1.91 (m, 9H, NHCOCH<sub>3</sub>, 2H  $\beta$  Arg, H  $\beta$  Pro, H  $\beta$  Pro, H  $\gamma$  Pro, H  $\gamma$  Pro), 1.82–1.60 (m, 4H, H  $\gamma$  Pro, H  $\gamma$  Pro, 2H  $\gamma$  Arg), 1.55 (d, *J* = 7.0 Hz, 3H, 3H  $\beta$  Ala). <sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>)  $\delta$  (ppm): 176.8, 174.6, 173.9, 173.6, 171.9, 171.4, 171.1, 169.2 (8 CO), 156.8 (C  $\zeta$  Arg), 85.4 (C<sub>1</sub>), 71.9 (C<sub>5</sub>), 68.4 (C<sub>4</sub>), 67.6 (C<sub>3</sub>), 61.1 (C<sub>6</sub>), 60.3 (C  $\alpha$  Pro, C  $\alpha$  Pro), 53.7 (C  $\alpha$  Cys), 51.1 (C  $\alpha$  Arg), 50.1 (C  $\alpha$  Asp), 50.0 (C<sub>2</sub>), 48.0 (C  $\alpha$  Ala), 47.9 (C  $\delta$  Pro), 47.7 (C  $\delta$  Pro), 40.6 (C  $\delta$  Arg), 35.2 (C  $\beta$  Asp), 32.6 (C  $\beta$  Cys), 29.6 (C  $\beta$  Pro), 29.3 (C  $\beta$  Pro), 27.7 (C  $\gamma$  Pro), 24.7 (C  $\gamma$  Pro), 24.6 (C  $\beta$  Arg), 24.1 (C  $\gamma$  Arg), 21.9 (NHCOCH<sub>3</sub>), 15.1 (C  $\beta$  Ala).

Ala-Pro-Asp-hCys( $\alpha$ -S-GalNAc)-Arg-Pro-NH<sub>2</sub> (18). Following the SPPS methodology with Fmoc-Pro-OH (337 mg, 1 mmol), Fmoc-Arg(Pbf)-OH (649 mg, 1 mmol), Fmoc-hCys( $\alpha$ -S-(AcO)<sub>3</sub>GalNAc)-OH (135 mg, 0.2 mmol), Fmoc-Asp(O<sub>i</sub>Bu)-OH (412 mg, 1 mmol), Fmoc-Pro-OH (337 mg, 1 mmol), and Fmoc-Ala-OH (311 mg, 1 mmol), peptide 18 was obtained with a 67% yield after



purification by reversed phase HPLC and lyophilization ( $t_R$  20.8 min). HRMS (ESI+) ( $m/z$ ) 874.4084 [ $M + H$ ] $^+$ ; calcd  $C_{35}H_{60}N_{11}O_{13}S^+$  874.4087.  $^1H$  NMR (400 MHz,  $H_2O/D_2O$  9:1)  $\delta$  (ppm) amide protons: 8.63 (d,  $J = 7.0$  Hz, 1H, NHAsp), 8.39 (d,  $J = 7.6$  Hz, 1H, NHArg), 8.27 (d,  $J = 7.2$  Hz, 1H, NHhCys), 8.21 (d,  $J = 8.2$  Hz, 1H, NHCOCH<sub>3</sub>), 7.71 (s, 1H, CONH<sub>2</sub>), 7.21 (t,  $J = 5.2$  Hz, 1H, NH  $\epsilon$  Arg), 7.00 (s, 1H, CONH<sub>2</sub>), 6.67 (br s, 2H, NH<sub>2</sub>  $\eta$  Arg).  $^1H$  NMR (400 MHz, D<sub>2</sub>O)  $\delta$  (ppm): 5.50 (d,  $J = 5.4$  Hz, 1H, H1), 4.68–4.63 (m, 2H, H  $\alpha$  Asp, H  $\alpha$  hCys), 4.54–4.47 (m, 2H, H  $\alpha$  Pro, H  $\alpha$  Arg), 4.42–4.34 (m, 3H, H<sub>2</sub>, H  $\alpha$  Pro, H  $\alpha$  Ala), 4.30–4.27 (m, 1H, H<sub>5</sub>), 4.00 (s, 1H, H<sub>4</sub>), 3.89–3.64 (m, 7H, 2H  $\delta$  Pro, 2H  $\delta$  Pro, 2H<sub>6</sub>, H<sub>3</sub>), 3.23 (t,  $J = 6.7$  Hz, 2H, 2H  $\delta$  Arg), 3.01–2.85 (m, 2H, 2H  $\beta$  Asp), 2.67 (t,  $J = 6.9$  Hz, 2H, 2H  $\gamma$  hCys), 2.39–2.28 (m, 2H, H  $\beta$  Pro, H  $\beta$  Pro), 2.23–1.67 (m, 15H, H  $\beta$  Pro, H  $\beta$  Pro, 2H  $\gamma$  Pro, 2H  $\gamma$  Pro, 2H  $\beta$  hCys, 2H  $\beta$  Arg, 2H  $\gamma$  Arg, NHCOCH<sub>3</sub>), 1.55 (d,  $J = 6.9$  Hz, 3H, 3H  $\beta$  Ala).  $^{13}C$  NMR (100 MHz, CDCl<sub>3</sub>)  $\delta$  (ppm): 175.8, 174.5, 173.8, 173.5, 172.8, 172.1, 171.3, 169.3 (8 CO), 156.8 (C  $\zeta$  Arg), 84.8 (C1), 71.8 (C5), 68.5 (C4), 67.7 (C3), 61.2 (C6), 60.3 (C  $\alpha$  Pro, C  $\alpha$  Pro), 52.4 (C  $\alpha$  Arg), 51.2 (C  $\alpha$  hCys), 50.1 (C  $\alpha$  Asp), 50.0 (C2), 48.0 (C  $\alpha$  Ala), 47.9 (C  $\delta$  Pro), 47.7 (C  $\delta$  Pro), 40.5 (C  $\delta$  Arg), 35.0 (C  $\beta$  Asp), 31.3 (C  $\gamma$  Pro), 29.6 (C  $\beta$  Pro), 29.4 (C  $\beta$  Pro), 27.4 (C  $\gamma$  Pro), 27.1 (C  $\gamma$  hCys), 24.7 (C  $\beta$  Arg), 24.7 (C  $\beta$  hCys), 24.1 (C  $\gamma$  Arg), 21.9 (NHCOCH<sub>3</sub>), 15.1 (C  $\beta$  Ala).

Ala-Pro-Asp-hhCys( $\alpha$ -S-GalNAc)-Arg-Pro-NH<sub>2</sub> (19). Following the SPPS methodology with Fmoc-Pro-OH (337 mg, 1 mmol), Fmoc-Arg(Pbf)-OH (649 mg, 1 mmol), Fmoc-hhCys( $\alpha$ -S-(AcO)<sub>3</sub>GalNAc)-OH (140 mg, 0.2 mmol), Fmoc-Asp(O<sub>t</sub>Bu)-OH (412 mg, 1 mmol), Fmoc-Pro-OH (337 mg, 1 mmol), and Fmoc-Ala-OH (311 mg, 1 mmol), peptide 19 was obtained with a 59% yield after purification by reversed phase HPLC and lyophilization ( $t_R$  24.4 min). HRMS (ESI+) ( $m/z$ ) 888.4251 [ $M + H$ ] $^+$ ; calcd  $C_{36}H_{62}N_{11}O_{13}S^+$  888.4249.  $^1H$  NMR (400 MHz,  $H_2O/D_2O$  9:1)  $\delta$  (ppm) amide protons: 8.63 (d,  $J = 7.0$  Hz, 1H, NHAsp), 8.32 (d,  $J = 7.3$  Hz, 1H, NHArg), 8.27 (d,  $J = 7.1$  Hz, 1H, NHhhCys), 8.20 (d,  $J = 8.1$  Hz, 1H, NHCOCH<sub>3</sub>), 7.69 (s, 1H, CONH<sub>2</sub>), 7.20 (t,  $J = 5.2$  Hz, 1H, NH  $\epsilon$  Arg), 7.00 (s, 1H, CONH<sub>2</sub>), 6.67 (br s, 2H, NH<sub>2</sub>  $\eta$  Arg).  $^1H$  NMR (400 MHz, D<sub>2</sub>O)  $\delta$

(ppm): 5.49 (d,  $J = 5.4$  Hz, 1H, H1), 4.68–4.64 (m, 2H, H  $\alpha$  Asp, H  $\alpha$  Arg), 4.50–4.46 (m, 1H, H  $\alpha$  Pro), 4.41–4.30 (m, 4H, H2, H  $\alpha$  Pro, H  $\alpha$  Ala, H  $\alpha$  hhCys), 4.25 (t,  $J = 5.9$  Hz, 1H, H5), 3.98 (d,  $J = 2.8$  Hz, 1H, H4), 3.83 (dd,  $J = 11.4, 3.1$  Hz, 1H, H3), 3.81–3.58 (m, 6H, 2H6, 2H  $\delta$  Pro, 2H  $\delta$  Pro), 3.27–3.16 (m, 2H, 2H  $\delta$  Arg), 2.95 (dd,  $J = 17.0, 6.5$  Hz, 1H, H  $\beta$  Asp), 2.85 (dd,  $J = 17.0, 7.1$  Hz, 1H, H  $\beta$  Asp), 2.70–2.55 (m, 2H, 2H  $\delta$  hhCys), 2.38–2.27 (m, 2H, H  $\beta$  Pro, H  $\beta$  Pro), 2.10–1.81 (m, 11H, NHCOCH<sub>3</sub>, 2H  $\beta$  hhCys, 2H  $\gamma$  hhCys, H  $\beta$  Pro, H  $\beta$  Pro, H  $\gamma$  Pro, H  $\gamma$  Pro), 1.80–1.58 (m, 6H, 2H  $\gamma$  Arg, 2H  $\beta$  Arg, H  $\gamma$  Pro, H  $\gamma$  Pro), 1.54 (d,  $J = 7.0$  Hz, 3H, 3H  $\beta$  Ala). <sup>13</sup>C NMR (100 MHz, D<sub>2</sub>O)  $\delta$  (ppm): 176.8, 174.5, 173.8, 173.5, 173.1, 172.0, 171.3, 169.2 (8 CO), 156.7 (C  $\zeta$  Arg), 83.6 (C1), 71.7 (C5), 68.5 (C4), 67.7 (C3), 61.2 (C6), 60.3 (C  $\alpha$  Pro), 60.3 (C  $\alpha$  Pro), 53.2 (C  $\alpha$  hhCys), 51.0 (C  $\alpha$  Arg), 50.1 (C  $\alpha$  Asp), 50.0 (C2), 48.0 (C  $\alpha$  Ala), 47.9 (C  $\delta$  Pro), 47.7 (C  $\delta$  Pro), 40.6 (C  $\delta$  Arg), 35.1 (C  $\beta$  Asp), 30.0 (C  $\gamma$  Pro), 29.6 (C  $\beta$  Pro, C  $\beta$  hhCys), 29.4 (C  $\beta$  Pro), 29.4 (C  $\delta$  hhCys), 27.4 (C  $\gamma$  Pro), 25.0 (C  $\beta$  Arg), 25.0 (C  $\gamma$  hhCys), 24.1 (C  $\gamma$  Arg), 21.9 (NHCOCH<sub>3</sub>), 15.1 (C  $\beta$  Ala).

Ala-Pro-Asp-Ser(O-mercaptopropyl- $\alpha$ -S-GalNAc)-Arg-Pro-NH<sub>2</sub> (20).

Following the SPPS methodology with Fmoc-Pro-OH (337 mg, 1 mmol), Fmoc-Arg(Pbf)-OH (649 mg, 1 mmol), Fmoc-Ser-(mercaptopropyl- $\alpha$ -S-(AcO)<sub>3</sub>GalNAc)-OH (146 mg, 0.2 mmol), Fmoc-Asp(O<sub>t</sub>Bu)-OH (412 mg, 1 mmol), Fmoc-Pro-OH (337 mg, 1 mmol), and Fmoc-Ala-OH (311 mg, 1 mmol), peptide 20 was obtained with a 67% yield after purification by reversed phase HPLC and lyophilization ( $t_r$  25.7 min). HRMS (ESI+) ( $m/z$ ) 459.7210 [ $M + 2H$ ]<sub>+2</sub>; calcd C<sub>37</sub>H<sub>65</sub>N<sub>11</sub>O<sub>14</sub>S<sub>2</sub>+ 459.7211. <sup>1</sup>H NMR (400 MHz, H<sub>2</sub>O/D<sub>2</sub>O 9:1)  $\delta$  (ppm) amide protons: 8.60 (d,  $J = 6.8$  Hz, 1H, NHAsp), 8.31–8.20 (m, 3H, NHSer, NHArg, NHCOCH<sub>3</sub>), 7.69 (s, 1H, CONH<sub>2</sub>), 7.25 (d,  $J = 6.8$  Hz, 1H, NH  $\epsilon$  Arg), 7.01 (s, 1H, CONH<sub>2</sub>), 6.68 (br s, 2H, NH<sub>2</sub>  $\eta$  Arg). <sup>1</sup>H NMR (400 MHz, D<sub>2</sub>O)  $\delta$  (ppm): 5.50 (d,  $J = 5.4$  Hz, 1H, H1), 4.76–4.68 (m, 2H, H  $\alpha$  Asp, H  $\alpha$  Arg), 4.53 (t,  $J = 5.0$  Hz, 1H, H  $\alpha$  Ser), 4.51–4.47 (m, 1H, H  $\alpha$  Pro), 4.42–4.32 (m, 3H, H2, H  $\alpha$  Pro, H  $\alpha$  Ala), 4.26 (t,  $J = 6.0$  Hz, 1H, H5), 3.99 (d,  $J = 2.8$  Hz, 1H, H4), 3.83 (dd,  $J = 11.4, 3.0$  Hz, 1H, H3), 3.81–3.70 (m, 6H, 2H6, H  $\delta$  Pro, H  $\delta$  Pro, 2H  $\beta$  Ser),

3.69–3.55 (m, 4H, H  $\delta$  Pro, H  $\delta$  Pro, 2H  $\delta$  Ser), 3.27–3.17 (m, 2H, 2H  $\delta$  Arg), 2.98 (dd,  $J = 17.1, 6.4$  Hz, 1H, H  $\beta$  Asp), 2.89 (dd,  $J = 17.1, 6.9$  Hz, 1H, H  $\beta$  Asp), 2.72–2.56 (m, 2H, H  $\zeta$  Ser), 2.39–2.26 (m, 2H, H  $\beta$  Pro, H  $\beta$  Pro), 2.11–1.80 (m, 12H, NHCOCH<sub>3</sub>, 2H  $\beta$  Arg, 2H  $\epsilon$  Ser, H  $\beta$  Pro, H  $\beta$  Pro, 2H  $\gamma$  Pro, H  $\gamma$  Pro), 1.79–1.59 (m, 3H, H  $\gamma$  Pro, 2H  $\gamma$  Arg), 1.55 (d,  $J = 7.0$  Hz, 3H, 3H  $\beta$  Ala). <sup>13</sup>C NMR (100 MHz, D<sub>2</sub>O)  $\delta$  (ppm): 176.8, 174.5, 173.8, 173.6, 172.2, 171.1, 170.0, 169.3 (8 CO), 156.7 (C  $\zeta$  Arg), 83.9 (C1), 71.7 (C5), 69.5 (C  $\delta$  Ser), 69.1 (C  $\beta$  Ser), 68.5 (C4), 67.7 (C3), 61.2 (C6), 60.4 (C  $\alpha$  Pro), 60.4 (C  $\alpha$  Pro), 53.6 (C  $\alpha$  Ser), 51.0 (C  $\alpha$  Arg), 50.2 (C2), 50.0 (C  $\alpha$  Asp), 48.1 (C  $\alpha$  Ala), 47.9 (C  $\delta$  Pro), 47.8 (C  $\delta$  Pro), 40.6 (C  $\delta$  Arg), 35.2 (C  $\beta$  Asp), 29.6 (C  $\beta$  Pro), 29.4 (C  $\beta$  Pro), 28.7 (C  $\epsilon$  Ser), 27.7 (C  $\gamma$  Pro), 26.9 (C  $\zeta$  Ser), 24.7, 24.6 (C  $\beta$  Arg, C  $\gamma$  Pro), 24.0 (C  $\gamma$  Arg), 21.9 (NHCOCH<sub>3</sub>), 15.1 (C  $\beta$  Ala).

Ala-Pro-Asp-Cys(S-mercaptopropyl- $\alpha$ -S-GalNAc)-Arg-Pro-NH<sub>2</sub> (21).

Following the SPPS methodology with Fmoc-Pro-OH (337 mg, 1 mmol), Fmoc-Arg(Pbf)-OH (649 mg, 1 mmol), Fmoc-Cys-(mercaptopropyl- $\alpha$ -S-(AcO)<sub>3</sub>GalNAc)-OH (149 mg, 0.2 mmol), Fmoc-sp(O<sub>t</sub>Bu)-OH (412 mg, 1 mmol), Fmoc-Pro-OH (337 mg, 1 mmol), and Fmoc-Ala-OH (311 mg, 1 mmol), peptide 21 was obtained with a 63% yield after purification by reversed phase HPLC and lyophilization ( $t_r$  30.3 min). HRMS (ESI+) ( $m/z$ ) 934.4123 [ $M + H$ ]<sup>+</sup>; calcd C<sub>37</sub>H<sub>64</sub>N<sub>11</sub>O<sub>13</sub>S<sub>2</sub> + 934.4121. <sup>1</sup>H NMR (400 MHz, H<sub>2</sub>O/D<sub>2</sub>O 9:1)  $\delta$  (ppm) amide protons: 8.56 (d,  $J = 7.3$  Hz, 1H, NHAsp), 8.31 (d,  $J = 7.5$  Hz, 1H, NHArg), 8.18 (d,  $J = 7.4$  Hz, 1H, NHCys), 8.13 (d,  $J = 8.0$  Hz, 1H, NHCOCH<sub>3</sub>), 7.59 (s, 1H, CONH<sub>2</sub>), 7.09 (t,  $J = 5.4$  Hz, 1H, NH  $\epsilon$  Arg), 6.91 (s, 1H, CONH<sub>2</sub>), 6.57 (br s, 1H, NH<sub>2</sub>  $\eta$  Arg). <sup>1</sup>H NMR (400 MHz, D<sub>2</sub>O)  $\delta$  (ppm): 5.51 (d,  $J = 5.4$  Hz, 1H, H1), 4.74–4.66 (m, 2H, H  $\alpha$  Asp, H  $\alpha$  Arg), 4.55–4.47 (m, 2H, H  $\alpha$  Cys, H  $\alpha$  Pro), 4.41–4.31 (m, 3H, H2, H  $\alpha$  Pro, H  $\alpha$  Ala), 4.26 (t,  $J = 6.0$  Hz, 1H, H5), 3.99 (d,  $J = 2.4$  Hz, 1H, H4), 3.86–3.79 (m, 2H, H3, H  $\delta$  Pro), 3.77 (d,  $J = 6.2$  Hz, 2H, 2H6), 3.75–3.69 (m, 1H, H  $\delta$  Pro), 3.69–3.60 (m, 2H, 2H  $\delta$  Pro), 3.26–3.18 (m, 2H, 2H  $\delta$  Arg), 3.02–2.94 (m, 2H, H  $\beta$  Cys, H  $\beta$  Asp), 2.91–2.84 (m, 2H, H  $\beta$  Asp, H  $\beta$  Cys), 2.77–2.63 (m, 4H, 2H  $\zeta$  Cys, 2H  $\delta$  Cys), 2.39–2.27 (m, 2H, H  $\beta$  Pro, H  $\beta$  Pro),

2.10–1.92 (m, 9H, NHCOCH<sub>3</sub>, 2H  $\beta$  Arg, 2H  $\gamma$  Pro, 2H  $\beta$  Pro), 1.91–1.82 (m, 3H, 2H  $\epsilon$  Cys, H  $\gamma$  Pro), 1.80–1.61 (m, 3H, H  $\gamma$  Pro, 2H  $\gamma$  Arg), 1.54 (d,  $J = 7.0$  Hz, 3H, 3H  $\beta$  Ala). <sup>13</sup>C NMR (100 MHz, D<sub>2</sub>O)  $\delta$  (ppm): 176.8, 174.5, 173.8, 173.5, 171.9, 171.5, 171.1, 169.3 (8 CO), 156.7 (C  $\epsilon$  Arg), 83.9 (C1), 71.7 (C5), 68.5 (C4), 67.6 (C3), 61.2 (C6), 60.4 (C  $\alpha$  Pro), 60.3 (C  $\alpha$  Pro), 52.9 (C  $\alpha$  Cys), 51.1 (C  $\alpha$  Arg), 50.2 (C2), 49.9 (C  $\alpha$  Asp), 48.0 (C  $\alpha$  Ala), 47.9 (C  $\delta$  Pro), 47.7 (C  $\delta$  Pro), 40.6 (C  $\delta$  Arg), 35.1 (C  $\beta$  Asp), 32.5 (C  $\beta$  Cys), 30.3 (C  $\delta$  Cys), 29.6 (C  $\beta$  Pro), 29.4 (C  $\beta$  Pro), 29.1 (C  $\zeta$  Cys), 28.6 (C  $\epsilon$  Cys), 27.6 (C  $\gamma$  Pro), 24.7, 24.6 (C  $\beta$  Arg, C  $\gamma$  Pro), 24.1 (C  $\gamma$  Arg), 21.9 (NHCOCH<sub>3</sub>), 15.1 (C  $\beta$  Ala).

Ala-Pro-Asp-Thr(O-mercaptopropyl- $\alpha$ -S-GalNAc)-Arg-Pro-NH<sub>2</sub> (22).  
 Following the SPPS methodology with Fmoc-Pro-OH (337 mg, 1 mmol), Fmoc-Arg(Pbf)-OH (649 mg, 1 mmol), Fmoc-Thr-(mercaptopropyl- $\alpha$ -S-(AcO)<sub>3</sub>GalNAc)-OH (149 mg, 0.2 mmol), Fmoc-Asp(O<sub>t</sub>Bu)-OH (412 mg, 1 mmol), Fmoc-Pro-OH (337 mg, 1 mmol), and Fmoc-Ala-OH (311 mg, 1 mmol), glycopeptide 22 was obtained with a 62% yield after purification by reversed phase HPLC and lyophilization ( $t_R$  29.7 min). HRMS (ESI+) ( $m/z$ ) 932.4506 [ $M + H$ ]<sup>+</sup>; calcd C<sub>38</sub>H<sub>66</sub>N<sub>11</sub>O<sub>14</sub>S<sub>4</sub> 932.4506. <sup>1</sup>H NMR (400 MHz, H<sub>2</sub>O/D<sub>2</sub>O 9:1)  $\delta$  (ppm) amide protons: 8.71 (d,  $J = 7.0$  Hz, 1H, NHAsp), 8.23 (d,  $J = 7.9$  Hz, 1H, NHCOCH<sub>3</sub>), 8.18 (d,  $J = 7.4$  Hz, 1H, NHArg), 8.06 (d,  $J = 8.2$  Hz, 1H, NHThr), 7.68 (s, 1H, CONH<sub>2</sub>), 7.21 (t,  $J = 4.5$  Hz, 1H, NH  $\epsilon$  Arg), 7.01 (s, 1H, CONH<sub>2</sub>), 6.67 (br s, 2H, NH<sub>2</sub>  $\eta$  Arg). <sup>1</sup>H NMR (400 MHz, D<sub>2</sub>O)  $\delta$  (ppm): 5.45 (d,  $J = 5.5$  Hz, 1H, H1), 4.78–4.74 (m, 1H, H  $\alpha$  Asp), 4.65 (dd,  $J = 8.1, 5.3$  Hz, 1H, H  $\alpha$  Arg), 4.46 (dd,  $J = 8.1, 6.3$  Hz, 1H, H  $\alpha$  Pro), 4.40–4.28 (m, 4H, H2, H  $\alpha$  Thr, H  $\alpha$  Pro, H  $\alpha$  Ala), 4.22 (t,  $J = 6.0$  Hz, 1H, H5), 4.01–3.94 (m, 2H, H  $\beta$  Thr, H4), 3.79 (dd,  $J = 11.4, 3.1$  Hz, 1H, H3), 3.76–3.55 (m, 7H, 2H6, 2H  $\delta$  Pro, 2H  $\delta$  Pro, H  $\delta$  Thr), 3.46–3.38 (m, 1H, H  $\delta$  Thr), 3.23–3.15 (m, 2H, 2H  $\delta$  Arg), 2.97 (dd,  $J = 17.1, 7.0$  Hz, 1H, H  $\beta$  Asp), 2.87 (dd,  $J = 17.1, 6.6$  Hz, 1H, H  $\beta$  Asp), 2.69–2.50 (m, 2H, 2H  $\zeta$  Thr), 2.36–2.22 (m, 2H, H  $\beta$  Pro, H  $\beta$  Pro), 2.08–1.71 (m, 13H, NHCOCH<sub>3</sub>, 2H  $\beta$  Arg, 2H  $\epsilon$  Thr, H  $\beta$  Pro, H  $\beta$  Pro, 2H  $\gamma$  Pro, 2H  $\gamma$  Pro), 1.70–1.57 (m, 2H, 2H  $\gamma$  Arg), 1.51 (d,  $J = 7.0$  Hz, 3H, 3H  $\beta$  Ala), 1.13 (d,  $J = 6.3$  Hz, 3H, 3H  $\gamma$  Thr). <sup>13</sup>C

NMR (100 MHz, D<sub>2</sub>O)  $\delta$  (ppm): 176.7, 174.5, 173.8, 173.5, 172.5, 171.2, 171.1, 169.2 (8 CO), 156.7 (C  $\zeta$  Arg), 83.9 (C1), 74.8 (C  $\beta$  Thr), 71.7 (C5), 68.4 (C4), 67.6 (C3), 67.4 (C  $\delta$  Thr), 61.1 (C6), 60.2 (C  $\alpha$  Pro, C  $\alpha$  Pro), 57.8 (C  $\alpha$  Thr), 51.1 (C  $\alpha$  Arg), 50.2 (C2), 49.9 (C  $\alpha$  Asp), 48.0 (C  $\alpha$  Ala), 47.9 (C  $\delta$  Pro), 47.6 (C  $\delta$  Pro), 40.6 (C  $\delta$  Arg), 35.0 (C  $\beta$  Asp), 29.6 (C  $\beta$  Pro), 29.4 (C  $\beta$  Pro), 29.0 (C  $\epsilon$  Thr), 27.6 (C  $\gamma$  Pro), 26.9 (C  $\zeta$  Thr), 24.7, 24.6 (C  $\beta$  Arg, C  $\gamma$  Pro), 24.1 (C  $\gamma$  Arg), 21.9 (NHCOCH<sub>3</sub>), 15.7 (C  $\gamma$  Thr), 15.0 (C  $\beta$  Ala).

2D NMR Experiments. Spectra were assigned using COSY and HSQC experiments. NOESY experiments were recorded on a 400 MHz spectrometer at 298 K and pH 5.5 in H<sub>2</sub>O/D<sub>2</sub>O (9:1). The experiments were conducted by using phase-sensitive ge-2D NOESY with WATERGATE for H<sub>2</sub>O/D<sub>2</sub>O (9:1) spectra. 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.

Molecular Dynamics (MD) Simulations with Time-Averaged Restraints (MD-tar). The simulations were carried out with the AMBER 12 package<sup>54</sup> with ff14SB,<sup>55</sup> GAFF,<sup>56</sup> and GLYCAMO6j<sup>57</sup> force fields. The parameters and charges for the unnatural amino acids were generated with the antechamber module of AMBER 12, using GAFF force field and AM1-BCC method for charges.<sup>58</sup> Prior to MDtar productive simulations, we performed an equilibration protocol consisting of an initial minimization of the water box of 5000 steps, followed by a 2500-step minimization of the whole system. Then the TIP3P water<sup>59</sup> box was heated at constant volume until 298 K using a time constant for the heat bath coupling of 1 ps. The equilibration finished with 200 ps of MD simulation without restraints at a constant pressure of 1 bar and turning on the Langevin temperature scaling with a collision frequency of 1 ps. Furthermore, nonbonded interactions were cutoff at 8.0 Å and updated every 25 steps. Periodic boundary conditions and the particle mesh Ewald method<sup>60</sup> were turned on in every step of the equilibration protocol to evaluate the long-range electrostatic forces using a grid spacing of approximately 1 Å. The NOE-derived distances shown in Figure 2 and in the Supporting Information were imposed as time-averaged constraint,

applying an  $r^{-6}$  averaging. The equilibrium distance range was set to  $r_{\text{exp}} - 0.2 \text{ \AA} \leq r_{\text{exp}} \leq r_{\text{exp}} + 0.2 \text{ \AA}$ . Trajectories were run at 298 K, with a decay constant of 2000 ps and a time step of 1 fs. The force constants  $r_{k2}$  and  $r_{k3}$  used in each case was  $10 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{\AA}^{-2}$ . The overall simulation length for the simulations was 20 ns. The coordinates were saved each 1 ps, obtaining MD trajectories of 20 000 frames each. A convergence within the equilibrium distance range was obtained in the simulations.

Molecular Dynamics (MD) Simulations on the Complex SM3:22. The crystal structure of glycopeptide 22 in complex with scFv-SM3 (PDB id 5fxc) was used as starting coordinates. The structure was conveniently modified to contain the missing linker as well as the GalNAc unit for the glycopeptide. The complex was then immersed in a water box with a  $10 \text{ \AA}$  buffer of TIP3P water molecules. A two-stage geometry optimization approach was performed. The first stage minimizes only the positions of solvent molecules, and the second stage is an unrestrained minimization of all atoms in the simulation cell. The systems were then gently heated by incrementing the temperature from 0 to 300 K under a constant pressure of 1 atm and periodic boundary conditions. Harmonic restraints of  $30 \text{ kcal}\cdot\text{mol}^{-1}$  were applied to the solute, and the Andersen temperaturecoupling scheme was used to control and equalize the temperature. The time step was kept at 1 fs during the heating stages, allowing potential inhomogeneities to self-adjust. Long-range electrostatic effects were modeled using the particle mesh Ewald method. An  $8 \text{ \AA}$  cutoff was applied to Lennard-Jones and electrostatic interactions. Each system was equilibrated for 2 ns with a 2 fs time step at a constant volume and temperature of 300 K. Production trajectories were then run for an additional 100 ns under the same simulation conditions.

Crystallization. Expression and purification of scFv-SM3 has been described previously by us.<sup>44</sup> Crystals were grown by sitting drop diffusion at  $18 \text{ }^\circ\text{C}$ . The drops were prepared by mixing  $0.5 \text{ }\mu\text{L}$  of protein solution, containing  $15 \text{ mg/mL}$  of scFv-SM3 and  $10 \text{ mM}$  of glycopeptide 22 with  $0.5 \text{ }\mu\text{L}$  of the mother liquor. Crystals of scFv-SM3 with the peptide above were grown in 20%

PEG 3350, 0.2 M disodium hydrogen phosphate. The crystals were cryoprotected in mother liquor containing 15% ethylene glycol and frozen in a nitrogen gas stream cooled to 100 K.

Structure Determination and Refinement. The data was processed and scaled using the XDS package<sup>61</sup> and CCP4 software;<sup>62</sup> relevant statistics are given in Supporting Information. The crystal structures were solved by molecular replacement with Phaser<sup>62</sup> and using the PDB entry 1SM3 as the template. Initial phases were further improved by cycles of manual model building in Coot<sup>63</sup> and refinement with REFMAC5.<sup>64</sup> The final models were validated with PROCHECK;<sup>65</sup> model statistics are given in the Supporting Information. Coordinates and structure factors have been deposited in the Worldwide Protein Data Bank (wwPDB); PDB id 5fxc.

## ASSOCIATED CONTENT

### \*<sub>s</sub>Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.joc.6b00833. 2D NOESY spectra of the neoglycopeptides, conformational analysis details, NMR spectra for all new compounds, CD spectra, HPLC chromatograms, and X-ray structure details (PDF)

## AUTHOR INFORMATION

### CORRESPONDING AUTHORS

\*E-mail: [jesusmanuel.peregrina@unirioja.es](mailto:jesusmanuel.peregrina@unirioja.es).

\*E-mail: [hector.busto@unirioja.es](mailto:hector.busto@unirioja.es).

\*E-mail: [francisco.corzana@unirioja.es](mailto:francisco.corzana@unirioja.es).

### PRESENT ADDRESS

§C.A.: CECB, Department of Chemistry, University of

Copenhagen, Universitetsparken 5, 2100 Copenhagen, Denmark.

### AUTHOR CONTRIBUTIONS

‡V.R.-O. and I.C. contributed equally.

### NOTES

The authors declare no competing financial interest.

### ACKNOWLEDGMENTS

We thank the Ministerio de Economía y Competitividad (projects CTQ2015-67727-R, CTQ2015-64597-P-C02-01, CTQ2013-44367-C2-2-P, and UNLR13-4E-1931). V.R.-O. and I.C. thank Gobierno de La Rioja and



Universidad de La Rioja for the FPI grants, respectively. We thank synchrotron radiation source ALBA (Barcelona) and beamline XALOC. This work was supported by Agencia Aragonesa para la Investigación y Desarrollo (ARAID), Diputación General de Aragón (DGA; B89 to R.H.-G.) and the EU Seventh Framework Programme (2007–2013) under BioStruct-X (grant agreement 283570 and BIOSTRUCTX 5186, to R.H.-G.).

## REFERENCES

- (1) (a) Gaidzik, N.; Westerlind, U.; Kunz, H. *Chem. Soc. Rev.* 2013, 42, 4421–4442. (b) Pinho, S. S.; Reis, C. A. *Nat. Rev. Cancer* 2015, 15, 540–555.
- (2) Nativi, C.; Renaudet, O. *ACS Med. Chem. Lett.* 2014, 5, 1176–1178.
- (3) Buskas, T.; Thompson, P.; Boons, G. J. *Chem. Commun.* 2009, 5335–5349.
- (4) (a) Wilson, R. M.; Danishefsky, S. J. *J. Am. Chem. Soc.* 2013, 135, 14462–14472. (b) Keding, S. J.; Danishefsky, S. J. *Proc. Natl. Acad. Sci. U. S. A.* 2004, 101, 11937–11942.
- (5) Ju, T.; Otto, V. I.; Cummings, R. D. *Angew. Chem., Int. Ed.* 2011, 50, 1770–1791.
- (6) Karsten, U.; Serttas, N.; Paulsen, H.; Danielczyk, A.; Goletz, S. *Glycobiology* 2004, 14, 681–692.
- (7) 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.
- (8) Palitzsch, B.; Gaidzik, N.; Stergiou, N.; Stahn, S.; Hartmann, S.; Gerlitzki, B.; Teusch, N.; Flemming, P.; Schmitt, E.; Kunz, H. *Angew. Chem., Int. Ed.* 2016, 55, 2894–2898.
- (9) Miles, D.; Roché, H.; Martin, M.; Perren, T. J.; Cameron, D. A.; Glaspy, J.; Dodwell, D.; Parker, J.; Mayordomo, J.; Tres, A.; Murray, J. L.; Ibrahim, N. K. *Oncologist* 2011, 16, 1092–1100.
- (10) Ohyama, C. *Int. J. Clin. Oncol.* 2008, 13, 308–313.
- (11) Richichi, B.; Thomas, B.; Fiore, M.; Bosco, R.; Qureshi, H.; Nativi, C.; Renaudet, O.; BenMohamed, L. *Angew. Chem., Int. Ed.* 2014, 53, 11917–11920.
- (12) Dondoni, A.; Marra, A. *Chem. Rev.* 2000, 100, 4395–4422.

- (13) Kuberan, B.; Sikkander, S. A.; Tomiyama, H.; Linhardt, R. J. *Angew. Chem., Int. Ed.* 2003, 42, 2073–2075.
- (14) Awad, L.; Madani, R.; Gillig, A.; Kolympani, M.; Philgren, M.; Muhs, A.; Gérard, C.; Vogel, P. *Chem. - Eur. J.* 2012, 18, 8578–8582.
- (15) Wu, X.; Lipinski, T.; Paszkiewicz, E.; Bundle, D. R. *Chem. --Eur. J.* 2008, 14, 6474–6482.
- (16) Bundle, D. R.; Rich, J. R.; Jacques, S.; Yu, H. N.; Nitz, M.; Ling, C.-C. *Angew. Chem., Int. Ed.* 2005, 44, 7725–7729.
- (17) 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.
- (18) Oberbillig, T.; Mersch, C.; Wagner, S.; Hoffmann-Röder, A. *Chem. Commun.* 2012, 48, 1487–1489.
- (19) Johannes, M.; Reindl, M.; Gerlitzki, B.; Schmitt, E.; Hoffmann-Röder, A. *Beilstein J. Org. Chem.* 2015, 11, 155–161.
- (20) (a) Vichier-Guerre, S.; Lo-Man, R.; Huteau, V.; Dériaud, E.; Leclerc, C.; Bay, S. *Bioorg. Med. Chem. Lett.* 2004, 14, 3567–3570. (b) Norgren, A. S.; Arvidsson, P. I. *Org. Biomol. Chem.* 2005, 3, 1359–1361.
- (21) (a) Wolfert, M. A.; Boons, G.-J. *Nat. Chem. Biol.* 2013, 9, 776–784. (b) Apostolopoulos, V.; Yuriev, E.; Lazoura, E.; Yu, M.; Ramsland, P. A. *Hum. Vaccines* 2008, 4, 400–409. (c) Glithero, A.; Tormo, J.; Haurum, J. S.; Arsequell, G. *Immunity* 1999, 10, 63–74. (d) Apostolopoulos, V.; Yuriev, E.; Ramsland, P. A.; Halton, J.; Osinski, C.; Li, W.; Plebanski, M.; Paulsen, H.; McKenzie, I. F. C. *Proc. Natl. Acad. Sci. U. S. A.* 2003, 100, 15029–15034. (e) Speir, J. A.; Abdel-Motal, U. M.; Jondal, M.; Wilson, I. A. *Immunity* 1999, 10, 51–61.
- (22) Bonauer, C.; Walenzyk, T.; König, B. *Synthesis* 2006, 2006, 1–20.
- (23) (a) Willey, J. M.; van der Donk, W. A. *Annu. Rev. Microbiol.* 2007, 61, 477–501. (b) Jack, R. W.; Jung, G. *Curr. Opin. Chem. Biol.* 2000, 4, 310–317. (c) Jung, G. *Angew. Chem., Int. Ed. Engl.* 1991, 30, 1051–168.
- (24) (a) Levengood, M. R.; van der Donk, W. A. *Nat. Protoc.* 2007, 1, 3001–3010. (b) Wang, J.; Schiller, S. M.; Schultz, P. G. *Angew. Chem., Int. Ed.* 2007, 46, 6849–851. (c) Bernardes, G. J. L.; Chalker, J. M.; Errey, J. C.; Davis, B. G. *J. Am. Chem. Soc.* 2008, 130, 5052–5053. (d) Guo, J.; Wang, J.; Lee, J. S.; Schultz, P. G. *Angew. Chem., Int. Ed.* 2008, 47,

6399–6401. (e) Chalker, J. M.; Bernardes, G. J. L.; Lin, Y. A.; Davis, B. G. *Chem. - Asian J.* 2009, 4, 630–640.

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

(26) (a) Zhu, Y.; van der Donk, W. A. *Org. Lett.* 2001, 3, 1189–119. (b) Galonic, D. P.; van der Donk, W. A.; Gin, D. Y. *Chem. - Eur. J.* 2003, 9, 5997–6006.

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

(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) (a) Floyd, N.; Vijaykrishnan, B.; Koeppe, J. R.; Davis, B. G. *Angew. Chem., Int. Ed.* 2009, 48, 7798–7802. (b) Fiore, M.; Lo Conte, M.; Pacifico, S.; Marra, A.; Dondoni, A. *Tetrahedron Lett.* 2011, 52, 444–447. (c) Dondoni, A. *Angew. Chem., Int. Ed.* 2008, 47, 8995–8997. (d) Griesbaum, K. *Angew. Chem., Int. Ed. Engl.* 1970, 9, 273–287. (e) Lázár, L.; Csávás, M.; Herczeg, M.; Herczegh, P.; Borbás, A. *Org. Lett.* 2012, 14, 4650–4653. (f) Markey, L.; Giordani, S.; Scanlan, E. M. *J. Org. Chem.* 2013, 78, 4270–4277. (g) Dénès, F.; Pichowicz, M.; Povie, G.; Renaud, P. *Chem. Rev.* 2014, 114, 2587–2693.

(30) (a) Zhu, X.; Schmidt, R. R. *Chem. - Eur. J.* 2004, 10, 875–887. (b) Galonic, D. P.; van der Donk, W. A.; Gin, D. Y. *J. Am. Chem. Soc.* 2004, 126, 12712–12713. (c) Galonic, D. P.; Ide, N. D.; van der Donk, W. A.; Gin, D. Y. *J. Am. Chem. Soc.* 2005, 127, 7359–7369. (d) Zhu, X.; Schmidt, R. R. *Tetrahedron Lett.* 2003, 44, 6063–6067.

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

(32) (a) Marra, A.; Staderini, S.; Berthet, N.; Dumy, P.; Renaudet, O.; Dondoni, A. *Eur. J. Org. Chem.* 2013, 2013, 1144–1149. (b) Hoyle, C. E.; Lowe, A. B.; Bowman, C. N. *Chem. Soc. Rev.* 2010, 39, 1355–1387. (c) Hoyle, C. E.; Bowman, C. N. *Angew. Chem., Int. Ed.* 2010, 49, 1540–1573.

(33) Kolb, H. C.; Finn, M. G.; Sharpless, K. B. *Angew. Chem., Int. Ed.* 2001, 40, 2004–2021.

(34) (a) Knapp, S.; Myers, D. S. *J. Org. Chem.* 2002, 67, 2995–2999. (b) Rojas, V.; Carreras, J.; Corzana, F.; Avenoza, A.; Busto, J. H.; Peregrina, J. M. *Carbohydr. Res.* 2013, 373, 1–8.

- (35) (a) Tasdelen, M. A.; Yagci, Y. *Angew. Chem., Int. Ed.* 2013, 52, 5930–5938. (b) Saada, M.-C.; Ombouma, J.; Montero, J.-L.; Supuran, C. T.; Winum, J.-Y. *Chem. Commun.* 2013, 49, 5699–5701.
- (c) Dondoni, A.; Marra, A. *Chem. Soc. Rev.* 2012, 41, 573–586.
- (36) (a) Triola, G.; Brunsveld, L.; Waldmann, H. *J. Org. Chem.* 2008, 73, 3646–3649. (b) Merbouh, N.; Wallner, F. K.; Cociorva, O. M.; Seeberger, P. H. *Org. Lett.* 2007, 9, 651–653.
- (37) Povie, G.; Tran, A. T.; Bonnaff, D.; Habegger, J.; Hu, Z.; Le Narvor, C.; Renaud, P. *Angew. Chem., Int. Ed.* 2014, 53, 3894–3898.
- (38) (a) Tadd, A. C.; Meinander, K.; Luthman, K.; Wallén, E. A. *J. Org. Chem.* 2011, 76, 673–675. (b) O'Donnell, M. J.; Delgado, F. *Tetrahedron* 2001, 57, 6641–6650.
- (39) Calce, E.; Leone, M.; Monfregola, L.; De Luca, S. *Org. Lett.* 2013, 15, 5354–5357.
- (40) Knerr, P. J.; Tzekou, A.; Ricklin, D.; Qu, H.; Chen, H.; van der Donk, W. A.; Lambris, J. D. *ACS Chem. Biol.* 2011, 6, 753–760.
- (41) Knerr, P. J.; van der Donk, W. A. *J. Am. Chem. Soc.* 2012, 134, 7648–7651.
- (42) Pearlman, D. J. *Biomol. NMR* 1994, 4, 1–16.
- (43) 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.
- (44) 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.; Peregrina, J. M.; Hurtado-Guerrero, R.; Corzana, F. *Angew. Chem., Int. Ed.* 2015, 54, 9830–9834.
- (45) Dyson, H. J.; Wright, P. E. *Annu. Rev. Biophys. Biophys. Chem.* 1991, 20, 519–538.
- (46) (a) 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. (b) Live, D. H.; Kumar, R. A.; Beebe, X.; Danishefsky, S. J. *Proc. Natl. Acad. Sci. U. S. A.* 1996, 93, 12759–12761. (c) O'Connor, S. E.; Imperiali, B. *Chem. Biol.* 1998, 5, 427–437.
- (47) Kinarsky, L.; Suryanarayanan, G.; Prakash, O.; Paulsen, H.; Clausen, H.; Hanisch, F.-G.; Hollingsworth, M. A.; Sherman, S. *Glycobiology* 2003, 13, 929–939.

- (48) Dziadek, S.; Griesinger, C.; Kunz, H.; Reinscheid, U. M. *Chem. - Eur. J.* 2006, 12, 4981–4993.
- (49) 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.
- (50) 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.
- (51) 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.
- (52) Matsushita, T.; Ohyabu, N.; Fujitani, N.; Naruchi, K.; Shimizu, H.; Hinou, H.; Nishimura, S.-I. *Biochemistry* 2013, 52, 402–414.
- (53) Mazeau, K.; Tvaroska, I. *Carbohydr. Res.* 1992, 225, 27–41.
- (54) Case, D. A.; Darden, T. A.; Cheatham, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Goetz, A. W.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wolf, R. M.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M. J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. *AMBER 12*; 2012, University of California, San Francisco.
- (55) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. J. *Chem. Theory Comput.* 2015, 11, 3696–3713.
- (56) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* 2004, 25, 1157–1174.
- (57) 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.
- (58) Jakalian, A.; Jack, D. B.; Bayly, C. I. *J. Comput. Chem.* 2002, 23, 1623–1641.
- (59) Kiyohara, K.; Gubbins, K.; Panagiotopoulos, A. *Mol. Phys.* 1998, 94, 803–808.
- (60) Darden, T.; York, D.; Pedersen, L. J. *Chem. Phys.* 1993, 98, 10089–10092.
- (61) Kabsch, W. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 2010, 66, 125–132.

(62) (a) 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. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 2011, 67, 235–242. (b) Collaborative Computational Project. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 1994, 50, 760–763.

(63) Emsley, P.; Cowtan, K. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 2004, 60, 2126–2132.

(64) 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. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 2011, 67, 355–367.

(65) Laskowski, R. A.; Moss, D. S.; Thornton, J. M. *J. Mol. Biol.* 1993, 231, 1049–1067.

## Article IV: The Use of Fluoroproline in MUC1 Antigen Enables Efficient Detection of Antibodies in Patients with Prostate Cancer

Víctor J. Somovilla,<sup>†,‡</sup> Iris A. Bermejo,<sup>†</sup> Inês S. Albuquerque,<sup>§</sup> Nuria Martínez-Sáez,<sup>†,‡</sup> Jorge Castro-López,<sup>||</sup> Fayna García-Martín,<sup>⊥</sup> Ismael Compañón,<sup>†</sup> Hiroshi Hinou,<sup>⊥</sup> Shin-Ichiro Nishimura,<sup>⊥</sup> Jesús Jiménez-Barbero,<sup>#</sup> Juan L. Asensio,<sup>¶</sup> Alberto Avenoza,<sup>†</sup> Jesús H. Busto,<sup>†</sup> Ramón Hurtado-Guerrero,<sup>||,△</sup> Jesús M. Peregrina,<sup>†</sup> Gonçalo J. L. Bernardes,<sup>\*,§,▽</sup> and Francisco Corzana<sup>\*,†</sup>

<sup>†</sup>Departamento de Química, Universidad de La Rioja, Centro de Investigación en Síntesis Química, 26006 Logroño, Spain.

<sup>‡</sup>Department of Chemical Biology and Drug Discovery, Utrecht Institute for Pharmaceutical Sciences, Bijvoet Center for Biomolecular Research, Utrecht University, Universiteitsweg 99, Utrecht, The Netherlands.

<sup>§</sup>Instituto de Medicina Molecular, Faculdade de Medicina da, Universidade de Lisboa, Avenida Professor Egas Moniz, 1649-028, Lisboa, Portugal.

<sup>||</sup>Institute of Biocomputation and Physics of Complex Systems (BIFI), University of Zaragoza, BIFI-IQFR (CSIC), Zaragoza, Spain.

<sup>⊥</sup>Graduate School and Faculty of Advanced Life Science, Field of Drug Discovery Research, Hokkaido University, N21 W11, Sapporo 001-0021, Japan.

<sup>#</sup>(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.

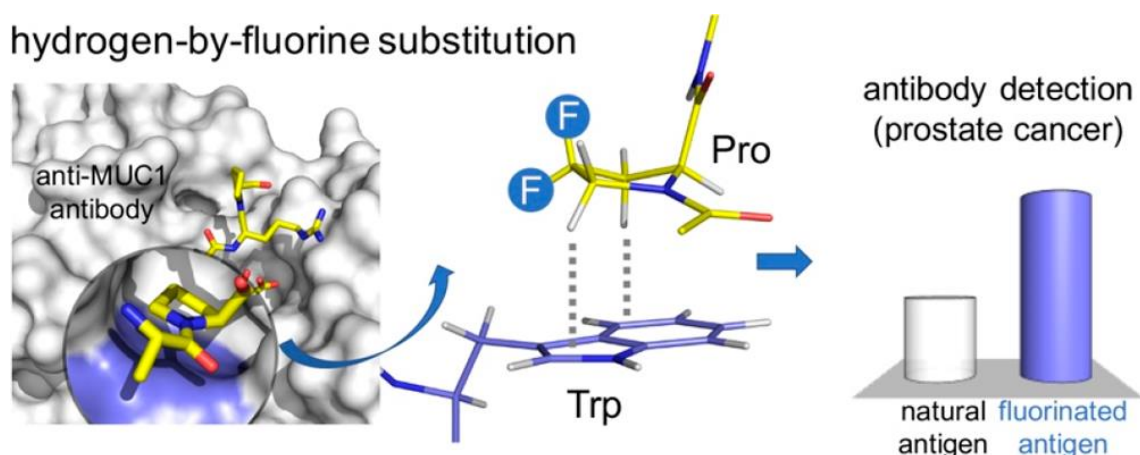
<sup>¶</sup>Instituto de Química Orgánica General, IQOG-CSIC, 28006 Madrid, Spain.

<sup>△</sup>Fundación ARAID, 50018 Zaragoza, Spain.

<sup>▽</sup>Department of Chemistry, University of Cambridge, Lensfield Road, CB2 1EW Cambridge, U.K.

### 3.12. 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 (4S)-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.





### 3.13. INTRODUCTION

MUC1 is a glycoprotein overexpressed in around 80% of human cancers.<sup>1-3</sup> 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.<sup>4-7</sup> Because of this unique feature, extensive efforts have been made toward the development of cancer vaccines based on MUC1 fragments.<sup>8-12</sup> 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.<sup>13-18</sup> Consequently, efforts have 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,<sup>19</sup> 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.<sup>20</sup> 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,<sup>21,22</sup> 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,<sup>23</sup> 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( $\alpha$ -O-GalNAc)RP in complex with the SM3,<sup>22</sup> which is a therapeutic antibody used in the treatment of cancer.<sup>21</sup> 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.<sup>24</sup> Interestingly, recent studies have shown that CH/ $\pi$  bonds can be significantly enhanced by simply increasing the polarization of the interacting CH moieties.<sup>25–28</sup> 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 (4S)-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/ $\pi$  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).

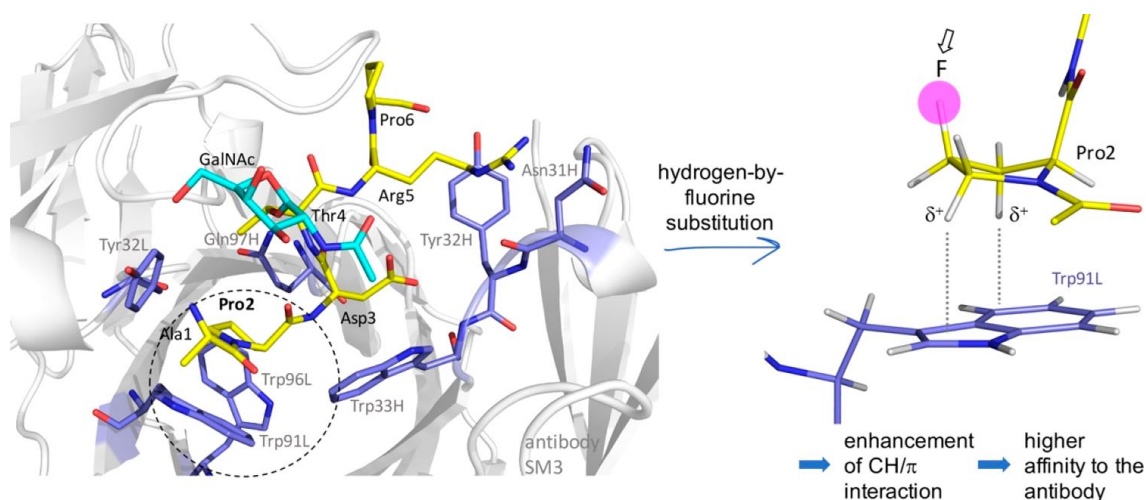


Figure 1. Hydrogen-by-fluorine Pro replacement strategy to improve antibody–antigen affinity. Crystal structure of glycopeptide APDT( $\alpha$ -O-GalNAc)RP in complex with antibody SM3 (pdb ID: 5a2k) together with the strategy proposed in this work to design effective antigens based on MUC1.

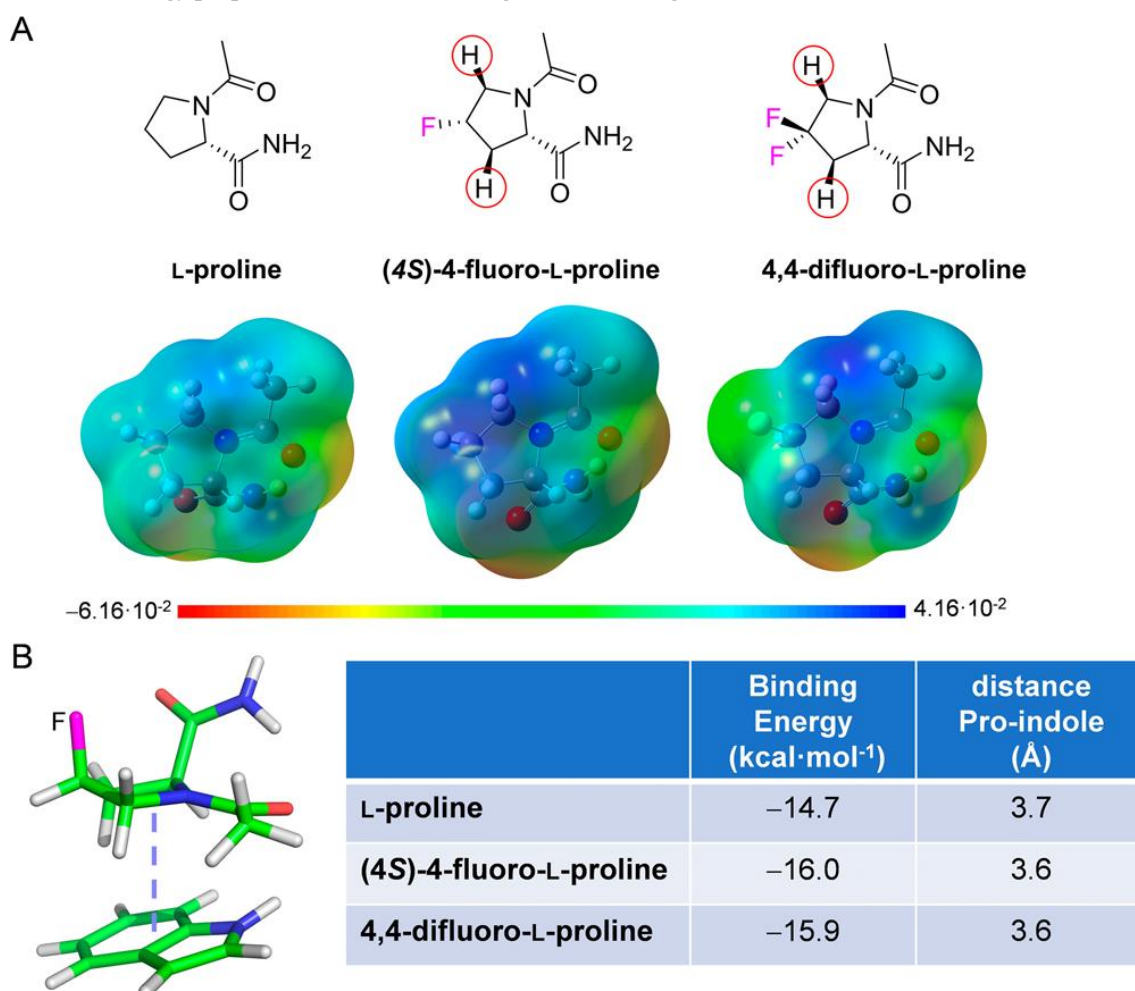


Figure 2. Effect of fluorine atoms on the electrostatic potential and on CH/ $\pi$  stability. (A) Electrostatic potential surfaces (in au) calculated at the Mo6-2X/6-31+G(d,p) level in a vacuum, showing the re faces of the Pro derivatives. Blue/red indicates positive/negative potentials. (B) Binding energy calculated at the Mo6-2X/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.

### 3.14. RESULTS AND DISCUSSION

To provide theoretical support for the hypothesis depicted in Figure 1, we calculated the electrostatic potential surface of the 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.<sup>29</sup> 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 monosubstituted derivative (4S)-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<sup>30</sup> (Supporting Information).

We then evaluated the impact of the hydrogen-by-fluorine substitutions on the antigen association to antibody scFv-SM3<sup>22</sup> 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).<sup>22,23</sup> This trend is also noted for the non-natural derivatives. To our delight, surrogates fP\* and 2fP\*, which contain a (4S)-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<sup>31,32</sup> to determine the affinity of these MUC1 variants with commercially available antibodies SM3 and VU-3C6<sup>23</sup> (Figure 4B). The results obtained from the microarray experiments were in line with the BLI data, showing a clear increase

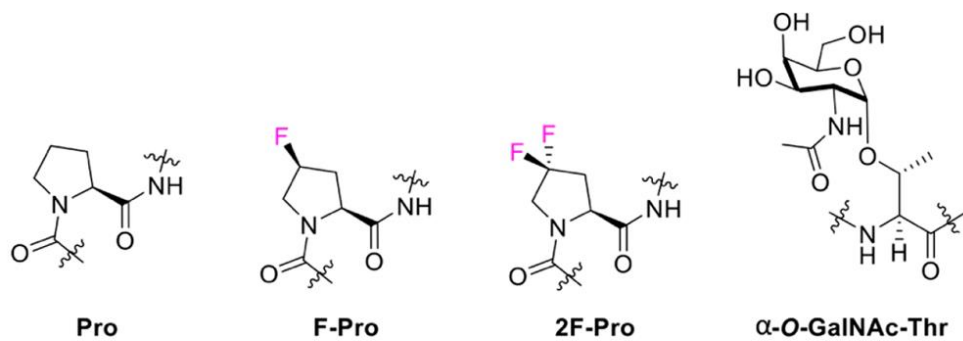
in binding affinity for the non-natural variants. This result may indicate that both SM3 and VU-3C6 antibodies recognize the antigen with an equivalent binding mode.

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 GVTSAfPDT\*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( $\alpha$ -O-GalNAc)RP in complex with the same antibody.<sup>21,22</sup> 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.<sup>21,22</sup> 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  $\sim 68$  and  $91^\circ$ , respectively. This conformation is similar to that found for a Tn-glycopeptide in complex with SM3 (pdb ID: 5a2k)<sup>22</sup> 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 (4S)-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 the improvement in the proline/tryptophan stacking accounts for the increased stability of the complexes formed by the nonnatural 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 =  $4.0 \pm 0.4$  Å), in agreement with the X-ray structure described in Figure 5, and for compound 2fP\*' (distance Pro-Trp91L =  $4.1 \pm 0.6$  Å). In contrast, the distance found for the natural glycopeptide P\*' was  $4.7 \pm 0.4$  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,<sup>19</sup> 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.



*MUC1 tandem repeat fragment*

**P**, X=Pro, Y=Thr  
**P\***, X=Pro, Y= $\alpha$ -O-GalNAc-Thr  
**fP**, X=F-Pro, Y=Thr  
**fP\***, X=F-Pro, Y= $\alpha$ -O-GalNAc-Thr  
**2fP**, X=2F-Pro, Y=Thr  
**2fP\***, X=2F-Pro, Y= $\alpha$ -O-GalNAc-Thr

Figure 3. MUC1-like peptides and glycopeptides synthesized and studied in this work.

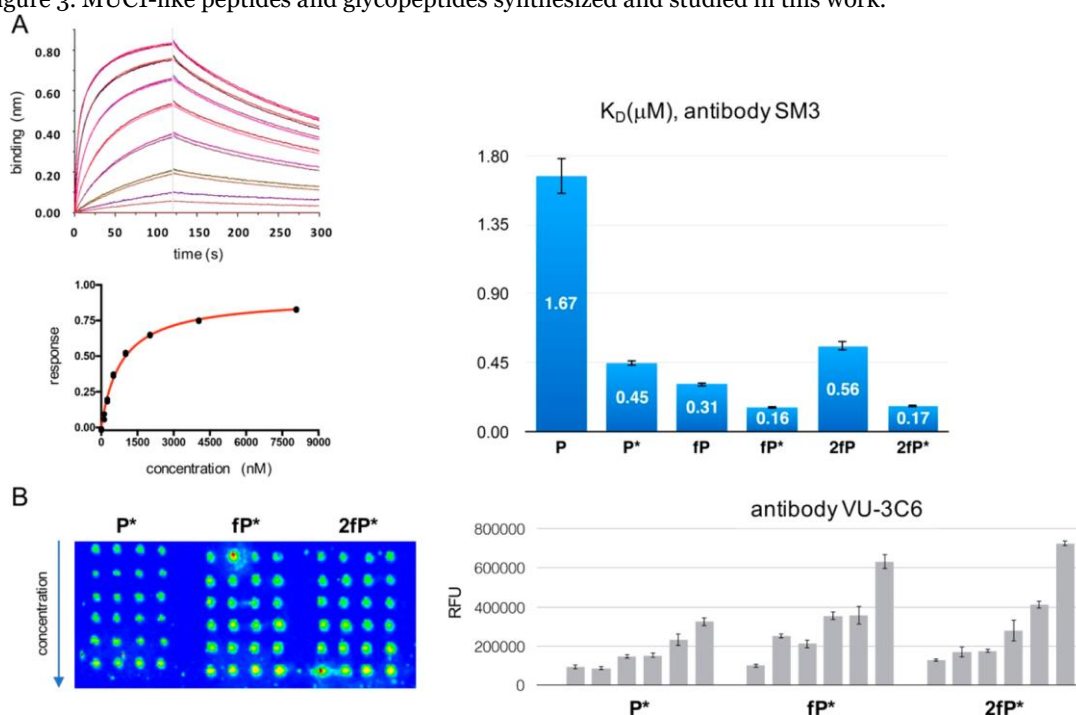


Figure 4. Binding of the glycopeptides to anti-MUC1 antibodies. (A) BLI curves and fit obtained for glycopeptide fP\* and antibody scFv-SM3, together with the  $K_D$  constants derived from BLI experiments for all studied MUC1-related compounds. (B) Interaction of the anti-MUC1 antibody VU-3C6 with the glycopeptides using a microarray platform. Compounds were printed onto an aminoxy-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).

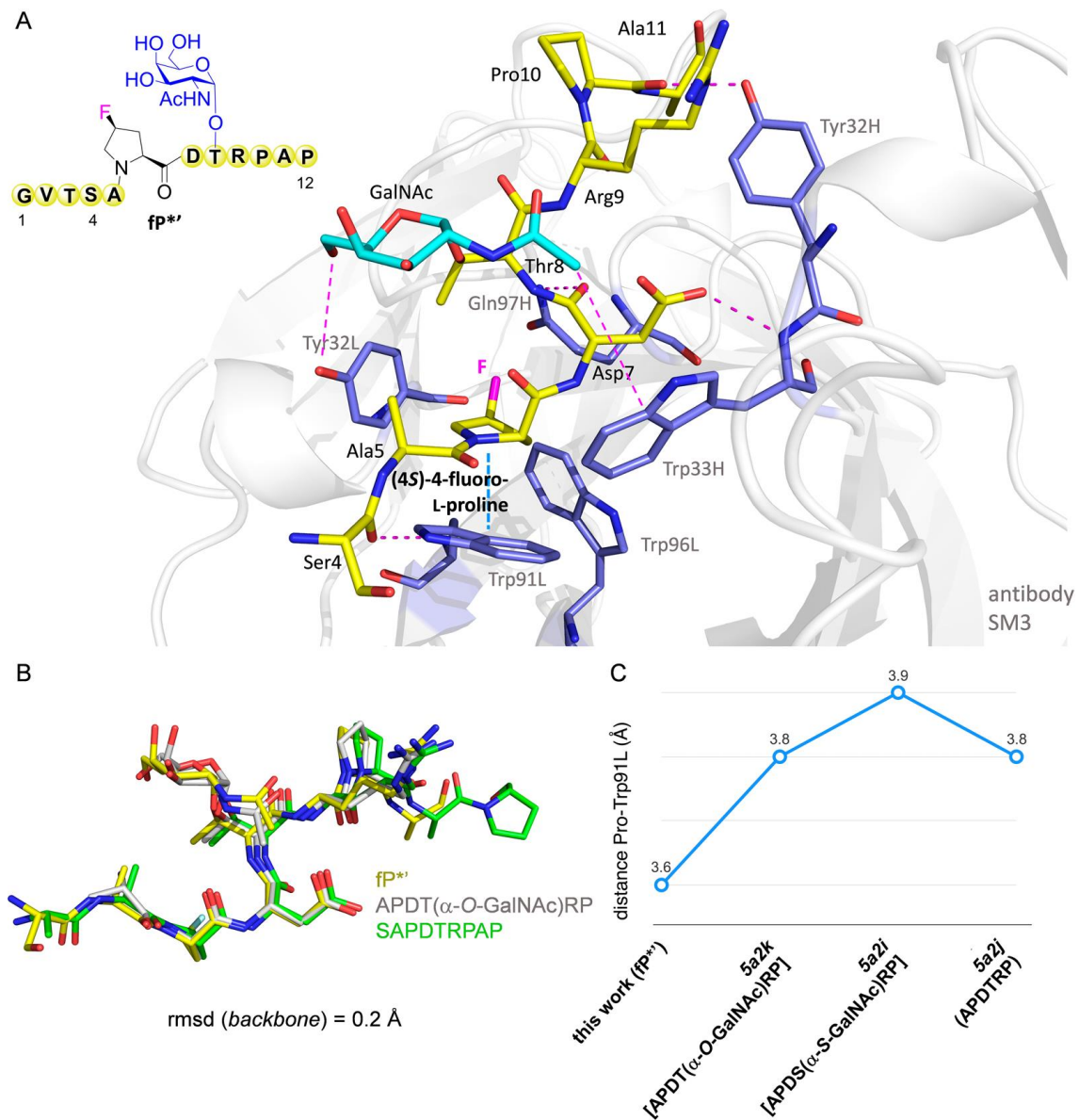


Figure 5. Crystal structure of fP\* bound to scFv-SM3. (A) Key binding interactions of glycopeptide fP\* with scFv-SM3 mAb, as observed in the X-ray crystal structure (pdb ID: 5OWP). Peptide backbone carbon atoms are shown in yellow. GalNAc carbon atoms are shown in cyan. Carbon atoms of key residues 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( $\alpha$ -O-GalNAc)RP, together with SAPDTRPAP peptide<sub>21</sub> 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.



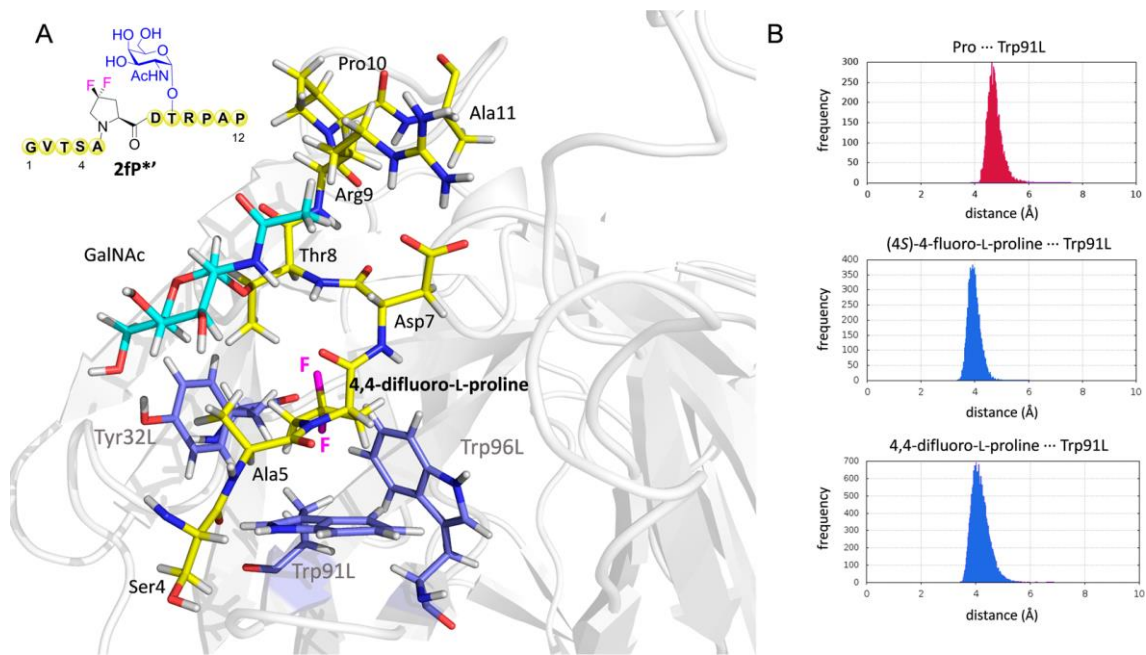


Figure 6. MD simulations of glycopeptide 2fP\* in complex with scFv-SM3. (A) Representative frame of the 200 ns MD simulations performed on 2fP\* in complex with antibody scFv-SM3. Only peptide fragment Ser4-Ala11 is shown for clarity. Peptide backbone carbon atoms are shown in yellow. GalNAc carbon 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.

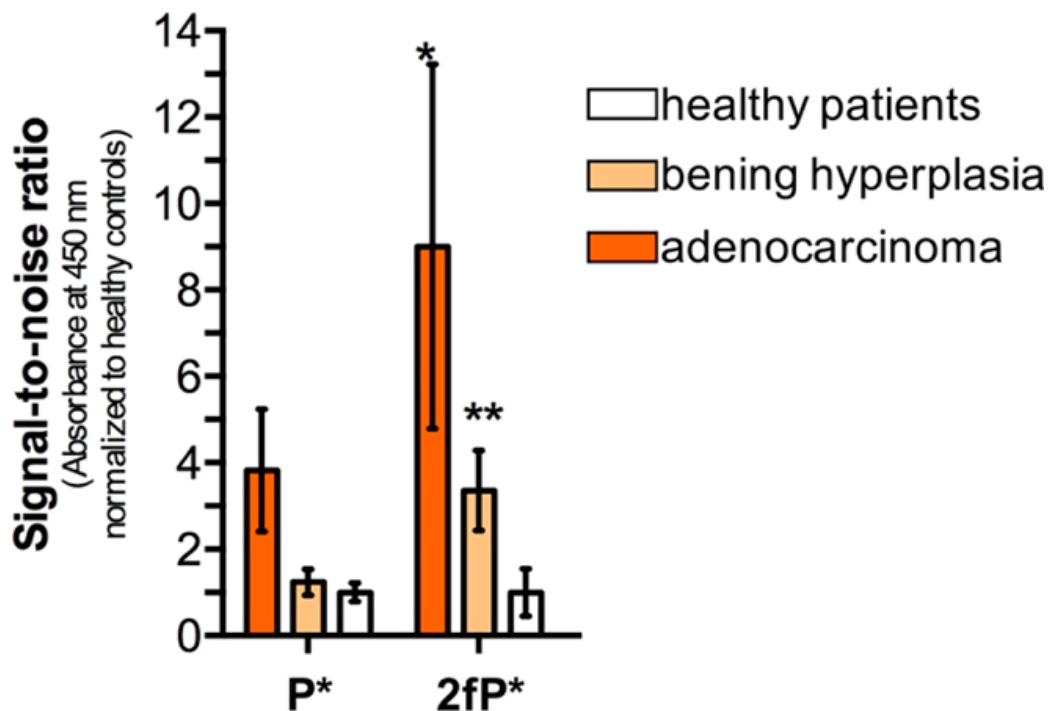


Figure 7. Detection of circulating anti-MUC1 antibodies in serum of patients presenting benign and malignant prostate cancer. Binding affinity of human circulating antibodies against MUC1 (P\*) and synthetic variant 2fP\*, in the context of prostate adenocarcinoma and benign prostatic hyperplasia. Bars show average  $\pm$  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).

### 3.15. 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 (4S)-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 display a significant potential as diagnostic tools to detect anti-MUC1 antibodies in prostate cancer patients.

## ASSOCIATED CONTENT

### \*<sub>s</sub>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 Io4 (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.; Buetti, 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.; Snijdwint, 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-González, J.; Madariaga, D.; Compañón, I.; Somovilla, V. J.; Salvadó, M.; Asensio, J. L.; Jiménez-Barbero, J.; Avenoza, A.; Busto, J. H.; Bernardes, G. J. L.; Peregrina, J. M.; Hurtado-Guerrero, R.; Corzana, F. *Angew. Chem., Int. Ed.* 2015, 54, 9830–9834.
- (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.

## Article V: Water Sculpts the Distinctive Shapes and Dynamics of the Tumor-Associated Carbohydrate Tn Antigens: Implications for Their Molecular Recognition

Iris A. Bermejo,<sup>†,□</sup> Imanol Usabiaga,<sup>‡,□</sup> Ismael Compañón,<sup>†,□</sup> Jorge Castro-López,<sup>§</sup> Aran Insausti,<sup>‡,¶</sup> José A. Fernández,<sup>‡</sup> Alberto Avenoza,<sup>†</sup> Jesús H. Busto,<sup>†</sup> Jesús Jiménez-Barbero,<sup>||,⊗,∇</sup> Juan L. Asensio,<sup>⊥</sup> Jesús M. Peregrina,<sup>†</sup> Gonzalo Jiménez-Osés,<sup>†</sup> Ramón Hurtado-Guerrero,<sup>\*,§,#</sup> Emilio J. Cocinero,<sup>\*,‡,¶</sup> and Francisco Corzana<sup>\*,†</sup>

<sup>†</sup>Departamento de Química, Centro de Investigación en Síntesis Química, Universidad de La Rioja, 26006 Logroño, Spain.

<sup>‡</sup>Departamento de Química Física, Facultad de Ciencia y Tecnología, Universidad del País Vasco (UPV-EHU), 48080 Bilbao, Spain.

<sup>§</sup>Institute of Biocomputation and Physics of Complex Systems (BIFI), University of Zaragoza, BIFI-IQFR (CSIC), 50018 Zaragoza, Spain.

<sup>||</sup>CIC bioGUNE, Bizkaia Technology Park, Building 801A, 48170 Derio, Spain.

<sup>⊗</sup>Ikerbasque, Basque Foundation for Science, Maria Diaz de Haro 13, 48009 Bilbao, Spain.

<sup>∇</sup>Department of Organic Chemistry II, Faculty of Science & Technology, University of the Basque Country, 48940 Leioa, Spain.

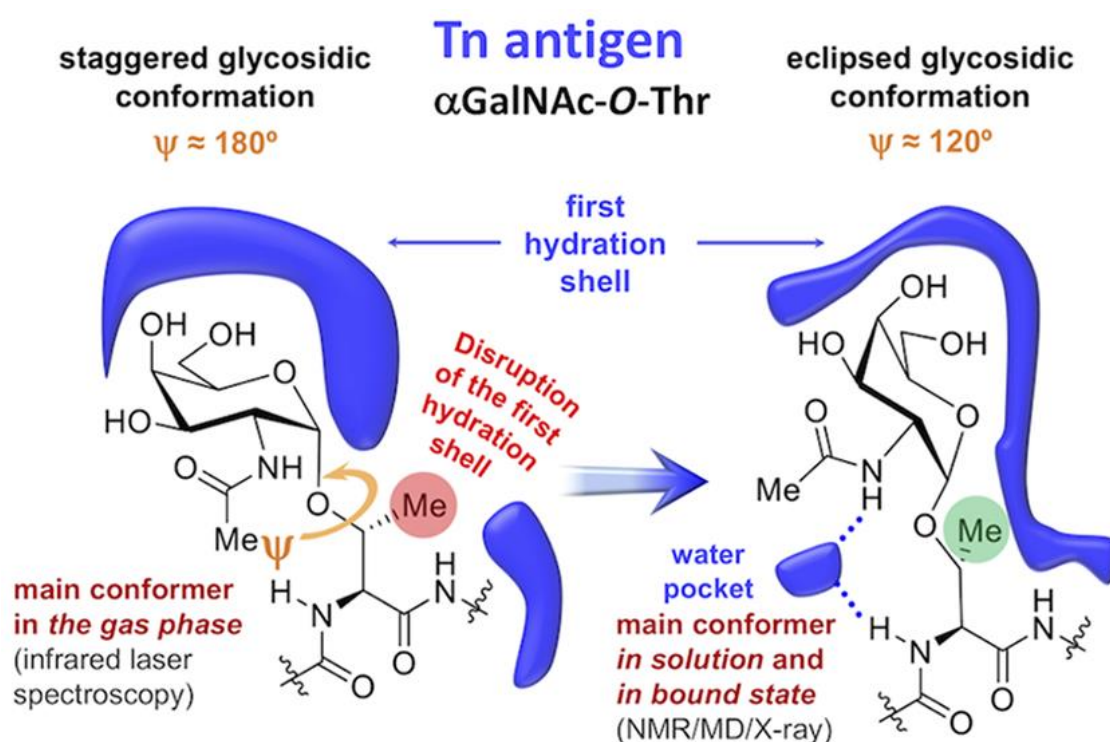
<sup>⊥</sup>Instituto de Química Orgánica General, IQOG-CSIC. 28006 Madrid, Spain.

<sup>#</sup>Fundación ARAID, 50018 Zaragoza, Spain.

<sup>¶</sup>Biofisika Institute (CSIC, UPV/EHU), University of the Basque Country (UPV/EHU), 48940 Leioa, Spain.

### 3.16. ABSTRACT

The tumor-associated carbohydrate Tn antigens include two variants,  $\alpha$ GalNAc-O-Thr and  $\alpha$ GalNAc-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  $\alpha$ GalNAc-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.





### 3.17. INTRODUCTION

The Tn antigens are among the most specific human tumor-associated carbohydrate antigens (TACAs), present in approximately 90% of tumors.<sup>1,2</sup> In general, the aggressiveness of the carcinoma and the occurrence of these antigens has a clear direct correlation,<sup>3</sup> promoting their use as biomarkers and potential therapeutic targets against cancer.<sup>4</sup> 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.<sup>5</sup>

In general, the Tn antigen is referred to as GalNAc  $\alpha$ -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.<sup>6-12</sup>

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.<sup>6</sup> On the other hand, several anti-Tn antibodies show a clear preference for glycopeptides containing the Tn-Ser antigen.<sup>13</sup> A recent study 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.<sup>9</sup> In the context of antifreeze glycoproteins, which consist of the tandem repeating (Thr-Ala-Ala)<sub>n</sub> polypeptide glycosylated with Gal  $\beta$ -(1,3)-GalNAc, the replacement of a threonine by a serine residue eliminates the antifreeze activity.<sup>14</sup> 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.<sup>4</sup>

Our conformational analysis based on NMR data combined with experiment-guided Molecular Dynamics (MD) simulations,<sup>15</sup> showed that Tn-Thr (Figure 1b) is rather rigid in solution,<sup>8</sup> with its O-glycosidic linkage in the so-called “eclipsed” conformation ( $\phi \approx 80^\circ$ ,  $\psi \approx 120^\circ$ ), and its side chain fixed at a

$\chi_1$  torsional angle value around  $60^\circ$  (Figure 1a). Conversely, Tn-Ser (Figure 1b) displays the typical *exo*-anomeric/*syn* conformation for the glycosidic linkage with  $\phi$  and  $\psi$  values  $\sim 80^\circ$  and  $\sim 180^\circ$ , respectively. This latter system is more flexible and exhibits the three possible staggered conformers for the side chain in solution.<sup>7</sup> 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.<sup>8</sup> 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  $^{15}\text{N}$ - and  $^{13}\text{C}$ -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.<sup>8</sup> 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.

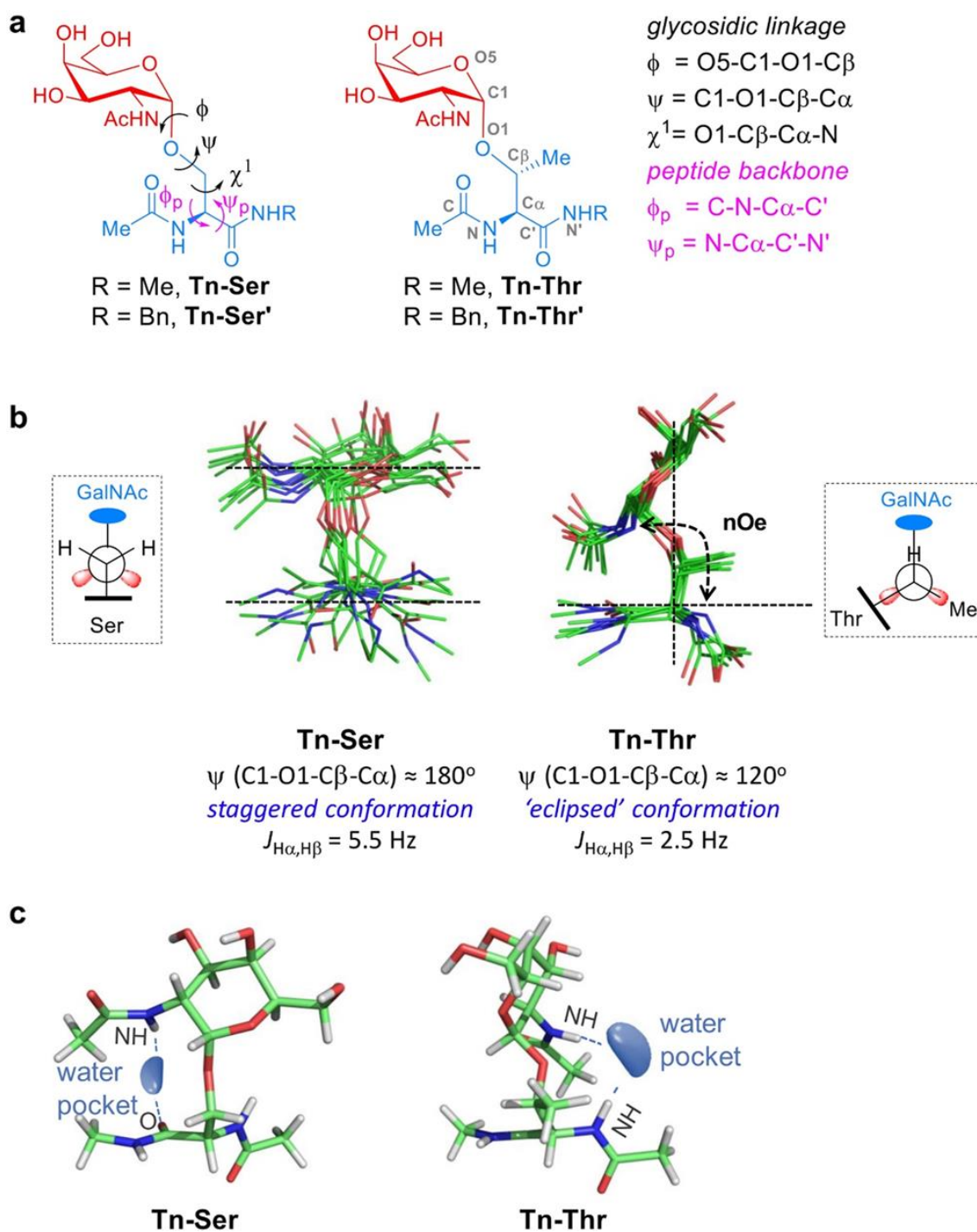


Figure 1. Conformational behavior of the antigens Tn-Ser and Tn-Thr in water. (a) The two Tn-antigens studied in this work, together with the definition of the most relevant torsional angles and atom labels. (b) Major conformations in solution for the Tn antigen with either a serine (Tn-Ser) or a threonine (Tn-Thr) derived from experiment-guided molecular dynamics (MD) simulations.<sup>7,8</sup> The values of the  $\psi$  torsion angle of the glycosidic linkage and of the associated coupling constant  $J_{\text{H}\alpha,\text{H}\beta}$  are shown. The Newman projections for the  $\text{C}\beta\text{-O1}$  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.<sup>7,8</sup> Antigens Tn-Ser and Tn-Thr accommodate different water pockets owing to their distinct conformational behavior in solution.

### 3.18. 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.<sup>16–18</sup> 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<sup>15</sup> (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  $\sim 3370$  cm<sup>-1</sup> 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

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  $158 \pm 11^\circ$  for the Tn-Ser' antigen and  $153 \pm 12^\circ$  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  $\sim 3200\text{--}3350\text{ cm}^{-1}$  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 $\cdots$ O=C) is displaced about  $300\text{ cm}^{-1}$  and it is broadened around  $200\text{ cm}^{-1}$ . Although NH groups display similar behavior, they display narrower bands and smaller displacements than OH groups.<sup>16b, c</sup> In this case, the acetamide NH of GalNAc (strong NH<sub>GalNAc</sub> $\cdots$ O=C) is displaced around  $100\text{ cm}^{-1}$ , and it is broadened about  $100\text{ cm}^{-1}$ . The remaining hydroxyl groups of the sugar are engaged in weak hydrogen bonds, characterized by bands within the  $3500\text{--}3700\text{ cm}^{-1}$  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,<sup>9</sup> 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.<sup>8</sup> 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  $\psi \approx 165^\circ$ ), 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<sub>2</sub>O 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  $\psi \approx 120^\circ$  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 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  $60^\circ$  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 reported,<sup>6a</sup> the structure of the complex between the Thr containing glycopeptide 1 (Figure 4a) when bound to an anti-MUC1 antibody (SM3)<sup>21</sup> 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.<sup>8</sup> 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<sup>6b</sup> (compounds 2 and 3 in Figure 4a; see 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<sup>6,22</sup> 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<sub>22</sub> 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.<sup>8,25</sup> 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 residue in the fluorinated variants, and is similar to that found in the crystal structure reported for the naked peptide<sup>21</sup> (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 in the solid state for glycopeptides 2



and 3, with  $\psi$  close to  $120^\circ$  (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  $90^\circ$ . 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  $120^\circ$  in solution.<sup>6</sup> Moreover, the glycosidic linkage of Tn-Thr antigen in complex with other proteins/enzymes can adopt variable values for  $\psi \approx 90^\circ$  to  $130^\circ$  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<sub>2</sub>F and CHF<sub>2</sub> 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,<sup>26–28</sup> 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<sup>29</sup> 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.

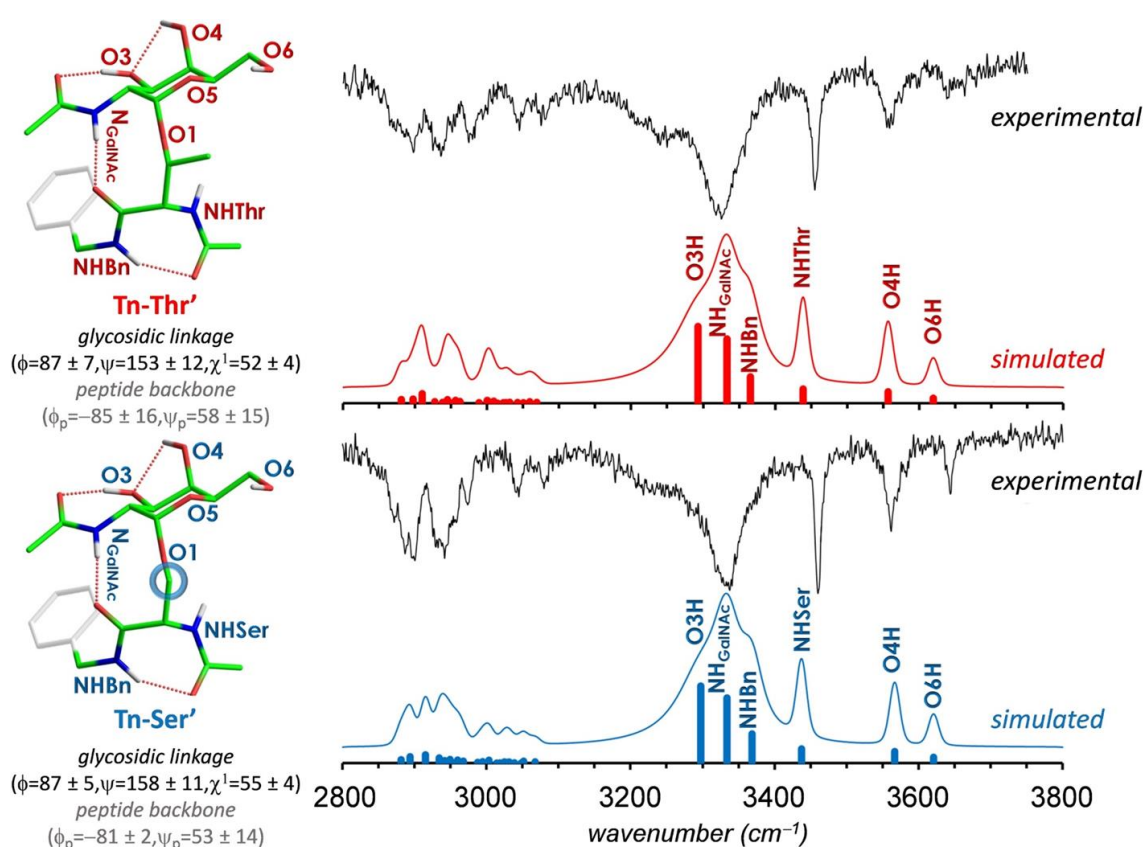


Figure 2. Infrared ion-dip (IRID) spectra of antigens Tn-Ser' and Tn-Thr' in the gas phase. Experimental and simulated infrared ion-dip (IRID) spectra of derivatives Tn-Thr' (upper panel) and Tn-Ser' (lower panel), together with representative minimum free energy conformers calculated for both compounds at the M06-2X/6-31+G(d) level.<sup>20</sup> 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).

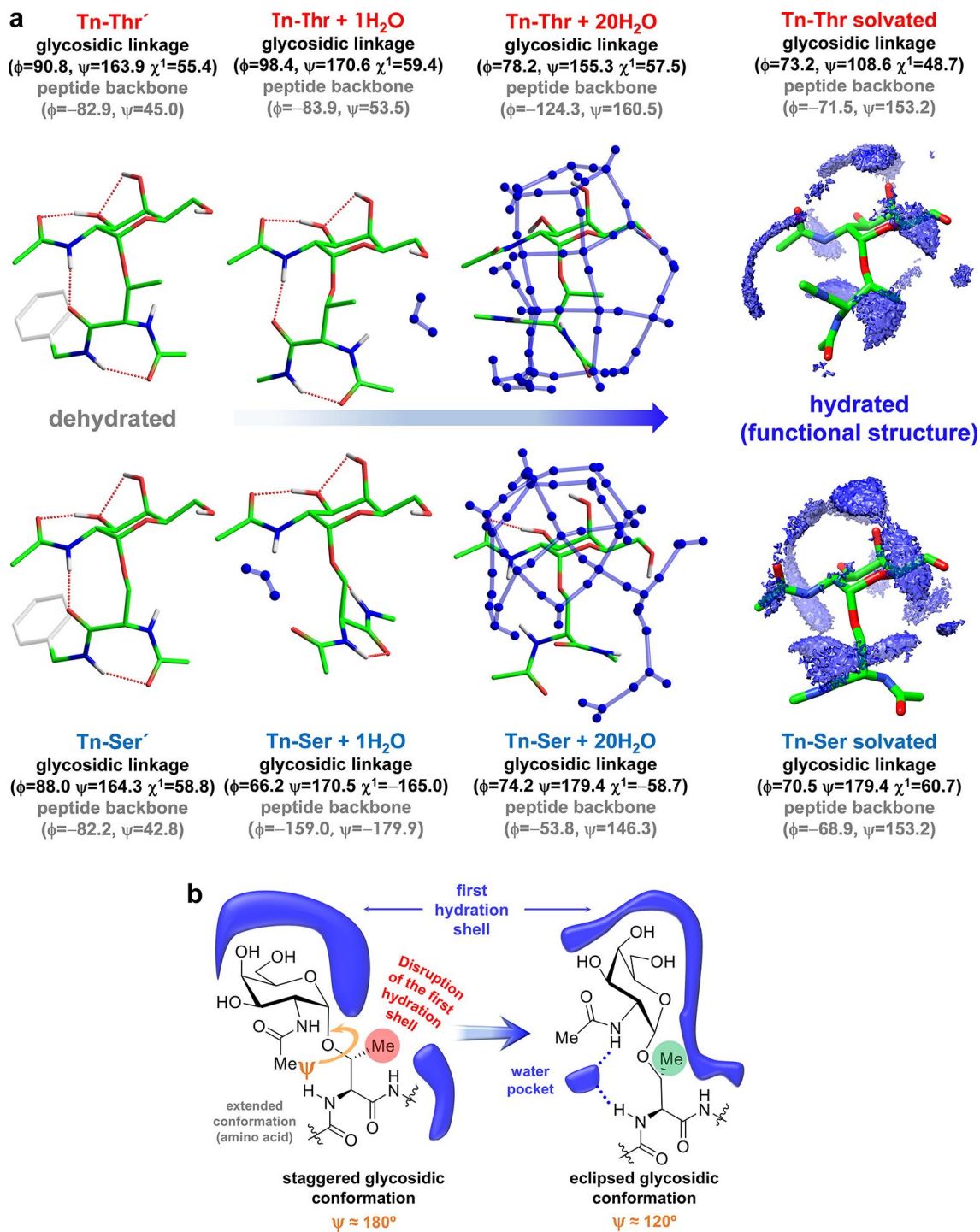


Figure 3. Gradual solvation and first hydration shell of the Tn antigens. (a) Lowest energy conformers calculated at the Mo6-2X/6-31+G(d) level<sup>20</sup> for Tn-Thr (upper panel) and Tn-Ser (lower panel) with discrete water molecules (1 and 20), together with the averaged first hydration shell derived 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.

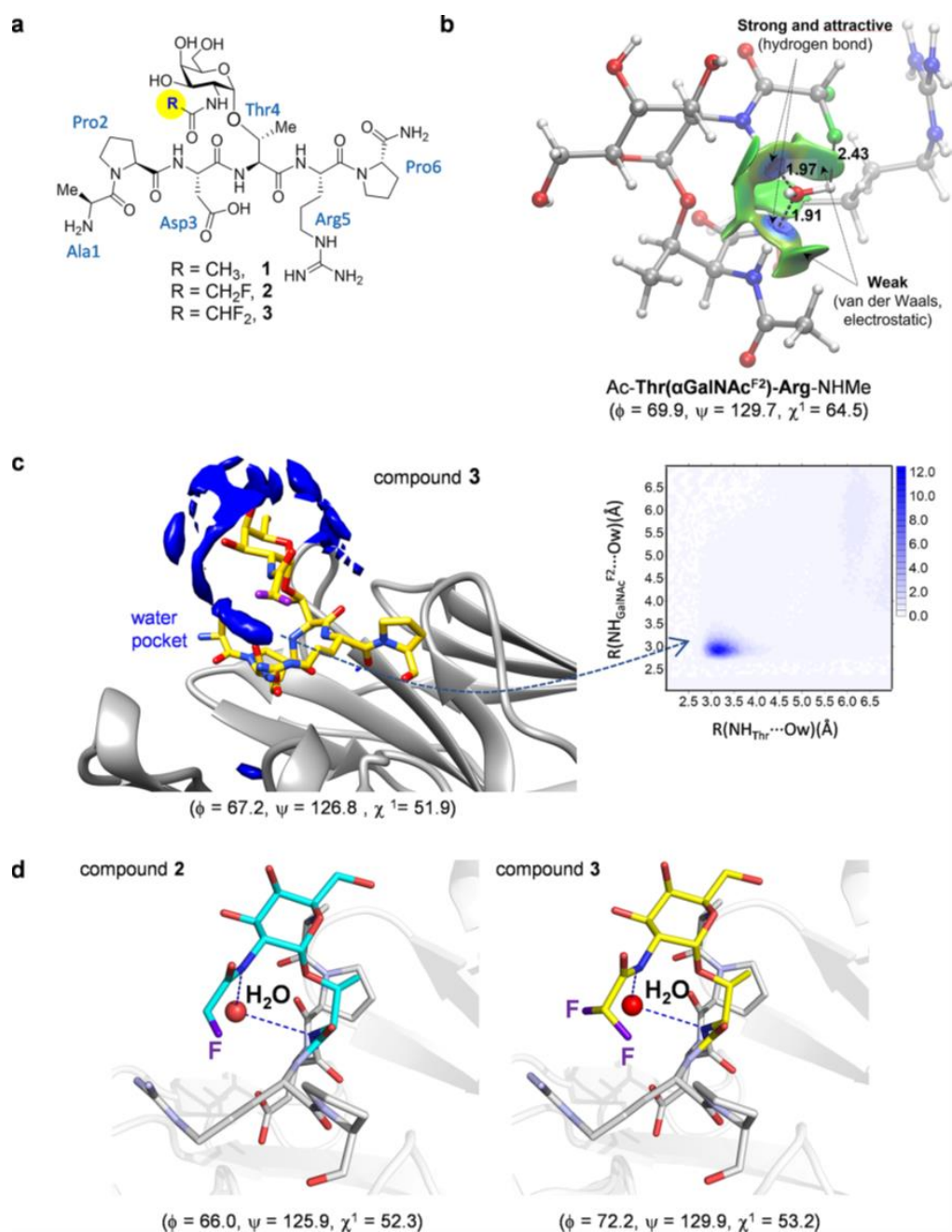


Figure 4. Use of fluorinated derivatives to trap bridging water molecules in the solid state. (a) Natural and fluorinated glycopeptides synthesized in this work, comprising the peptide epitope recognized by anti-MUC1 antibodies.<sup>22</sup> (b) Non-covalent interactions for the water molecule bound to the hydrophilic pocket of model glycopeptide Ac-Thr(αGalNAcF<sub>2</sub>)-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.<sup>20,23</sup> (c) Representation of the first hydration shell around the fluorinated Tn antigen derived from 200 ns MD simulations performed on glycopeptide 3 in the SM<sub>3</sub>-bound state. The 2D radial distribution function<sup>24</sup> 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-SM<sub>3</sub> 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).

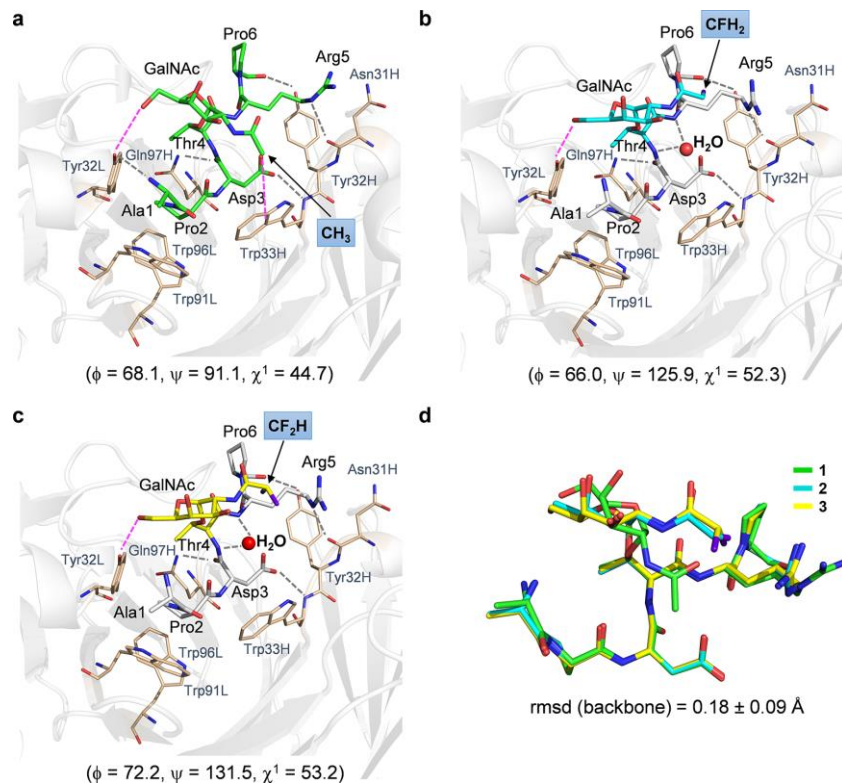


Figure 5. Analysis of the X-ray structures of glycopeptides 1 (ref 6), 2, and 3 in complex with scFv-SM3. Key binding interactions of glycopeptides 1 (a), 2 (b), and 3 (c) with the antibody, as observed in the X-ray crystal structures (PDB IDs: 5A2K, 6FZR, and 6FZQ, respectively). Pink dashed lines indicate 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.

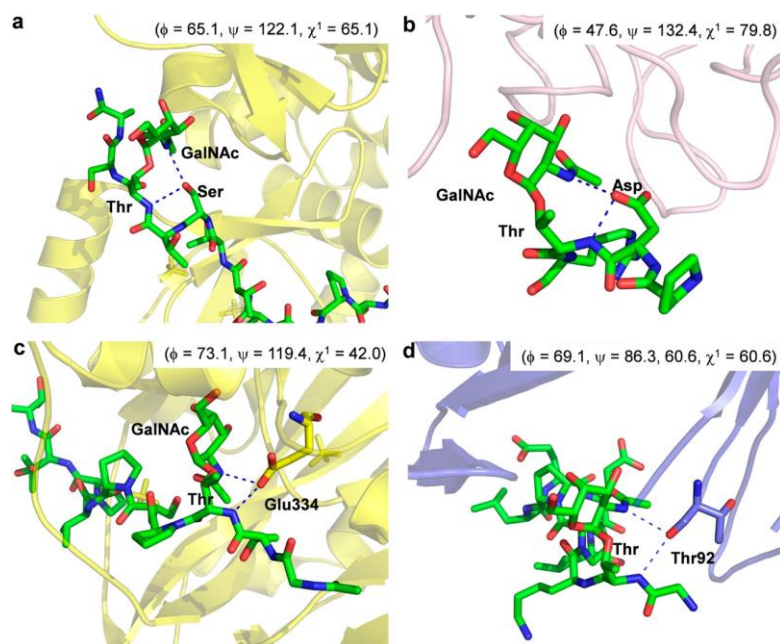


Figure 6. Molecular recognition of glycopeptides bearing the Tn-Thr antigen. The bridging water molecule is replaced by an atom of either the glycopeptide or the receptor. (a) Zoom-in of the crystal structure of the active form of GalNAc-Transferase 2 (GalNAc-T2) in complex with UDP and the glycopeptide MUC5AC-13, showing the lectin domain (PDB ID: 5AJP).<sup>27</sup> (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).<sup>30</sup> (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).<sup>27</sup> (d) Zoom-in of the crystal structure of the antibody 237 in complex with its glycopeptide epitope (PDB ID: 3IET).<sup>28</sup> In all cases, the geometry of the glycosidic linkage is shown in parentheses.

### 3.19. 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<sub>β</sub> disturbs the proper solvation of the “native” gas-phase 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,<sup>6,13</sup> enzymes (GalNAc-transferases),<sup>26,27</sup> and lectins.<sup>30</sup> 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,<sup>31-33</sup> 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

### \*<sub>s</sub>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.; Heimbürg-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-González, J.; Madariaga, D.; Compañón, I.; Somovilla, V. J.; Salvadó, M.; Asensio, J. L.; Jiménez-Barbero, J.; Avenoza, A.; Busto, J. H.; Bernardes, G. J. L.; Peregrina, J. M.; Hurtado-Guerrero, R.; Corzana, F. *Angew. Chem., Int. Ed.* 2015, 54, 9830–9834. (b) Unione, L.; Alcalá, M.; Echeverría, 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-González, 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.; Rodríguez, 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.; Fukushima, 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.; García-Martin, F.; Marcelo, F. *J. Am. Chem. Soc.* 2015, 137, 12438–12441.
- (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.; Vocado, 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-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.
- (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.

Article VI: Structure-Based Design of Potent Tumor-Associated Antigens:  
Modulation of Peptide Presentation by Single-Atom O/S or O/Se  
Substitutions at the Glycosidic Linkage

Ismael Compañón,<sup>†,+</sup> Ana Guerreiro,<sup>‡,+</sup> Vincenzo Mangini,<sup>§,+</sup> Jorge Castro-López,<sup>||</sup> Margarita Escudero-Casao,<sup>⊥</sup> Alberto Avenzoa,<sup>†</sup> Jesús H. Busto,<sup>†</sup> Sergio Castellón,<sup>⊥</sup> Jesús Jiménez-Barbero,<sup>#,∇,¶</sup> Juan L. Asensio,<sup>•</sup> Gonzalo Jiménez-Osés,<sup>†,#</sup> Omar Boutureira,<sup>⊥</sup> Jesús M. Peregrina,<sup>†</sup> Ramón Hurtado-Guerrero,<sup>||</sup> Roberto Fiammengo,<sup>\*</sup> §Gonçalo J. L. Bernardes,<sup>\*,‡,°</sup> and Francisco Corzana<sup>\*,†</sup>

<sup>†</sup>Departamento de Química, Universidad de La Rioja, Centro de Investigación en Síntesis Química, 26006 Logroño, Spain

<sup>‡</sup>Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, Avenida Professor Egas Moniz, 1649-028 Lisboa, Portugal

<sup>§</sup>Center for Biomolecular Nanotechnologies@UniLe, Istituto Italiano di Tecnologia (IIT), 73010 Arnesano, Lecce, Italy

<sup>||</sup>Institute of Biocomputation and Physics of Complex Systems (BIFI), University of Zaragoza, BIFI-IQFR (CSIC), Fundación

ARAID, 50018 Zaragoza, Spain

<sup>⊥</sup>Departament de Química Analítica i Química Orgànica, Facultat de Química, Universitat Rovira i Virgili, 43007 Tarragona, Spain

<sup>#</sup>CIC bioGUNE, Bizkaia Technology Park, Building 801A, 48170 Derio, Spain

<sup>∇</sup>Ikerbasque, Basque Foundation for Science, Maria Diaz de Haro 13, 48009 Bilbao, Spain

<sup>¶</sup>Department of Organic Chemistry II, Faculty of Science & Technology, University of the Basque Country, 48940 Leioa, Spain

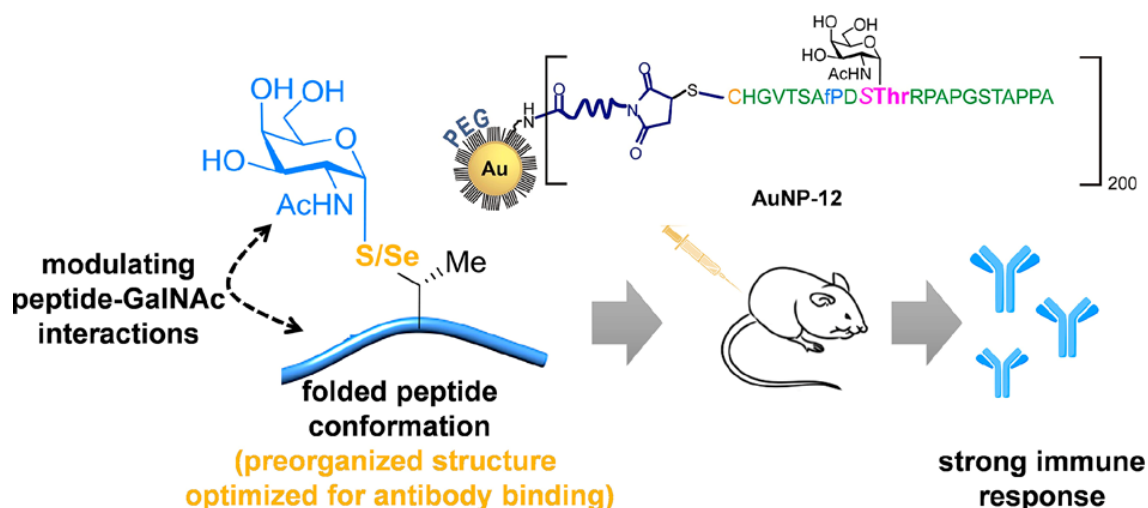
<sup>•</sup>Instituto de Química Orgánica General, IQOG-CSIC, 28006 Madrid, Spain

<sup>°</sup>Department of Chemistry, University of Cambridge, Lensfield Road, CB2 1EW Cambridge, U.K.

### 3.20. ABSTRACT:

GalNAc-glycopeptides derived from mucin MUC1 are an important class of tumor-associated antigens.  $\alpha$ -Oglycosylation forces the peptide to adopt an extended conformation in solution, which is far from the structure observed in complexes with a model anti-MUC1 antibody. Herein, we propose a new strategy for designing potent antigen mimics based on modulating peptide/carbohydrate interactions by means of O  $\rightarrow$  S/Se replacement at the glycosidic linkage. These minimal chemical modifications bring about two key structural changes to the glycopeptide. They increase the carbohydrate-peptide distance and change the orientation and dynamics of the glycosidic linkage. As a result, the peptide acquires a preorganized and optimal structure suited for antibody binding. Accordingly, these new glycopeptides display improved binding toward a representative anti-MUC1 antibody relative to the native antigens. To prove the potential of these glycopeptides as tumor-associated MUC1 antigen mimics, the derivative bearing the S-glycosidic linkage was conjugated to gold nanoparticles and tested as an immunogenic formulation in mice without any adjuvant, which resulted in a significant humoral immune response. Importantly, the mice antisera recognize cancer cells in biopsies of breast cancer patients with high selectivity. This finding demonstrates that the antibodies elicited against the mimetic antigen indeed recognize the naturally occurring antigen in its physiological context. Clinically, the exploitation of tumor-associated antigen mimics may contribute to the development of cancer vaccines and to the improvement of cancer diagnosis based on anti-MUC1 antibodies. The methodology presented here is of general interest for applications because it may be extended to modulate the affinity of biologically relevant glycopeptides toward their receptors.

### Structure-based design of potent MUC1-like antigens



#### 3.21. INTRODUCTION

MUC1 mucin is an O-glycoprotein overexpressed in many tumor tissues.<sup>1-4</sup> Although in healthy cells the backbone of this protein is decorated with complex glycans, in cancer cells this backbone carries rather simple and truncated oligosaccharides. Consequently, different tumor-associated carbohydrate antigens, such as the Tn determinant ( $\alpha$ -O-GalNAc-Ser/Thr),<sup>5</sup> are presented to the immune system and can be identified by anti-MUC1 antibodies. Peptide fragment Ala-Pro-Asp-Thr-Arg-Pro, which includes the immunodominant PDTRP region of MUC1 tandem repeats,<sup>6</sup> constitutes the minimum epitope recognized by these antibodies.<sup>7</sup> Partially glycosylated MUC1 derivatives have been used to prepare immunogenic formulations for the development of therapeutic cancer vaccines.<sup>8-12</sup> Similarly, unnatural glycopeptides that mimic tumor-associated MUC1 can find application as biosensors for the detection of cancerous cells.<sup>13</sup> An intriguing observation about the structural characteristics of these peptides is that  $\alpha$ -O-glycosylation with GalNAc forces the underlying peptide into an extended conformation in solution as a result of stabilizing interactions, which include direct<sup>14</sup> or water-mediated hydrogen bonds, between the peptide and the carbohydrate moiety (Figure 1A).<sup>15-17</sup> In contrast, the X-ray structure of the glycopeptide epitope bound to an anti-MUC1 antibody (SM3)<sup>18</sup> revealed a folded conformation around the glycosylated Thr (Figure 1A).<sup>19</sup> In this case, the sugar shifts the structure of the peptide in solution away from that adopted upon antibody binding. This

conformational entropic penalty is, however, compensated for by favorable enthalpic contributions (hydrogen bonds and CH/ $\pi$  stabilizing interactions)<sup>20,21</sup> between the sugar moiety and the antibody. As a result, a modest net increase in binding affinity (around 3-fold) is observed for the glycosylated versus the nonglycosylated peptide.

Herein, we propose a rational approach based on singleatom substitution (O  $\rightarrow$  S/Se) at the glycosidic linkage to obtain potent antigens with an improved affinity toward anti-MUC1 antibodies (Figure 1B). This simple modification increases the distance between the sugar and the peptide fragment—sulfur (or selenium) is larger than oxygen—which in turn minimizes the exo-anomeric effect<sup>22</sup> and alters the flexibility and the most stable conformation of the glycosidic linkage toward the one optimized for the antibody. Overall, these glycopeptides adopt a distinct structure in solution, which differs markedly from their oxygenated counterparts, thus avoiding the subsequent entropic cost associated with the extended-to-folded conformational transition of the glycopeptide in the bound state. In this work, we describe the strong binding of these glycopeptides to a model MUC1 antibody and demonstrate the possibility of using them as tumor-associated MUC1 mimics when they are incorporated into immunogenic formulations. In fact, the antibodies elicited in mice selectively recognize the naturally occurring tumor-associated MUC1 epitopes displayed on cancer cells in biopsies of breast cancer patients.

### 3.22. RESULTS AND DISCUSSION

Synthesis and Conformational Analysis of the Unnatural Glycopeptides in Solution. Glycopeptides 2\* and 3\* (Figure 2A) were designed to feature S-( $\alpha$ -D-GalNAc)-thiothreonine (SThr\*) and Se-( $\alpha$ -D-GalNAc)-selenothreonine (SeThr\*), respectively, as the fourth residue of the Ala-Pro-Asp-Xaa-Arg-Pro epitope. Naturally occurring glycopeptide 1\*, which has a threonine residue at this position, and nonglycosylated variants 1 and 2 were prepared for comparison. Although the synthesis of the amino acid thiothreonine (SThr) has been previously described,<sup>24–26</sup> the preparation of conveniently protected thiothreonine and selenothreonine derivatives as well as glycopeptides 2\* and 3\* has not yet been reported. As an example, the synthesis of building block 4, which is ready to be used in solid-phase peptide synthesis (SPPS), is shown in

Scheme 1. Conveniently protected threonine 5 was reacted with triphenylphosphine and iodine in the presence of imidazole as a base to afford iodo-derivative 6 with a total inversion of configuration at the  $\beta$ -carbon.<sup>27</sup> In parallel, selenosugar 9 was prepared in two steps from peracetylated compound 7. In the first step, 7 was treated with Woollin's reagent and pyridine to give selenazoline 8 in 70% yield. The hydrolysis of 8 with trifluoroacetic acid (TFA) in water afforded selenosugar 9 in moderate yield and as a dimer because of the formation of a diselenide bond. The key step in the synthesis of building block 4 is the nucleophilic attack of 9, previously reduced in situ with sodium borohydride, at iodo-derivative 6. This reaction proceeded in 51% yield with a total inversion of the configuration at the  $\beta$ -carbon of the selenothreonine surrogate while the  $\alpha$ - configuration at the anomeric carbon was completely preserved, as determined by <sup>1</sup>H NMR spectroscopy. (See the Synthesis section in the Supporting Information.) Subsequent deprotection steps gave the desired compound, 4, in 42% overall yield from 9. A similar strategy was used to prepare the building block of SThr\*. (See the Synthesis section in the Supporting Information.)

All (glyco)peptides were synthesized using microwaveassisted SPPS (MW-SPPS) by following our reported protocol<sup>13</sup> (Supporting Information). Next, we performed a thorough conformational analysis of unnatural glycopeptides 2\* and 3\* in solution by combining NMR spectroscopic measurements with molecular dynamics (MD) simulations (Figure 2B–D; see also Figures S1 and S2). The lack of a ROESY cross-peak between the NH of the unnatural SThr4 residue (or SeThr4) and the NH of GalNAc (Figure 2B, left panel), characteristic of the eclipsed conformation of the glycosidic linkage in GalNAc-Thr,<sup>16,17</sup> together with the presence of a cross-peak between the NH of SThr4 (or SeThr4) and H1 of GalNAc (Figure 2B, right panel), suggests a different conformation for the S- and Se-containing glycosidic linkages in 2\* and 3\*, respectively, with regard to GalNAc-Thr (Figure 2B). Clear structural differences between glycopeptides 1\* and the two surrogates, 2\* and 3\*, are also observable in their peptide backbone. In particular, the sequential NH-NH ROESY cross-peak that connects residues 4 and 5 in compounds 2\* and 3\* hints

at a folded conformation of the glycosylated SThr and SeThr residues (Figure 2B, left panel).<sup>28</sup>

The relevant proton–proton distances for the conformational analysis derived from the ROESY spectrum of each compound were then used as time-averaged restraints<sup>29,30</sup> in experiment-guided MD simulations in accordance with our well-established protocol.<sup>31,32</sup> The good agreement between the experimental and theoretically derived distances validates our calculations (Table S1 and Figures S4 and S5). According to the MD simulations, the S- and Se-glycosidic linkages of 2\* and 3\* display a unique conformation centered at  $\phi/\psi \approx 65^\circ/70^\circ$ , which agrees with the exo-anomeric effect<sup>22</sup> and deviates from the more eclipsed arrangement observed for 1\* ( $\phi/\psi \approx 65^\circ/120^\circ$ )<sup>16</sup> (Figure 2C and Figure S6). It is important to note that this conformer lies at one of the local minima calculated for methyl 4-thio- $\alpha$ -maltoside<sup>33</sup> and is similar to that explored by an unnatural Tn antigen with a cysteine residue previously prepared in our laboratory.<sup>34</sup> The side chain of the unnatural residues in 2\* and 3\* is rather rigid in solution, with conformers characterized by  $\chi_1 = 60^\circ$ . The slightly different geometry and flexibility of S- and Se-glycosidic linkages relative to the O-glycosidic linkage, together with the larger size of the S and Se atoms, precludes an effective interaction between the peptide backbone and the carbohydrate. In fact, neither significant hydrogen bonds nor water pockets were observed between these moieties. This finding emphasizes the synergistic roles of the methyl group of the threonine and the glycosidic oxygen atom in defining the conformational preference of the natural Tn-Thr antigen. Regarding the peptide backbone, compounds 2\* and 3\* showed conformations characterized by a folded structure around unnatural residues SThr4 and SeThr4, respectively (Figure 2 and Figures S4 and S5). This arrangement of the peptide differs from that previously reported for 1\* (Figure 2D), which displays a mostly  $\beta$ -sheet-like extended conformation in solution (Figure 2D) owing to water-mediated hydrogen bonds between the peptide and GalNAc.<sup>17,32</sup> The different arrangement of the backbone was also supported by the CD spectra (Figure 2E). Furthermore, according to unrestrained 1  $\mu$ s MD simulations in explicit water, nonglycosylated peptide 2 exhibits a random coil conformation in solution (Figure S3), which is different from the structure adopted by 2\*. Thus, despite the larger distance between the



carbohydrate and the peptide backbone in glycopeptides 2\* and 3\*, our results suggest that the sugar moiety still plays a role as a structural modulator, which presumably may reduce the conformational space accessible to the peptide backbone.

Overall, unnatural glycopeptides 2\* and 3\* display markedly different conformations in solution relative to that of naturally occurring counterpart 1\* that are induced by the replacement of a single atom in these compounds (O → S/Se). In particular, the conformational preference at both the glycosidic linkage and the unnatural residue (SThr4 or SeThr4) is shifted toward those of 1\* bound to an anti-MUC1 monoclonal antibody.<sup>19</sup> Thus, the energy cost associated with a conformational change in the glycopeptide from extended in solution to folded in the bound state is expected to be minimized (*vide infra*).

Conformational Analysis of Unnatural Glycopeptides 2\* and 3\* Bound to scFv-SM3. Crystals suitable for the X-ray diffraction analysis of a recombinantly expressed singlechain variable fragment of the SM3 antibody (scFv-SM3) complexed with 2\* and 3\* were obtained. The X-ray structure of these complexes, solved at high resolution (<2.0 Å, Table S2 and Figure 3 and Figure S7; PDB IDs: 5N7B and 6FRJ) revealed that the conformation of the bound peptide was nearly identical to that adopted by 1\* when bound to scFv-SM3 (Figure 3C). This result demonstrates that the antibody recognizes a well-defined epitope conformation, regardless of the nature of the glycosylated amino acid, characterized by torsion angles at the glycosylated residue typical of folded structures ( $\phi$  and  $\psi$  close to  $-88^\circ$  and  $10^\circ$ , respectively). As detailed above, this conformation is also adopted in solution by the peptide backbone of glycopeptides 2\* and 3\* (Figure 2D and Figures S4 and S5). As for glycopeptide 1\*, the stabilizing contacts in complexes 2\*/scFv-SM3 and 3\*/scFv-SM3 involve several hydrogen bonds, some of which are mediated by water molecules, as well as several stacking interactions (Figure 3A,B).

Of note, two distinct binding modes are observed for glycopeptide 2\* in complex with scFv-SM3 that differ solely in the geometry of the glycosidic linkage. Binding mode A is characterized by a glycosidic linkage with  $\phi/\psi = 87^\circ/74^\circ$ . This conformer corresponds to the structure adopted by 2\* in solution

(Figures 2C and 3A). Alternatively, in binding mode B, with glycosidic linkage angles of  $\varphi/\psi \approx 90^\circ/-90^\circ$ , the glycopeptide structure is stabilized by an intramolecular hydrogen bond between the NH of SThr4 and the endocyclic oxygen (O5) of GalNAc (Figure 3A). This binding mode was also found for the serine and cysteine variants of the immunodominant PDTRP region of MUC1.<sup>19</sup> Although binding mode A allows the N-acetyl group of GalNAc to stack with the aromatic ring of a tryptophan residue (Trp33H) of scFv-SM3, mode B impedes any direct contact between the sugar and the antibody. The electron density observed for the GalNAc moiety of glycopeptide 3\* is rather weak, which may suggest the simultaneous presence of both binding modes observed for derivative 2\* (Figure 3B and Figure S7).

Extensive MD simulations performed on the 2\*/scFv-SM3 complex supported that both binding modes (A and B) are stable in solution (Figure S8 and S9). Interestingly, quantum mechanical (QM) calculations performed on abbreviated models of glycopeptides 1\* and 2\* (compounds 1' and 2', respectively; see Tables S3 and S4 and Figure S10) indicate that the larger repulsion between the  $\beta$ -methyl group of Thr and H1 of GalNAc in glycopeptide 1\*, as a result of the smaller size of the oxygen atom, together with the more distorted geometry of the intramolecular hydrogen bond between O5 (GalNAc) and NH (Thr) leads to the lack of binding mode B in the naturally occurring glycopeptide.

Affinity of Unnatural Glycopeptides 2\* and 3\* for scFv-SM3. A detailed conformational analysis of glycopeptides 2\* and 3\* both in solution and bound to scFv-SM3 in comparison to that assumed by 1\* in solution indicates that the structure of these peptides is preorganized for binding, which is not the case for 1\*. Accordingly, tighter binding would be expected for the unnatural derivatives (vide supra).

To confirm this hypothesis, their binding affinities ( $K_D$ ) for scFv-SM3 were measured by using surface plasmon resonance (SPR) assays (Figure 4 and Figures S11–S15). The highest affinities were observed for unnatural glycopeptides 2\* and 3\* (with  $K_D = 168$  and  $193 \mu\text{M}$  at  $25^\circ\text{C}$ , respectively).

Notably, an improved affinity (~20-fold) was obtained relative to unglycosylated epitope 1. The variation in the affinity of natural glycopeptide 1\* with temperature is higher than for the unnatural counterparts. This result may indicate the existence of an extra entropic penalty associated with the binding of 1\* (Figure 4B) and highlights in this respect the inherently different conformational behavior of unnatural glycopeptides 2\* and 3\*, as already concluded through NMR experiments and MD simulations.

Preparation and in Vivo Studies of a Cancer Vaccine Based on an Engineered Glycopeptide. As discussed above, partially glycosylated peptides with sequences derived from MUC1 are an exciting niche of research for the development of therapeutic cancer vaccines. As yet, none of them has so far succeeded in clinical trials, underlining the difficulty of inducing effective and durable immunological responses to a self-antigen such as tumor-associated MUC1.<sup>12</sup> Additional research is needed to understand how to break the tolerance to self-antigens, which includes a knowledge of how vaccine formulation is processed by the immune system and how (glyco)peptide antigens are presented by major histocompatibility complexes (MHCs) I and II.<sup>35-38</sup>

The results presented in this work prove that a single atom substitution at the glycosidic linkage has a remarkable impact on the structure of the glycopeptide in solution, especially at the glycosylated residue, which in turn may significantly affect the peptide presentation and overall vaccine efficacy. Additionally, natural glycopeptides may suffer degradation from endogenous glycosidases,<sup>39,40</sup> which alters their effectiveness as immunizing antigens,<sup>41</sup> while S-glycoside analogs have improved stability.<sup>42,43</sup> These two considerations prompted us to test whether structurally engineered glycopeptide 12 could be used as tumor-associated antigen mimic through a nanoparticle-based immunogenic formulation (Figure 5A). Glycopeptide 12 comprises the complete tandem repeat sequence of MUC1 and features the SThr\* residue described above. Additionally, this glycopeptide displays a (4S)-4-fluoro-L-proline (fPro) residue that replaces the Pro moiety at the beginning of the PDTRP epitope sequence. The motivation to select this doubly engineered glycopeptide was to combine the entropic benefit induced by SThr\* by preorganizing the epitope structure for optimal binding and the beneficial enthalpic effect produced by

fPro by enhancing antigen–antibody interactions.<sup>13</sup> Moreover, one of us has previously shown that PEGylated AuNPs could be used as efficient antigen carriers to establish humoral immunity against the tumor-associated form of MUC1 in mice, and the elicited antibodies recognized the natural antigen on human breast cancer cells.<sup>44</sup> These promising results led us to conjugate MUC1 antigen mimic glycopeptide 12 to AuNPs in accordance with the strategy previously described (AuNP-12, Figure 5A, Figure S16, and Table S5).<sup>44</sup> On this occasion, the synthesis effort was greatly reduced by omitting the extension of the glycopeptide with a CD4 T-cell peptide epitope, and the immunogenic formulation was administered to the mice without any additional adjuvant.

The success of the conjugation reactions was easily confirmed through gel electrophoresis analysis, in which conjugated AuNP-12 is characterized by a reduced electrophoretic mobility relative to the precursor, linker-functionalized AuNPs (Figure 5B). Additionally, a significant increase in the hydrodynamic diameter of the nanoparticles was observed with dynamic light scattering (DLS, Table S5) upon conjugation. Peptide loading was determined by amino acid analysis to be ~200 glycopeptides/AuNP.

Next, a standard immunization strategy was followed to test the immunogenic potential of AuNP-12 *in vivo*. Thus, a group of five BALB/c mice were immunized with a prime dose followed by three equal booster doses of AuNP-12 (each dose corresponds to 2 µg of the glycopeptide) at 21-day intervals, whereas a control group was treated with phosphate-buffered saline (PBS) as shown in Figure 5B. A week after the last booster dose, the mice were sacrificed, and the serum was harvested. Analyses of the antisera showed that AuNP-12 can elicit a significant anti-MUC1 IgG antibody response. The total antibody end point titers (Figure S18) were better than those observed for the previously reported AuNP-based vaccine candidate in the presence of complete Freund's adjuvant.<sup>44</sup> This result demonstrates that this adjuvant is fully dispensable for the administration of our AuNP-based vaccine candidate, which is therefore self-adjuvating in its own right. Non-negligible IgM titers were also observed (although these were significantly lower than IgG titers), which suggests that glycopeptide 12 on AuNPs can induce class-switch recombination even without the use of a “universal” CD4 Tcell peptide. Next, the antibody isotypes in the

antisera were evaluated. IgG1 was the predominant antibody in all antisera (Figure 5C), which suggests that Th2-type immune responses were predominantly induced by AuNP-12 in these mice.

Finally, IgG2a, IgG2b, and carbohydrate-related IgG3 antibodies<sup>45</sup> were detected in all animals, albeit weakly. To confirm that the elicited antibodies were able to recognize the native tumor-associated MUC1 antigen on human cancer cells selectively, two human cancer cell lines (MCF-7 and T47D) and the human embryonic kidney cell line (HEK293T) were stained with the mice antisera and analyzed by flow cytometry (Figure 6A). Indeed, the antisera reacted strongly with MCF-7 and T47D cells, which express tumor-associated MUC1 on their surface. Conversely, negligible low binding was observed for HEK293T cells, which is consistent with the lack of MUC1 on their surface. These results are in good agreement with those obtained from confocal microscopy (Figure 6B) that show the presence of the MUC1 antigen on the surface of MCF-7 and T47D cells (green color) but not on HEK293T cells. Notably, the antisera also positively stained cancer cells from biopsies of breast cancer patients (right panel in Figure 6C and Figure S19), but no staining is observed in the case of cells from healthy patients. Thus, these results demonstrate the antigen mimic potential of unnatural glycopeptide 12.

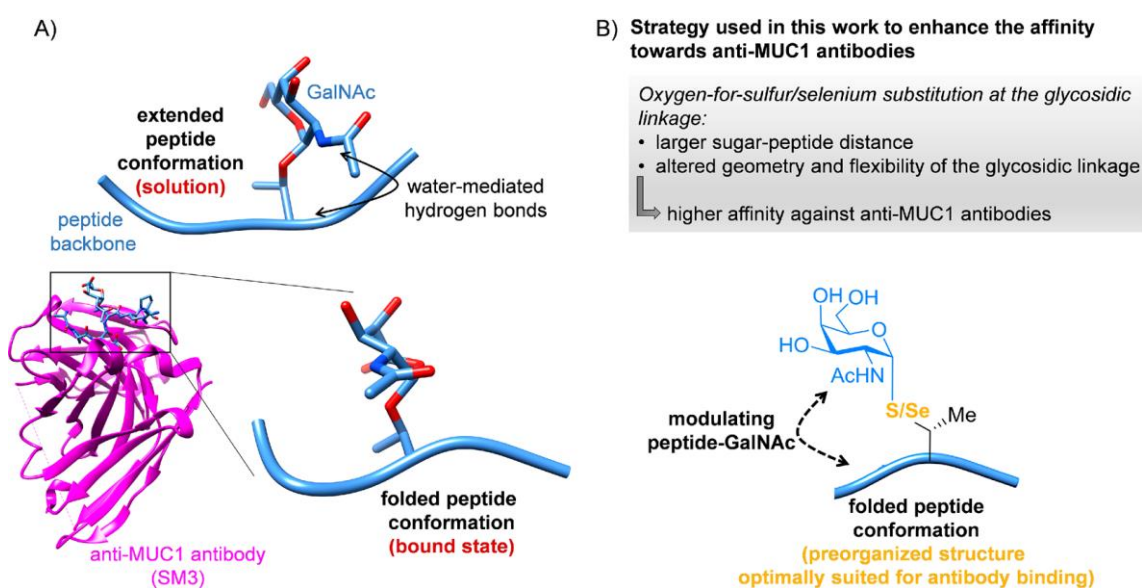


Figure 1. (A) Major conformation of a tumor-associated MUC1 glycopeptide in solution (top) and bound to the SM3 antibody (bottom). (B) Proposed strategy to allow the peptide backbone to adopt a preorganized structure in solution.

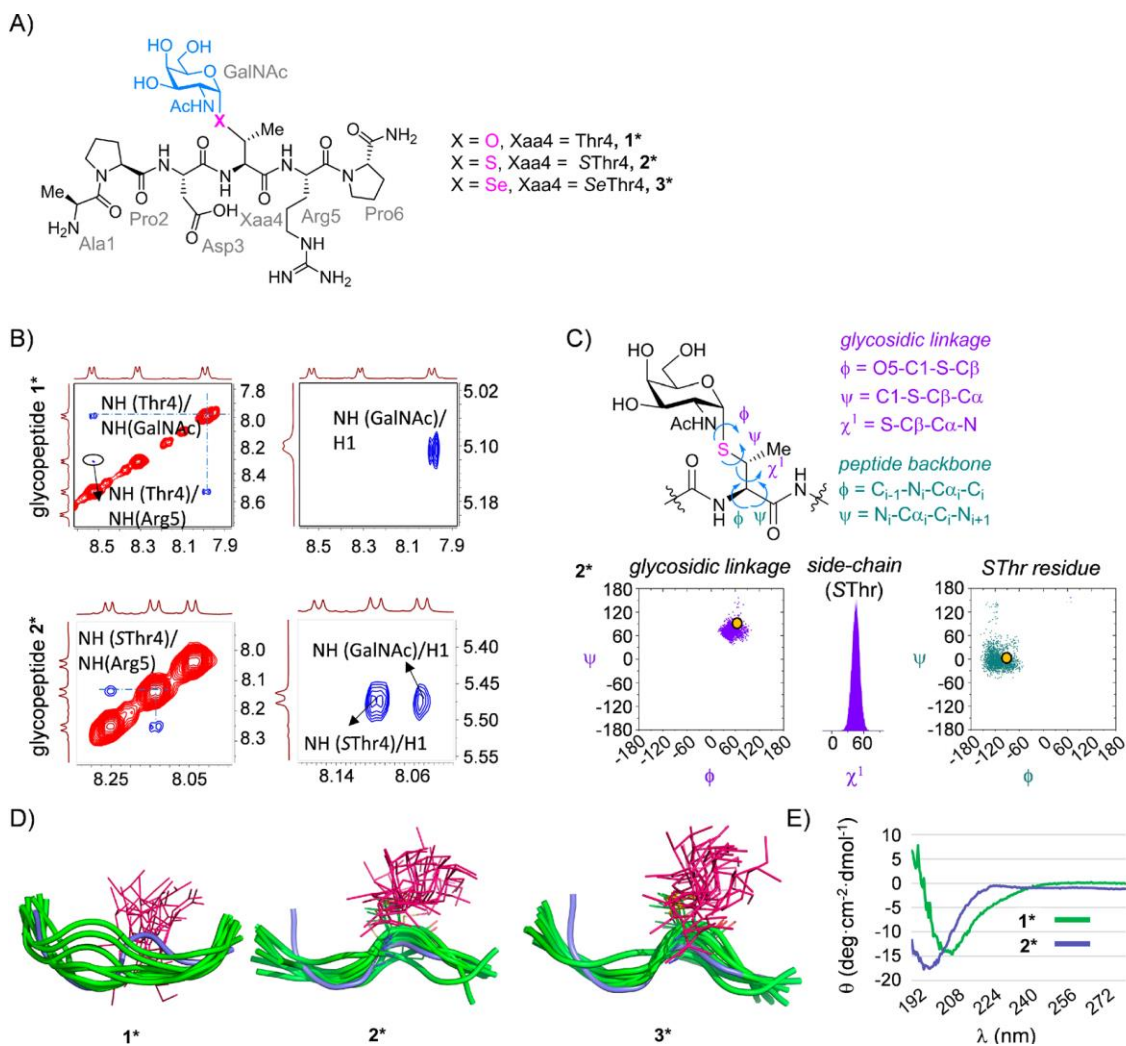


Figure 2. (A) Glycopeptides synthesized and studied in this work, comprising the minimum epitope recognized by most anti-MUC1 antibodies.<sup>7</sup>(B) Sections of the 500 ms 2D-ROESY spectrum (400 MHz) in H<sub>2</sub>O/D<sub>2</sub>O (9:1) at 298 K and pH 6.5 for glycopeptides 1\* (upper panel) and 2\* (lower panel) that show the amide region. Diagonal peaks are in red. ROE contacts are represented as blue cross-peaks. A second set of signals is observed, corresponding to the cis configuration of the amide bond of proline residues.<sup>23</sup> (C) Geometry and flexibility at the glycosidic linkage and peptide backbone for the unnatural SThr residue of glycopeptide 2\* in solution derived from 20 ns experiment-guided MD simulations. The yellow circles correspond to the conformation found in the crystal structure of glycopeptide 1\* bound to a single-chain variable fragment of the SM3 antibody (scFv-SM3; PDB ID: 5A2K). (D) Structural ensembles derived from 20 ns experiment-guided MD simulations for compounds 1\*, 2\*, and 3\* in solution, together with the conformation of the peptide backbone of 1\* (in blue) found by X-ray crystallography to be bound to the scFv-SM3 antibody (PDB ID: 5A2K). (E) Circular dichroism (CD) spectra of compounds 1\* and 2\* (0.25 mM in sodium phosphate buffer, pH 7.5, 20°C).

Scheme 1. Synthesis Route Followed for the Preparation of Building Block 4<sup>44</sup>

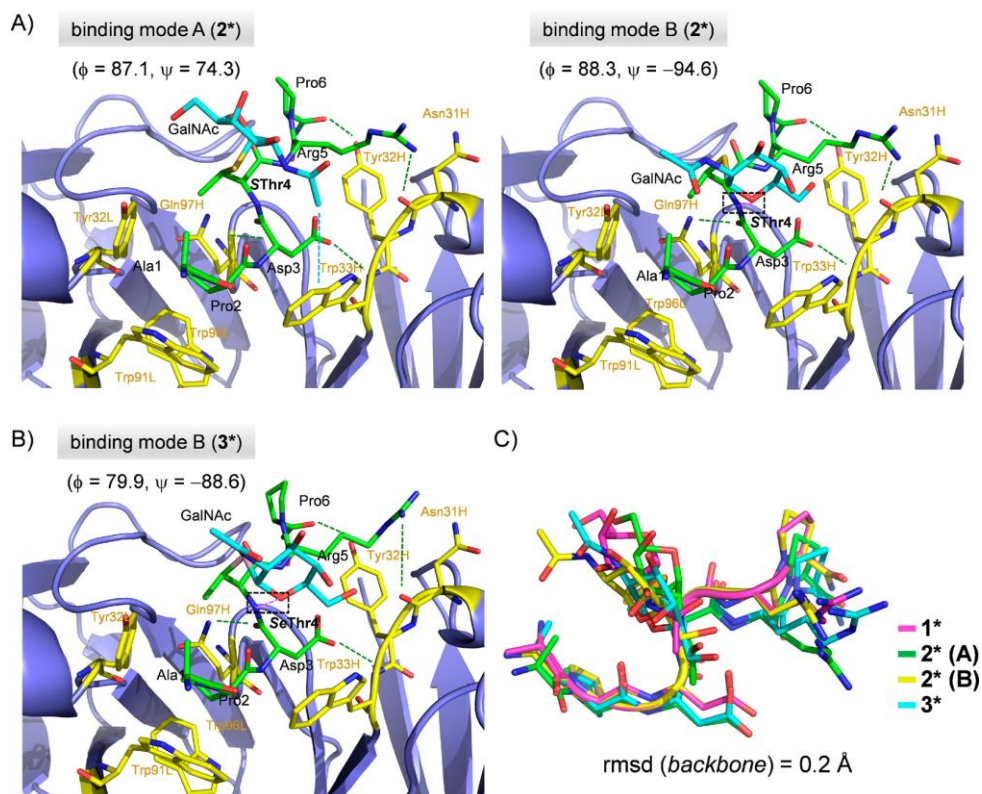
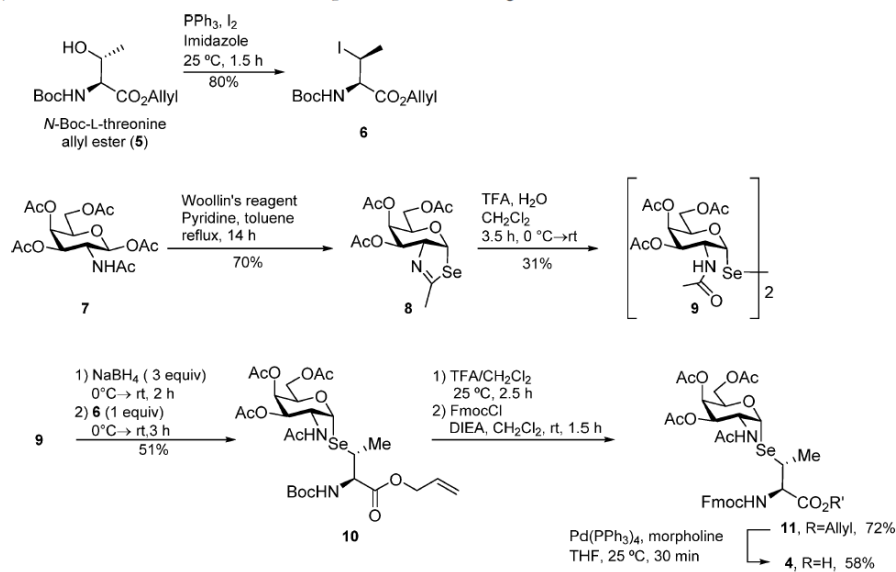


Figure 3. X-ray structures of glycopeptides (A) 2\* and (B) 3\* bound to the scFv-SM3 antibody (PDB IDs: 5N7B and 6FRJ). Glycopeptide carbon atoms are shown in green. Carbon atoms of key residues of scFv-SM3 are colored yellow. Green dashed lines indicate hydrogen bonds between peptide backbones and the scFv-SM3 antibody. Pink dashed lines indicate the hydrogen bond between the NH of SThr (or SeThr) and O5 (dashed boxes). The blue dashed line indicates a CH/ $\pi$  interaction between the N-acetyl group of GalNAc and Trp33H in binding mode A. Note that the density corresponding to the GalNAc moiety in glycopeptide 3\* is only partial (Figure S7), strongly suggesting the existence of local flexibility. (C) Superposition of glycopeptides 1\*, 2\*, and 3\* in complex with the scFv-SM3 antibody, which shows that the antibody recognizes the same conformation for the peptide backbone, regardless of the nature of the glycosylated residue.

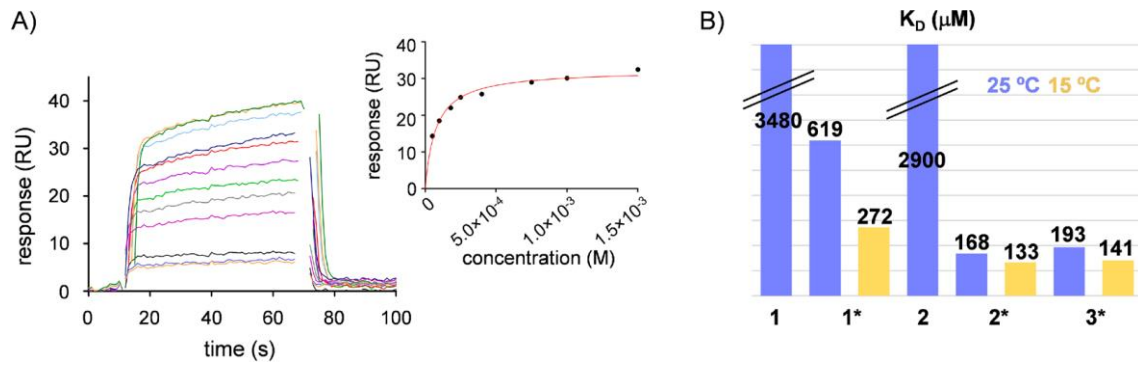


Figure 4. (A) SPR curves and the response–concentration fit obtained for the binding of 2\* to scFv-SM3. (B)  $K_D$  constants derived from SPR experiments for the studied (glyco)peptides.

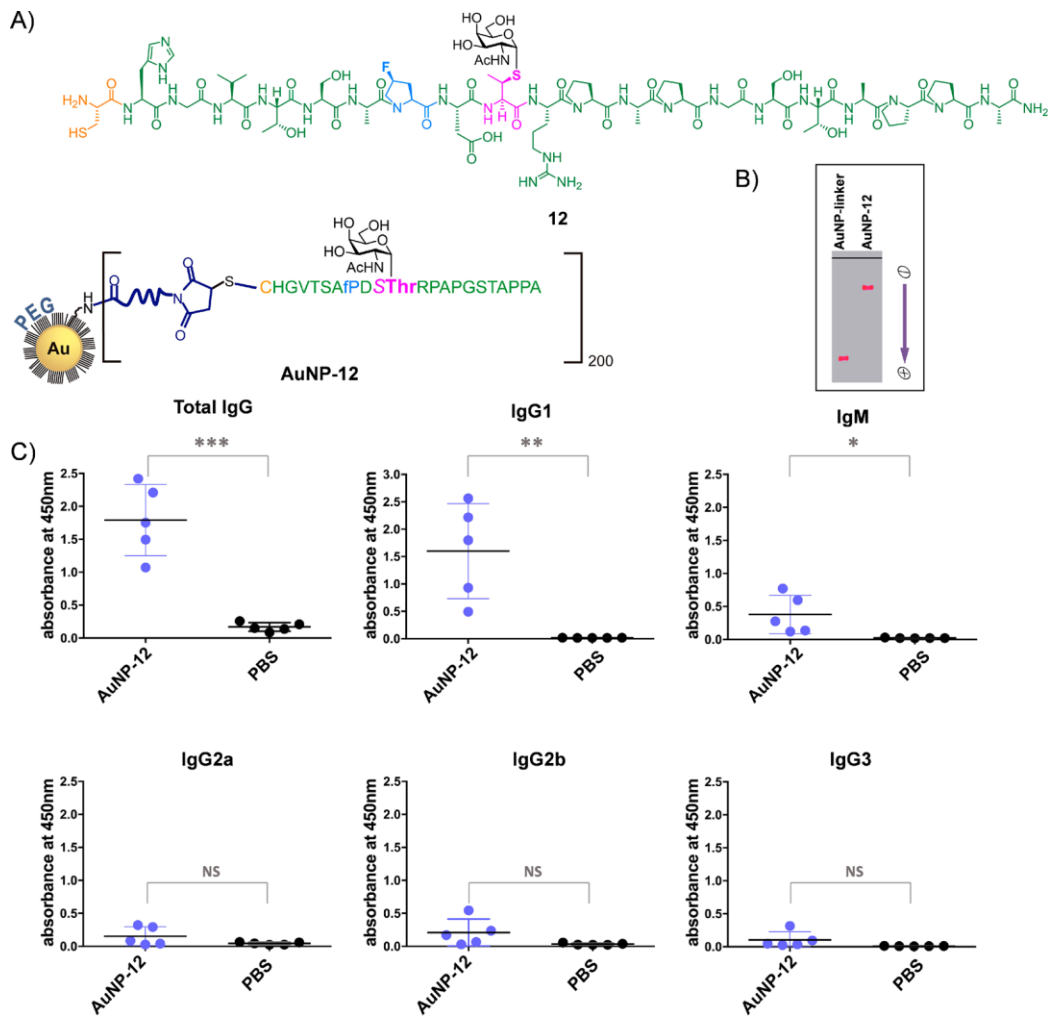


Figure 5. (A) Schematic representation of the vaccine candidate containing engineered MUC1-like glycopeptide 12 attached to the surface of gold nanoparticles (AuNP-12). (B) Agarose gel electrophoresis of the AuNP-linker (loaded with the SM(PEG)<sub>2</sub>-linker; see the Supporting Information) and AuNP-12. (C) Total and subtyping (IgG1, IgG2a/b, IgG3, and IgM) anti-MUC1 antibodies after immunizing mice ( $n = 5$ ) with AuNP-12. The ELISA plates were coated with glycopeptide 12 conjugated to bovine serum albumin. Horizontal lines indicate the mean for the group of five mice. Asterisks indicate statistically significant differences (\*\*\* $P < 0.005$ , \*\* $P < 0.01$ , \* $P < 0.05$ ), and NS indicates no significant difference.



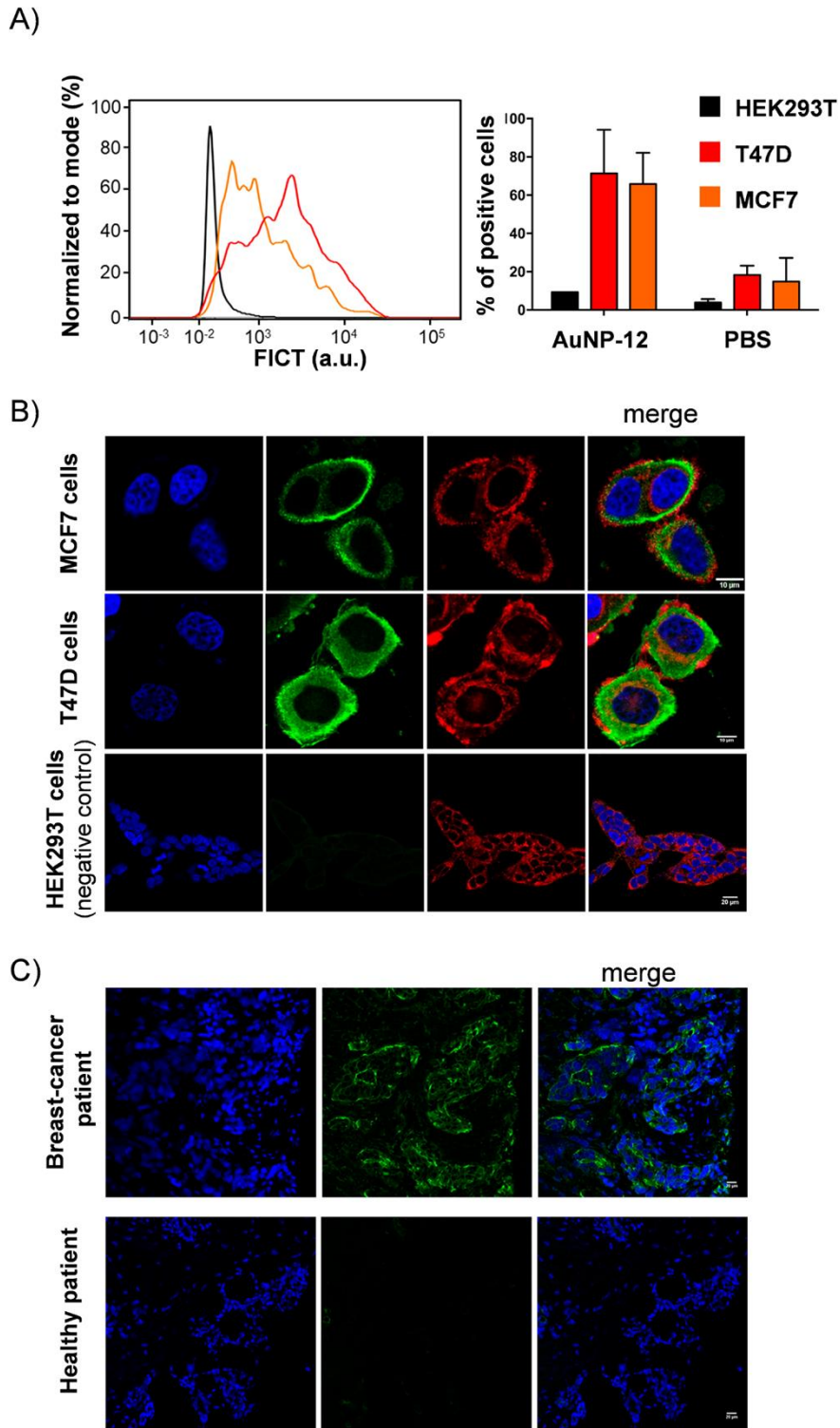


Figure 6. (A) Staining of living cells with the antisera of mice immunized with AuNP-12 analyzed by flow cytometry: HEK293T (black line), MCF7 (orange line), and T47D (red line). Staining with a 1:100 dilution of sera and visualization with a mouse secondary  $\alpha$ -IgG-488 antibody. (B) Confocal microscopy images show that mice antisera after vaccination with AuNP-12 do not stain HEK293T cells as expected because these cells do not express tumor-associated MUC1 on their surface. On the contrary, breast cancer cells MCF7 and T47D expressing tumor-associated MUC1 are positively stained by mice antisera. Blue = Hoechst (nuclei); green = secondary antimouse IgG Alexa 488 (tumor-associated MUC1); and red = CellMask Deep Red (membrane dye). (C) The antisera of mice vaccinated with AuNP-12 positively stain tissue biopsies from breast cancer patients. Blue = Hoechst (nuclei); green = secondary antimouse IgG Alexa 488 (tumor-associated MUC1).

### 3.23. CONCLUSIONS

Our experimental evidence strongly suggests that it is possible to fine tune the conformational preferences of GalNAc-containing glycopeptides in solution by employing a simple oxygen-for-sulfur or oxygen-for-selenium substitution at the glycosidic linkage. These simple chemical modifications have a significant structural impact allowing the peptide backbone to adopt a preorganized structure that is optimally suited for antibody binding, as confirmed by the improved binding affinity to a model anti-MUC1 antibody. Additionally, the potential of a dually modified glycopeptide (fPro and SThr) as a tumor-associated MUC1 antigen mimic has been demonstrated *in vivo*. Significantly, the antisera of mice vaccinated with AuNP-12 recognize cancer cells with high selectivity in biopsies of breast cancer patients. This result confirms that the antibodies generated against the engineered antigen are able to recognize the naturally occurring antigen in its physiological context. Finally, we envision the strategy presented here to be of general interest because it may be applied to modulate the affinity of biologically relevant glycopeptides toward their receptors.

#### ASSOCIATED CONTENT

##### \* Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.8b13503.

Synthesis and characterization of glycopeptides 2\* and 3\* and of the AuNP-based vaccine, conformational analysis in solution of glycopeptides 2\* and 3\*, details of the X-ray structure of 2\* and 3\* bound to scFv-SM3, SPR curves and the response–concentration fit obtained for the binding of the (glyco)peptides, Cartesian coordinates, electronic energies, Gibbs free energies and lowest frequencies of the DFT-calculated structures, additional molecular dynamics simulations figures, immunization protocol, antibody titers and antibody isotypes, antibody reactivity toward human cancer cell lines determined by flow cytometry analysis and analyzed by confocal microscopy, and studies on cancer cells from breast cancer patients (PDF)

## AUTHOR INFORMATION

### Corresponding Authors

\*roberto.fiammengo@iit.it.

\*gb453@cam.ac.uk; gbernardes@medicina.ulisboa.pt.

\*francisco.corzana@unirioja.es.

### ORCID

Jesús Jiménez-Barbero: 0000-0001-5421-8513

Gonzalo Jiménez-Osés: 0000-0003-0105-4337

Omar Boutureira: 0000-0002-0768-8309

Jesús M. Peregrina: 0000-0003-3778-7065

Ramón Hurtado-Guerrero: 0000-0002-3122-9401

Roberto Fiammengo: 0000-0002-6087-6851

Gonçalo J. L. Bernardes: 0000-0001-6594-8917

Francisco Corzana: 0000-0001-5597-8127

### AUTHOR CONTRIBUTIONS

+These authors contributed equally.

### NOTES

The authors declare no competing financial interest.

### ACKNOWLEDGMENTS

We thank the Ministerio de Economía y Competitividad (projects CTQ2015-67727-R, CTQ2013-44367-C2-2-P, CTQ2015-64597-C2-1P, CTQ2017-

89750-R, CTQ2017-90088-R, CTQ2015-70524-R, and BFU2016-75633-P) and the Italian Ministry of Education, University and Research (MIUR) (PRIN 2015 contract no. 2015RNWJAM). I.C. thanks Universidad de La Rioja for an FPI grant. A.G. and G.J.L.B. thank FCT Portugal (Ph.D. studentship, SFRH/BD/115932/2016 and FCT Investigator IF/00624/2015, respectively). G.J.L.B. holds a Royal Society URF (UF110046 and URF\R\180019) and an ERC StG (TagIt, GA no. 676832). F.C. and G.J.L.B. thank the EU (Marie-Sklodowska Curie ITN, Protein Conjugates, GA no. 675007). O.B. and G.J.-O. are Ramón y Cajal Fellows (RYC-2015-17705 and RYC-2013-14706, respectively). R.H-G. thanks Agencia Aragonesa para la Investigación y Desarrollo (ARAID) and the Diputación General de Aragón (DGA, E34\_R17) for financial support. The research that led to these results also attracted funding from the FP7 (2007-2013) under BioStruct-X (grant agreement no. 283570 and BIOSTRUCTX\_5186). We also thank the ALBA synchrotron radiation source, in particular, the beamline XALOC facility. We thank Emanuele Papini and Regina Tavano (University of Padua, Italy) for insightful discussions about the immunogenicity of glycopeptidefunctionalized AuNPs and Vito Maggi and Valentina Rosato (IIT-Lecce) for expert technical assistance. Computational resources of CESGA (Santiago de Compostela) and BERONIA (Universidad de La Rioja) were used in this work. The authors thank Vikki Cantrill for her help with the preparation and editing of this article.

## REFERENCES

- (1) Hollingsworth, M. A.; Swanson, B. J. Mucins in Cancer: Protection and Control of the Cell Surface. *Nat. Rev. Cancer* 2004, 4, 45–60.
- (2) Kufe, D. W. Mucins in Cancer: Function, Prognosis and Therapy. *Nat. Rev. Cancer* 2009, 9, 874–885.
- (3) Senapati, S.; Das, S.; Batra, S. K. Mucin-Interacting Proteins: From Function to Therapeutics. *Trends Biochem. Sci.* 2010, 35, 236–245.
- (4) Taylor-Papadimitriou, J.; Burchell, J. M. Mucins and Cancer; Taylor-Papadimitriou, J., Burchell, J. M., Eds.; Future Medicine: London, 2013.

(5) Ju, T.; Otto, V. I.; Cummings, R. D. The Tn Antigen-Structural Simplicity and Biological Complexity. *Angew. Chem., Int. Ed.* 2011, 50, 1770–1791.

(6) Schuman, J.; Campbell, A. P.; Koganty, R. R.; Longenecker, B. M. Probing the Conformational and Dynamical Effects of O-Journal Glycosylation Within the Immunodominant Region of a MUC1 Peptide Tumor Antigen. *J. Pept. Res.* 2003, 61, 91–108.

(7) Karsten, U.; Serttas, N.; Paulsen, H.; Danielczyk, A.; Goletz, S. Binding Patterns of DTR-Specific Antibodies Reveal a Glycosylation-Conditioned Tumor-Specific Epitope of the Epithelial Mucin (MUC1). *Glycobiology* 2004, 14, 681–692.

(8) 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.

(9) Buskas, T.; Thompson, P.; Boons, G.-J. Immunotherapy for Cancer: Synthetic Carbohydrate-Based Vaccines. *Chem. Commun.* 2009, 105, 5335–5349.

(10) 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.

(11) Gaidzik, N.; Westerlind, U.; Kunz, H. The Development of Synthetic Antitumour Vaccines from Mucin Glycopeptide Antigens. *Chem. Soc. Rev.* 2013, 42, 4421–4442.

(12) Hossain, M. K.; Wall, K. A. Immunological Evaluation of Recent MUC1 Glycopeptide Cancer Vaccines. *Vaccines* 2016, 4, 25–13.

(13) Somovilla, V. J.; Bermejo, I. A.; Albuquerque, I. S.; Martínez-Sáez, N.; Castro-López, J.; Garcia-Martin, F.; Compañón, I.; Hinou, H.; Nishimura, S.-I.; Jiménez-Barbero, J.; Asensio, J. L.; Avenoza, A.; Busto, J. H.; Hurtado-Guerrero, R.; Peregrina, J. M.; Bernardes, G. J. L.; Corzana, F. The Use of Fluoroproline in MUC1 Antigen Enables Efficient Detection of Antibodies in Patients with Prostate Cancer. *J. Am. Chem. Soc.* 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. Principles of Mucin Architecture: Structural Studies on Synthetic Glycopeptides Bearing Clustered Mono-, Di-, Tri-, and Hexasaccharide Glycodomains. *J. Am. Chem. Soc.* 2002, 124, 9833–9844.

(15) Corzana, F.; Busto, J. H.; Jiménez-Oses, G.; Asensio, J. L.; Jiménez-Barbero, J.; Peregrina, J. M.; Avenoza, A. New Insights into  $\alpha$ -GalNAc-Ser Motif: Influence of Hydrogen

Bonding versus Solvent Interactions on the Preferred Conformation. *J. Am. Chem. Soc.* 2006, 128, 14640–14648.

(16) Corzana, F.; Busto, J. H.; Jiménez-Oses, G.; García de Luis, M.; 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.

(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-Oses, G.; Hurtado-Guerrero, R.; Cocinero, E. J.; Corzana, F. Water Sculpt the Distinctive Shapes and Dynamics of the Tumor-Associated Carbohydrate Tn Antigens: Implications for Their Molecular Recognition. *J. Am. Chem. Soc.* 2018, 140, 9952–9960.

(18) Burchell, J.; Gendler, S.; Taylor-Papadimitriou, J.; Girling, A.; Lewis, A.; Millis, R.; Lampert, D. Development and Characterization of Breast Cancer Reactive Monoclonal Antibodies Directed to the Core Protein of the Human Milk Mucin. *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. 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.

(20) Asensio, J. L.; Ardá, A.; Cañada, F. J.; Jiménez-Barbero, J. Carbohydrate-Aromatic Interactions. *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. Modulating Weak Interactions for Molecular Recognition: A Dynamic Combinatorial Analysis for Assessing the Contribution of Electrostatics to the Stability of CH/π Bonds in Water. *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. Conformational Selection of Glycomimetics at Enzyme Catalytic Sites: Experimental Demonstration of the Binding of Distinct High-Energy Distorted Conformations of C-, S-, and O-Glycosides by *E. Coli* β-Galactosidases. *J. Am. Chem. Soc.* 2002, 124, 4804–4810.

(23) Dziadek, S.; Griesinger, C.; Kunz, H.; Reinscheid, U. M. Synthesis and Structural Model of an  $\alpha$  (2,6)-Sialyl-T Glycosylated MUC1 Eicosapeptide under Physiological Conditions. *Chem. - Eur. J.* 2006, 12, 4981–4993.

(24) Gracia-Vitoria, J.; Osante, I.; Cativiela, C. Stereoselective Synthesis of Modified Cysteines. *Tetrahedron: Asymmetry* 2017, 28, 215–245.

(25) Chen, S.; Fahmi, N. E.; Nangreave, R. C.; Mehellou, Y.; Hecht, S. M. Synthesis of pdCpAs and Transfer RNAs Activated with Thiiothreonine and Derivatives. *Bioorg. Med. Chem.* 2012, 20, 2679–2689.

(26) Narayan, R. S.; VanNieuwenhze, M. S. Versatile and Stereoselective Syntheses of Orthogonally Protected  $\beta$ -Methylcysteine and  $\beta$ -Methylanthionine. *Org. Lett.* 2005, 7, 2655–2658.

(27) Jobron, L.; Hummel, G. Solid-Phase Synthesis of New S Glycoamino Acid Building Blocks. *Org. Lett.* 2000, 2, 2265–2267.

(28) Dyson, H. J.; Wright, P. E. Defining Solution Conformations of Small Linear Peptides. *Annu. Rev. Biophys. Biophys. Chem.* 1991, 20, 519–538.

(29) Pearlman, D. A.; Kollman, P. A. Are Time-Averaged Restraints Necessary for Nuclear Magnetic Resonance Refinement? *J. Mol. Biol.* 1991, 220, 457–479.

(30) Pearlman, D. How Is an NMR Structure Best Defined? An Analysis of Molecular Dynamics Distance-Based Approaches. *J. Biomol. NMR* 1994, 4, 1–16.

(31) Fernández-Tejada, A.; Corzana, F.; Busto, J. H.; Jiménez-Osés, G.; Jiménez-Barbero, J.; Avenoza, A.; Peregrina, J. M. Insights into the Geometrical Features Underlying  $\beta$ -O-GlcNAc Glycosylation: Water Pockets Drastically Modulate the Interactions Between the Carbohydrate and the Peptide Backbone. *Chem. - Eur. J.* 2009, 15, 7297–7301.

(32) 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. Detection of Tumor-Associated Glycopeptides by Lectins: the Peptide Context Modulates Carbohydrate Recognition. *ACS Chem. Biol.* 2015, 10, 747–756.

(33) Mazeau, K.; Tvaroska, I. PCILO Quantum-Mechanical Relaxed Conformational Energy Map of Methyl 4-Thio- $\alpha$ -Maltoside in Solution. *Carbohydr. Res.* 1992, 225, 27–41.

(34) Aydiillo, C.; Compañón, I.; Avenzoza, 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.

(35) Lakshminarayanan, V.; Thompson, P.; Wolfert, M. A.; Buskas, T.; Bradley, J. M.; Pathangey, L. B.; Madsen, C. S.; Cohen, P. A.; Gendler, S. J.; Boons, G.-J. Immune Recognition of Tumor-Associated Mucin MUC1 Is Achieved by a Fully Synthetic Aberrantly Glycosylated MUC1 Tripartite Vaccine. *Proc. Natl. Acad. Sci. U. S. A.* 2012, *109*, 261–266.

(36) 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, D. B.; Wolfert, M. A.; Boons, G.-J.; Cohen, P. A.; Gendler, S. J. MUC1 Vaccines, Comprised of Glycosylated or Non-Glycosylated Peptides or Tumor-Derived MUC1, Can Circumvent Immunoediting to Control Tumor Growth in MUC1 Transgenic Mice. *PLoS One* 2016, *11*, No. e0145920.

(37) 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, U.; Finn, M. G.; Huang, X. Protective Epitope Discovery and Design of MUC1-Based Vaccine for Effective Tumor Protections in Immunotolerant Mice. *J. Am. Chem. Soc.* 2018, *140*, 16596–16609.

(38) Wolfert, M. A.; Boons, G.-J. Adaptive Immune Activation: Glycosylation Does Matter. *Nat. Chem. Biol.* 2013, *9*, 776–784.

(39) Richichi, B.; Thomas, B.; Fiore, M.; Bosco, R.; Qureshi, H.; Nativi, C.; Renaudet, O.; BenMohamed, L. A Cancer Therapeutic Vaccine Based on Clustered Tn-Antigen Mimetics Induces Strong Antibody-Mediated Protective Immunity. *Angew. Chem., Int. Ed.* 2014, *53*, 11917–11920.

(40) Nativi, C.; Renaudet, O. Recent Progress in Antitumoral Synthetic Vaccines. *ACS Med. Chem. Lett.* 2014, *5*, 1176–1178.

(41) Rich, J. R.; Bundle, D. R. S-Linked Ganglioside Analogues for Use in Conjugate Vaccines. *Org. Lett.* 2004, *6*, 897–900.

(42) Huo, C.-X.; Zheng, X.-J.; Xiao, A.; Liu, C.-C.; Sun, S.; Lv, Z.; Ye, X.-S. Synthetic and Immunological Studies of N-Acyl Modified S Linked STn Derivatives as Anticancer Vaccine Candidates. *Org. Biomol. Chem.* 2015, *13* (12), 3677–3690.

(43) Bousquet, E.; Spadaro, A.; Pappalardo, M. S.; Bernardini, R.; Romeo, R.; Panza, L.; Ronsisvalle, G. Synthesis and Immunostimulating Activity of a Thioglycolipopeptide



Glycomimetic as a Potential Anticancer Vaccine Derived From Tn Antigen. *J. Carbohydr. Chem.* 2000, 19, 527–541.

(44) Cai, H.; Degliangeli, F.; Palitzsch, B.; Gerlitzki, B.; Kunz, H.; Schmitt, E.; Fiammengo, R.; Westerlind, U. Glycopeptide-Functionalized Gold Nanoparticles for Antibody Induction Against the Tumor Associated Mucin-1 Glycoprotein. *Bioorg. Med. Chem.* 2016, 24, 1132–1135.

(45) Haji-Ghassemi, O.; Blackler, R. J.; Martin Young, N.; Evans, S. V. Antibody Recognition of Carbohydrate Epitopes. *Glycobiology* 2015, 25, 920–952.

## **DISCUSION Y CONCLUSIONES**

Numerosos estudios han observado la sobreexpresión del antígeno Tn en un gran número de cánceres. La formación de este antígeno ha demostrado ser o bien una consecuencia de la reubicación de la localización de las GalNAc-Ts o de la reciente hipermetilación del gen que codifica para la chaperona COSMC, que facilita el plegamiento de la glicosiltransferasa C1GalT1. (Ju, 2018)

Con el fin de plantar cara a estos problemas, se han desarrollado vacunas frente a estos antígenos en células cancerosas derivados de mucinas parcialmente glicosiladas (Wilson, 2013; Gaidzik, 2013). El problema principal de estas vacunas radica en que los péptidos generados químicamente pueden no comportarse de la misma forma en la que los antígenos lo hacen al presentarse al sistema inmunitario en su contexto biológico. Por este motivo, su eficacia se redujo considerablemente. (Beatty, 2015)

El estudio de las lectinas, proteínas con alta afinidad por carbohidratos, ha sido primordial para comprender cómo se realiza la interacción con carbohidratos (Bhutia *et al*, 2019). Este comportamiento ha servido para desarrollar tratamientos frente a patógenos invasivos que utilizan estas moléculas para su adhesión e invasividad. En péptidos glicosilados, las lectinas generalmente reconocen al carbohidrato y en menor medida o nada en absoluto a la secuencia aminoacídica (Hadjialirezaei, 2017). Conocer a nivel molecular como las lectinas reconocen el antígeno Tn permitiría la síntesis de lectinas con una alta especificidad por la molécula completa con fines de diagnóstico. Por el contrario, la mayoría de los anticuerpos hasta la fecha excepto el 5E5 presentan una alta especificidad por la cadena peptídica de los antígenos (Wandall, 2010). Por tanto, la combinación de un buen reconocimiento de la secuencia peptídica y del carbohidrato permitirían el desarrollo de vacunas específicas para estas enfermedades.

En esta tesis se ha pretendido caracterizar a nivel molecular la unión de la lectina a diferentes péptidos glicosilados (Artículo I). Se ha conseguido obtener la primera estructura tridimensional de una lectina en complejo con el epítipo que constituye el antígeno Tn (APDT\*R) y demostrar que la lectina lo reconoce a través de interacciones hidrofóbicas y por puentes de hidrógeno con el péptido y con el azúcar. Además, se han analizado tres fragmentos distintos

glicosilados, derivados de la secuencia de repetición en tándem de MUC1 que contiene 20 aminoácidos. Mediante el estudio de las constantes de afinidad y el análisis estructural, se ha descrito una pérdida de afinidad cuando este antígeno se encuentra en los fragmentos AHGVTSA o APGSTAP. Los residuos de Asp y Arg en la secuencia PDTR parecen fijar la conformación tanto del péptido como del carbohidrato para maximizar las interacciones con las lectinas. Estos estudios servirán como base para modificar nuestra lectina y aumentar la capacidad de reconocimiento del antígeno Tn, y por consiguiente poder usarla como herramienta de diagnóstico para detectar este antígeno.

Paralelamente se ha planteado el estudio de la unión de péptidos glicosilados derivados del antígeno Tn con la región variable (ScFv) del anticuerpo 1SM3 (Artículos II, III, IV, V y VI), el cual se había caracterizado hasta ahora únicamente frente al péptido sin glicosilar (Gaidzik *et al*, 2013). Este anticuerpo ha sido estudiado anteriormente por otros grupos (Meyer *et al*, 2003) y presenta mayor especificidad por el antígeno Tn que por la cadena peptídica sin glicosilar.

El dominio ScFv del anticuerpo 1SM3 ha resultado una buena elección a la hora de realizar las investigaciones, tanto por su fácil expresión como por su afinidad por el antígeno Tn, muy similar a la del anticuerpo completo. Se ha evaluado la secuencia peptídica reconocida por el anticuerpo observando la importancia de la presencia de Thr frente a Ser, así como la presencia del azúcar. Además, los ensayos de afinidad han demostrado que la Thr glicosilada presenta la mejor unión.

Del mismo modo y con el fin de mejorar la afinidad de los anticuerpos anti-Tn para facilitar su reconocimiento y especificidad, hemos optado por estrategias sintéticas en las que se han introducido diferentes variaciones en glicopéptidos derivados del antígeno Tn. Algunas de ellas se han concentrado en la sustitución del oxígeno del grupo hidroxilo de Ser o Thr por grupos S o Se (Artículo VI), en alejar el carbohidrato de la cadena peptídica mediante “linkers” (Artículo III) y en la modificación de la Pro en posición 7 por derivados de prolina modificados con flúor (Artículo IV). Estos derivados con flúor han mostrado una mejor presentación del antígeno al favorecer la interacción CH- $\pi$

con el Trp91L de 1SM3\_ScFv, otorgando una mayor estabilidad. En este caso, aunque no es el único, se ha analizado a nivel celular la eficacia de estos anticuerpos derivados de los antígenos no naturales en muestras de suero de ratones e incluso de pacientes de cáncer de próstata (Artículos IV y VI).

El resultado de la combinación de aunar técnicas de análisis de la afinidad con técnicas de análisis estructural ha permitido conocer cómo el antígeno es reconocido por el anticuerpo, tanto en sus interacciones a nivel molecular como en su disposición 3D en el anticuerpo. Las estructuras cristalinas han mostrado cómo el péptido se posiciona en el estado sólido (unido al anticuerpo), lo que ha llevado a pensar en la importancia de su dinámica en el estado libre y asociado para una interacción más óptima. Se han realizado estudios tanto en fase gas como en fase líquida y sólida para comparar esta disposición, y para ello se ha requerido el uso de derivados del antígeno con flúor. Con estos se ha llegado a la conclusión que el GalNAc con la Thr rota 60° en el espacio para posicionarse de forma que el metilo se repositone y permita que los grupos NH del azúcar y la Thr interaccionen mediante una molécula de agua. Por lo tanto, en la unión, el agua parece jugar un papel fundamental en la disposición y interacción antígeno-anticuerpo. (Artículo V)

Los resultados recogidos en esta tesis han ayudado, así, a comprender y racionalizar cómo la glicosilación puede modificar la estructura de un antígeno y los efectos que esas modificaciones pueden tener sobre la afinidad de unión antígeno-anticuerpo, lo que a su vez puede ser empleado para desarrollar una generación de alta afinidad de anticuerpos terapéuticos. La dirección de este estudio ha apuntado hacia un diseño innovador, basado en la estructura de MUC1, y ha sentado las bases necesarias para el desarrollo de una nueva generación de vacunas contra el cáncer como las que se están desarrollando actualmente.

Considerando los objetivos iniciales planteados al inicio de la tesis, las conclusiones que se recogen de los resultados obtenidos en cada uno de los artículos son las siguientes:

**ARTICULO I:** Detection of Tumor-Associated Glycopeptides by Lectins: The Peptide Context Modulates Carbohydrate Recognition, 2014.

- Se ha logrado cristalizar la lectina SBA con distintos antígenos derivados de MUC1, esto supone la primera estructura tridimensional del antígeno Tn en complejo con una lectina.
- Los residuos de Asp y Arg en la secuencia PDTR parecen fijar la conformación tanto del péptido como del carbohidrato para maximizar las interacciones con las lectinas. Se observa una pérdida de afinidad cuando este antígeno se encuentra en los fragmentos AHGVTSA o APGSTAP.

**ARTICULO II:** Deciphering the Non-Equivalence of Serine and Threonine O-Glycosylation Points: Implications for Molecular Recognition of the Tn Antigen by an anti-MUC1 Antibody, 2015.

- Se ha conseguido expresar en gran cantidad el 1SM3\_ScFv usando *Pichia pastoris* como sistema de expresión.
- Se ha demostrado que el fragmento ScFv se comporta de manera muy similar al anticuerpo completo. La constante de asociación mostrada por ambos en los ensayos de afinidad resulta muy similar.
- Se ha logrado cristalizar el anticuerpo 1SM3\_ScFv con diferentes péptidos. Se ha observado una mejor afinidad de 1SM3 con Thr glicosilada frente a Thr sin glicosilar. Este hecho puede atribuirse a un débil enlace de hidrógeno entre O6 y Tyr32L, y a un contacto hidrófobo entre el grupo metilo de la unidad GalNAc y Trp33H.

**ARTICULO III:** Design of  $\alpha$ -S-Neoglycopeptides Derived from MUC1 with a Flexible and Solvent-Exposed Sugar Moiety, 2016.

- Se ha cristalizado el fragmento 1SM3\_ScFv conjuntamente con derivados del antígeno Tn que poseen “linkers” de distintas longitudes que unen el azúcar (GalNAc) al péptido a través de un átomo de azufre.

- Se han sintetizado varios derivados del antígeno sulfa-Tn con “linkers” variables entre el glicano y el esqueleto del péptido. La presencia de estos “linkers” entre la cadena principal del péptido y el azúcar impide las interacciones intramoleculares, lo que favorece potencialmente la presentación del azúcar para los eventos de reconocimiento molecular.

**ARTICULO IV:** The Use of Fluoroproline in MUC1 Antigen Enables Efficient Detection of Antibodies in Patients with Prostate Cancer, 2017.

- Se ha cristalizado el fragmento 1SM3\_ScFv conjuntamente con derivados del antígeno Tn que contienen fluoroprolinas.
- Se ha diseñado y sintetizado un nuevo conjunto de derivados MUC1 no naturales que comprenden un residuo (4S)-4-fluoro-L-prolina o 4,4-difluoro-L-prolina en el dominio más inmunogénico. La sustitución de hidrógeno por flúor mejora la interacción clave CH- $\pi$ , que es crucial para mejorar la unión del antígeno al anticuerpo.
- Se ha demostrado la acción como biomarcador de los antígenos generados en este artículo, mejorando la detección de anticuerpos anti-MUC1 en muestras de suero de pacientes con cáncer de próstata.

**ARTICULO V:** Water Sculpts the Distinctive Shapes and Dynamics of the Tumor-Associated Carbohydrate Tn Antigens: Implications for Their Molecular Recognition, 2018.

- Se ha cristalizado el fragmento 1SM3\_ScFv conjuntamente con derivados del antígeno Tn de Thr con uno o dos átomos de flúor unidos a la amida del azúcar.
- Se han utilizado técnicas de análisis estructural, como RMN, conjuntamente con DM para analizar la estructura de los antígenos en solución. Ambos glicopéptidos se acomodan en su posición estéricamente mas cómoda, siendo paralela para el antígeno Tn con Ser, y casi perpendicular en el caso del antígeno Tn con Thr.
- Se ha demostrado mediante el uso de técnicas de análisis estructural en fase gas, como IR, conjuntamente con DM que, en los antígenos con Ser y

Thr glicosilados, el azúcar tiene la misma conformación paralela con el péptido.

- Se ha observado en experimentos en fase sólida, como en el derivado del antígeno Tn de Thr, el agua tiene una importancia crucial a la hora de la presentación del antígeno al anticuerpo. Se produce una rotación del azúcar, favorecida por el desplazamiento de un grupo metilo de Thr, a una posición estable. La rotación genera un bolsillo de agua que estabiliza la molécula al formar interacciones con el azúcar y la treonina del antígeno.

**ARTICULO VI:** Structure-Based Design of Potent Tumor-Associated Antigens: Modulation of Peptide Presentation By Single-Atom O/S or O/Se Substitutions at the Glycosidic Linkage, 2019.

- Se ha cristalizado el fragmento 1SM3\_ScFv conjuntamente con derivados del antígeno Tn en los que un átomo de oxígeno del hidroxilo de la Thr se sustituye por un átomo de azufre o de selenio. Adicionalmente estos glicopéptidos están modificados con prolina sustituida por (4S)-4-fluoro-L-prolina.
- Se ha demostrado que una sustitución de la prolina sustituida por (4S)-4-fluoro-L-prolina combinado con la sustitución del oxígeno por azufre u oxígeno por selenio en la Thr permite que el esqueleto del péptido adopte una estructura preorganizada que sea óptima para unión de anticuerpos.
- Se han utilizado nanopartículas de oro para anclar los derivados peptídicos. Los antisueros de ratones vacunados con AuNP-12 reconocen las células cancerosas con alta selectividad.



**REFERENCIAS**  
**BIBLIOGRAFICAS**

- A.K., Hult, Svensson L., Stamps R., Jorgensen R., Rydberg L., and Henry S.M. 2011. "Forssman Expression on Human Red Cells: Biochemical and Genetic Basis of a Novel Histo-Blood Group System Candidate." *Transfusion*. [https://doi.org/http://dx.doi.org/10.1111/j.1537-2995.2011.03301\\_1.x](https://doi.org/http://dx.doi.org/10.1111/j.1537-2995.2011.03301_1.x).
- Arnaud, Julie, Aymeric Audfray, and Anne Imberty. 2013. "Binding Sugars: From Natural Lectins to Synthetic Receptors and Engineered Neolectins." *Chemical Society Reviews*. <https://doi.org/10.1039/c2cs35435g>.
- Bennett, Eric P., Ulla Mandel, Henrik Clausen, Thomas A. Gerken, Timothy A. Fritz, and Lawrence A. Tabak. 2012. "Control of Mucin-Type O-Glycosylation: A Classification of the Polypeptide GalNAc-Transferase Gene Family." *Glycobiology*. <https://doi.org/10.1093/glycob/cwr182>.
- Bhutia, Sujit K., Prashanta K. Panda, Niharika Sinha, Prakash P. Praharaj, Chandra S. Bhol, Debasna P. Panigrahi, Kewal K. Mahapatra, et al. 2019. "Plant Lectins in Cancer Therapeutics: Targeting Apoptosis and Autophagy-Dependent Cell Death." *Pharmacological Research*. <https://doi.org/10.1016/j.phrs.2019.04.001>.
- Brown, Alan. 2009. "Analysis of Cooperativity by Isothermal Titration Calorimetry." *International Journal of Molecular Sciences*. <https://doi.org/10.3390/ijms10083457>.
- Carpino, Louis A. 1987. "The 9-Fluorenylmethyloxycarbonyl Family of Base-Sensitive Amino-Protecting Groups." *Accounts of Chemical Research*. <https://doi.org/10.1021/ar00143a003>.
- Cereghino, Joan Lin, and James M. Cregg. 2000. "Heterologous Protein Expression in the Methylophilic Yeast *Pichia Pastoris*." *FEMS Microbiology Reviews*. [https://doi.org/10.1016/S0168-6445\(99\)00029-7](https://doi.org/10.1016/S0168-6445(99)00029-7).
- Choi, H., and J. V. Aldrich. 1993. "Comparison of Methods for the Fmoc Solid-phase Synthesis and Cleavage of a Peptide Containing Both Tryptophan and Arginine." *International Journal of Peptide and Protein Research*. <https://doi.org/10.1111/j.1399-3011.1993.tb00350.x>.
- Dhar, Poshmaal, and Julie McAuley. 2019. "The Role of the Cell Surface Mucin MUC1 as a Barrier to Infection and Regulator of Inflammation." *Frontiers in Cellular and Infection Microbiology*. <https://doi.org/10.3389/fcimb.2019.00117>.
- Dokurno, Pawel, Paul A. Bates, Heather A. Band, Lorna M.D. Stewart, John M. Lally, Joy M. Burchell, Joyce Taylor-Papadimitriou, David Snary, Michael J.E. Sternberg, and Paul S. Freemont. 1998. "Crystal Structure at 1.95 Å Resolution of the Breast Tumour-Specific Antibody SM3 Complexed with Its Peptide Epitope Reveals Novel Hypervariable Loop Recognition." *Journal of Molecular Biology*. <https://doi.org/10.1006/jmbi.1998.2209>.
- Emsley, J. W., and J. Feeney. 1995. "Milestones in the First Fifty Years of NMR." *Progress in Nuclear Magnetic Resonance Spectroscopy*. [https://doi.org/10.1016/0079-6565\(95\)01023-8](https://doi.org/10.1016/0079-6565(95)01023-8).
- Emsley, J. W., and J. Feeney. 2007. "Forty Years of Progress in Nuclear Magnetic Resonance Spectroscopy." *Progress in Nuclear Magnetic Resonance Spectroscopy*. <https://doi.org/10.1016/j.pnmrs.2007.01.002>.

- Emsley, Jim, and Jim Feeney. 2016. "Fifty Years of 'Progress in NMR Spectroscopy' - An Editorial from the Founding Editors." *Progress in Nuclear Magnetic Resonance Spectroscopy*, 2016. <https://doi.org/10.1016/j.pnmrs.2016.05.001>.
- Freyer, Matthew W., and Edwin A. Lewis. 2008. "Isothermal Titration Calorimetry: Experimental Design, Data Analysis, and Probing Macromolecule/Ligand Binding and Kinetic Interactions." *Methods in Cell Biology*. [https://doi.org/10.1016/S0091-679X\(07\)84004-0](https://doi.org/10.1016/S0091-679X(07)84004-0).
- Fu, Lei Lei, Cheng Cheng Zhou, Shun Yao, Jia Ying Yu, Bo Liu, and Jin Ku Bao. 2011. "Plant Lectins: Targeting Programmed Cell Death Pathways as Antitumor Agents." *International Journal of Biochemistry and Cell Biology*. <https://doi.org/10.1016/j.biocel.2011.07.004>.
- Gilewski, Teresa A., Govind Ragupathi, Maura Dickler, Shemeeakah Powell, Sonal Bhuta, Kathy Panageas, R. Rao Koganty, et al. 2007. "Immunization of High-Risk Breast Cancer Patients with Clustered STn-KLH Conjugate plus the Immunologic Adjuvant QS-21." *Clinical Cancer Research*. <https://doi.org/10.1158/1078-0432.CCR-06-2189>.
- Gill, David J., Keit Min Tham, Joanne Chia, Shyi Chyi Wang, Catharina Steentoft, Henrik Clausen, Emilie A. Bard-Chapeau, and Frederic A. Bard. 2013. "Initiation of GalNAc-Type O-Glycosylation in the Endoplasmic Reticulum Promotes Cancer Cell Invasiveness." *Proceedings of the National Academy of Sciences* 110 (34): E3152–61. <https://doi.org/10.1073/pnas.1305269110>.
- Hadjialirezaei, Soosan, Gianfranco Picco, Richard Beatson, Joy Burchell, Bjørn Torger Stokke, and Marit Sletmoen. 2017. "Interactions between the Breast Cancer-associated MUC1 Mucins and C-Type Lectin Characterized by Optical Tweezers." *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0175323>.
- Hakomori, S. I. 2001. "Tumor-Associated Carbohydrate Antigens Defining Tumor Malignancy: Basis for Development of Anti-Cancer Vaccines." In *Advances in Experimental Medicine and Biology*.
- Handa, Kazuko, Mark R. Stroud, and Sen Itiroh Hakomori. 1997. "Sialosyl-Fucosyl Poly-LacNac without the Sialosyl-Le(x) Epitope as the Physiological Myeloid Cell Ligand in E-Selectin-Dependent Adhesion: Studies under Static and Dynamic Flow Conditions." *Biochemistry*. <https://doi.org/10.1021/bi971181t>.
- Hanes, Melinda S., Kelley W. Moremen, and Richard D. Cummings. 2017. "Biochemical Characterization of Functional Domains of the Chaperone Cosmc." *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0180242>.
- Hashim, Onn Haji, Jaime Jacqueline Jayapalan, and Cheng Siang Lee. 2017. "Lectins: An Effective Tool for Screening of Potential Cancer Biomarkers." *PeerJ*. <https://doi.org/10.7717/peerj.3784>.
- Häuselmann, Irina, and Lubor Borsig. 2014. "Altered Tumor-Cell Glycosylation Promotes Metastasis." *Frontiers in Oncology* 4. <https://doi.org/10.3389/fonc.2014.00028>.

- Hossain, Md, and Katherine Wall. 2016. "Immunological Evaluation of Recent MUC1 Glycopeptide Cancer Vaccines." *Vaccines*.  
<https://doi.org/10.3390/vaccines4030025>.
- Inamdar, Shashikala R., Mohammed Azharuddin Savanur, Sachin M. Eligar, Vishwanath B. Chachadi, Nagaraja N. Nagre, Chen Chen, Monica Barclays, et al. 2012. "The TF-Antigen Binding Lectin from *Sclerotium Rolfsii* Inhibits Growth of Human Colon Cancer Cells by Inducing Apoptosis in Vitro and Suppresses Tumor Growth in Vivo." *Glycobiology*. <https://doi.org/10.1093/glycob/cws090>.
- Invitrogen, Life Technologies. 2010. "User Manual - EasySelect™ Pichia Expression Kit." *Cat. No. K1740-01*. [https://doi.org/Manual part no. 25-0172](https://doi.org/Manual%20part%20no.%2025-0172).
- Ju, Tongzhong, Vivianne I. Otto, and Richard D. Cummings. 2011. "The Tn Antigen-Structural Simplicity and Biological Complexity." *Angewandte Chemie International Edition* 50 (8): 1770–91. <https://doi.org/10.1002/anie.201002313>.
- Karasek, F. W. 1973. "Liquid Chromatography: HPLC." *Research-Development*.
- Kato, Kentaro, Charlotte Jeanneau, Mads Agervig Tarp, Anna Benet-Pagès, Bettina Lorenz-Depiereux, Eric Paul Bennett, Ulla Mandel, Tim M. Strom, and Henrik Clausen. 2006. "Polypeptide GalNAc-Transferase T3 and Familial Tumoral Calcinosis." *Journal of Biological Chemistry* 281 (27): 18370–77.  
<https://doi.org/10.1074/jbc.M602469200>.
- King, May Jean, Anna Chan, Rachel Roe, Bryan F. Warren, Anne Dell, Howard R. Morris, David C.C. Bartolo, Paul Durdey, and Anthony P. Corfield. 1994. "Two Different Glycosyltransferase Defects That Result in GalNAc $\alpha$ -O-Peptide (Tn) Expression." *Glycobiology*. <https://doi.org/10.1093/glycob/4.3.267>.
- Kussrow, Amanda, Carolyn S. Enders, and Darryl J. Bornhop. 2012. "Interferometric Methods for Label-Free Molecular Interaction Studies." *Analytical Chemistry*.  
<https://doi.org/10.1021/ac202812h>.
- las Rivas, Matilde de, Erandi Lira-Navarrete, Thomas A. Gerken, and Ramon Hurtado-Guerrero. 2019. "Polypeptide GalNAc-Ts: From Redundancy to Specificity." *Current Opinion in Structural Biology*. <https://doi.org/10.1016/j.sbi.2018.12.007>.
- Leeuwen, E. B.M. van, S. Cloosen, B. L.M.G. Senden-Gijsbers, M. Agervig Tarp, U. Mandel, H. Clausen, M. J.E. Havenga, et al. 2006. "Expression of Aberrantly Glycosylated Tumor Mucin-1 on Human DC after Transduction with a Fiber-Modified Adenoviral Vector." *Cytherapy*.  
<https://doi.org/10.1080/14653240500513018>.
- Lemjabbar-Alaoui, Hassan, Andrew McKinney, Yi Wei Yang, Vy M. Tran, and Joanna J. Phillips. 2015. "Glycosylation Alterations in Lung and Brain Cancer." In *Advances in Cancer Research*. <https://doi.org/10.1016/bs.acr.2014.11.007>.
- Li, Yaohao, Amy H. Tran, Samuel J. Danishefsky, and Zhongping Tan. 2019. "Chemical Biology of Glycoproteins: From Chemical Synthesis to Biological Impact." In *Methods in Enzymology*. <https://doi.org/10.1016/bs.mie.2019.02.030>.

- Limpias Catalina, Pérez Gerardo, and Ricaurte Orlando Acosta Jinneth Vega Nohora. 2010. "DETECCIÓN DEL ANTÍGENO Tn EN TUMORES EPITELIALES CON LA LECTINA DE Vicia Villosa Isolectina B4." *Revista Facultad de Medicina*.
- Liu, Bo, He jiao Bian, and Jin ku Bao. 2010. "Plant Lectins: Potential Antineoplastic Drugs from Bench to Clinic." *Cancer Letters*.  
<https://doi.org/10.1016/j.canlet.2009.05.013>.
- Lloyd, Kenneth O., Joy Burchell, Valery Kudryashov, Beatrice W T Yin, and Joyce Taylor-Papadimitriou. 1996. "Comparison of O-Linked Carbohydrate Chains in MUC-1 Mucin from Normal Breast Epithelial Cell Lines and Breast Carcinoma Cell Lines: Demonstration of Simpler and Fewer Glycan Chains in Tumor Cells." *Journal of Biological Chemistry*. <https://doi.org/10.1074/jbc.271.52.33325>.
- Loureiro, Liliana R., Diana P. Sousa, Dylan Ferreira, Wengang Chai, Luís Lima, Carina Pereira, Carla B. Lopes, et al. 2018. "Novel Monoclonal Antibody L2A5 Specifically Targeting Sialyl-Tn and Short Glycans Terminated by Alpha-2-6 Sialic Acids." *Scientific Reports*. <https://doi.org/10.1038/s41598-018-30421-w>.
- Madsen, Caroline B., Kirstine Lavrsen, Catharina Steentoft, Malene B. Vester-Christensen, Henrik Clausen, Hans H. Wandall, and Anders Elm Pedersen. 2013. "Glycan Elongation Beyond the Mucin Associated Tn Antigen Protects Tumor Cells from Immune-Mediated Killing." Edited by Sophia N. Karagiannis. *PLoS ONE* 8 (9): e72413. <https://doi.org/10.1371/journal.pone.0072413>.
- Morrot, Alexandre. 2013. "The Role of Sialic Acid-Binding Receptors (Siglecs) in the Immunomodulatory Effects of Trypanosoma Cruzi Sialoglycoproteins on the Protective Immunity of the Host ." *Scientifica*.  
<https://doi.org/10.1155/2013/965856>.
- Nakada, H., M. Inoue, Y. Numata, N. Tanaka, I. Funakoshi, S. Fukui, A. Mellors, and I. Yamashina. 1993. "Epitopic Structure of Tn Glycophorin A for an Anti-Tn Antibody (MLS 128)." *Proceedings of the National Academy of Sciences* 90 (6): 2495-99. <https://doi.org/10.1073/pnas.90.6.2495>.
- Nakamoto, Kazuo. 2008. *Infrared and Raman Spectra of Inorganic and Coordination Compounds: Part A: Theory and Applications in Inorganic Chemistry: Sixth Edition. Infrared and Raman Spectra of Inorganic and Coordination Compounds: Part A: Theory and Applications in Inorganic Chemistry: Sixth Edition*. <https://doi.org/10.1002/9780470405840>.
- Nath, Sritama, and Pinku Mukherjee. 2014. "MUC1: A Multifaceted Oncoprotein with a Key Role in Cancer Progression." *Trends in Molecular Medicine*.  
<https://doi.org/10.1016/j.molmed.2014.02.007>.
- Newman, S, M Glosky, L Ghekiere, and A Alenty. 1977. "Quantitative Requirements for C3 to Induce Forssman Systemic Shock and Cutaneous Hemorrhagic Vasculitis in Guinea Pigs." *Journal of Allergy and Clinical Immunology* 59 (4): 327-33.  
[https://doi.org/10.1016/0091-6749\(77\)90055-0](https://doi.org/10.1016/0091-6749(77)90055-0).
- Olson, Fredrik J., Malin Bäckström, Hasse Karlsson, Joy Burchell, and Gunnar C. Hansson. 2005. "A MUC1 Tandem Repeat Reporter Protein Produced in CHO-K1 Cells Has Sialylated Core 1 O-Glycans and Becomes More Densely Glycosylated If

- Coexpressed with Polypeptide-GalNAc-T4 Transferase." *Glycobiology*.  
<https://doi.org/10.1093/glycob/cwh158>.
- Palomo, Jose M. 2014. "Solid-Phase Peptide Synthesis: An Overview Focused on the Preparation of Biologically Relevant Peptides." *RSC Advances*.  
<https://doi.org/10.1039/c4ra02458c>.
- Poiroux, Guillaume, Marguerite Pitié, Raphaël Culerrier, Elodie Lafont, Bruno Ségui, Els J. M. Van Damme, Willy J. Peumans, et al. 2011. "Targeting of T/Tn Antigens with a Plant Lectin to Kill Human Leukemia Cells by Photochemotherapy." Edited by Jean-Pierre Vartanian. *PLoS ONE* 6 (8): e23315.  
<https://doi.org/10.1371/journal.pone.0023315>.
- Prendergast, Jillian M., Ana Paula Galvao da Silva, David A. Eavarone, Darius Ghaderi, Mai Zhang, Dane Brady, Joan Wicks, Julie DeSander, Jeff Behrens, and Bo R. Rueda. 2017. "Novel Anti-Sialyl-Tn Monoclonal Antibodies and Antibody-Drug Conjugates Demonstrate Tumor Specificity and Anti-Tumor Activity." *MAbs*.  
<https://doi.org/10.1080/19420862.2017.1290752>.
- Pretsch, E, T Clerc, J Seibl, and W Simon. 1980. *Tablas Para La Elucidación Estructural de Compuestos Orgánicos Por Métodos Espectroscópicos*. Vertex.
- Radhakrishnan, Prakash, Sally Dabelsteen, Frey Brus Madsen, Chiara Francavilla, Katharina L. Kopp, Catharina Steentoft, Sergey Y. Vakhrushev, et al. 2014. "Immature Truncated O-Glycophenotype of Cancer Directly Induces Oncogenic Features." *Proceedings of the National Academy of Sciences* 111 (39): E4066–75.  
<https://doi.org/10.1073/pnas.1406619111>.
- Rafaelsen, Silje, Stefan Johansson, Helge Ræder, and Robert Bjerknes. 2014. "Long-Term Clinical Outcome and Phenotypic Variability in Hyperphosphatemic Familial Tumoral Calcinosis and Hyperphosphatemic Hyperostosis Syndrome Caused by a Novel GALNT3 Mutation; Case Report and Review of the Literature." *BMC Genetics*. <https://doi.org/10.1186/s12863-014-0098-3>.
- Robertson, E. G., M. R. Hockridge, P. D. Jelfs, and J. P. Simons. 2000. "IR-UV Ion-Dip Spectroscopy of n-Benzylformamide Clusters: Step Wise Hydration of a Model Peptide." *Journal of Physical Chemistry A*.
- Ruvolo, Peter P. 2016. "Galectin 3 as a Guardian of the Tumor Microenvironment." *Biochimica et Biophysica Acta - Molecular Cell Research*.  
<https://doi.org/10.1016/j.bbamcr.2015.08.008>.
- Schjoldager, Katrine T.B.G., and Henrik Clausen. 2012. "Site-Specific Protein O-Glycosylation Modulates Proprotein Processing - Deciphering Specific Functions of the Large Polypeptide GalNAc-Transferase Gene Family." *Biochimica et Biophysica Acta - General Subjects*.  
<https://doi.org/10.1016/j.bbagen.2012.09.014>.
- Schuldt, Linda, and Manfred S. Weiss. 2009. "Protein Crystallography: A Concise Guide, by Eaton E. Lattman and Patrick J. Loll." *Crystallography Reviews*.  
<https://doi.org/10.1080/08893110902723949>.
- Sears, P., and C. H. Wong. 1998. "Enzyme Action in Glycoprotein Synthesis." *Cellular and Molecular Life Sciences*. <https://doi.org/10.1007/s000180050146>.

- Sholl, David S., and Janice A. Steckel. 2009. *Density Functional Theory: A Practical Introduction*. *Density Functional Theory: A Practical Introduction*.  
<https://doi.org/10.1002/9780470447710>.
- Springer, G. 1984. "T and Tn, General Carcinoma Autoantigens." *Science* 224 (4654): 1198–1206. <https://doi.org/10.1126/science.6729450>.
- Steenfot, Catharina, Sergey Y. Vakhrushev, Hiren J. Joshi, Yun Kong, Malene B. Vester-Christensen, Katrine T B G Schjoldager, Kirstine Lavrsen, et al. 2013. "Precision Mapping of the Human O-GalNAc Glycoproteome through SimpleCell Technology." *EMBO Journal*. <https://doi.org/10.1038/emboj.2013.79>.
- Steenfot, Catharina, Eric P. Bennett, Katrine T.B.G. Schjoldager, Sergey Y. Vakhrushev, Hans H. Wandall, and Henrik Clausen. 2014. "Precision Genome Editing: A Small Revolution for Glycobiology." *Glycobiology*.  
<https://doi.org/10.1093/glycob/cwu046>.
- Takeuchi, Hideyuki, Kentaro Kato, Kaori Denda-Nagai, Franz Georg Hanisch, Henrik Clausen, and Tatsuro Irimura. 2002. "The Epitope Recognized by the Unique Anti-MUC1 Monoclonal Antibody MY.1E12 Involves Sialyl $\alpha$ 2-3galactosyl $\beta$ 1-3N-Acetylgalactosaminide Linked to a Distinct Threonine Residue in the MUC1 Tandem Repeat." *Journal of Immunological Methods*.  
[https://doi.org/10.1016/S0022-1759\(02\)00298-3](https://doi.org/10.1016/S0022-1759(02)00298-3).
- Tarp, Mads Agervig, and Henrik Clausen. 2008. "Mucin-Type O-Glycosylation and Its Potential Use in Drug and Vaccine Development." *Biochimica et Biophysica Acta - General Subjects*. <https://doi.org/10.1016/j.bbagen.2007.09.010>.
- Taylor-Papadimitriou, J., J. Burchell, D. W. Miles, and M. Dalziel. 1999. "MUC1 and Cancer." *Biochimica et Biophysica Acta - Molecular Basis of Disease*.  
[https://doi.org/10.1016/S0925-4439\(99\)00055-1](https://doi.org/10.1016/S0925-4439(99)00055-1).
- Varrot, Annabelle, Soorej M. Basheer, and Anne Imberty. 2013. "Fungal Lectins: Structure, Function and Potential Applications." *Current Opinion in Structural Biology*. <https://doi.org/10.1016/j.sbi.2013.07.007>.
- Voordijk, Sonja, Tomas Hansson, Donald Hilvert, and Wilfred F. Van Gunsteren. 2000. "Molecular Dynamics Simulations Highlight Mobile Regions in Proteins: A Novel Suggestion for Converting a Murine V(H) Domain into a More Tractable Species." *Journal of Molecular Biology*. <https://doi.org/10.1006/jmbi.2000.3890>.
- Wei, Meng Man, Yong Shi Wang, and Xin Shan Ye. 2018. "Carbohydrate-Based Vaccines for Oncotherapy." *Medicinal Research Reviews*.  
<https://doi.org/10.1002/med.21493>.
- Xu, Yanfei, Alessandro Sette, John Sidney, Sandra J. Gendler, and Alessandra Franco. 2005. "Tumor-Associated Carbohydrate Antigens: A Possible Avenue for Cancer Prevention." *Immunology and Cell Biology*. <https://doi.org/10.1111/j.1440-1711.2005.01347.x>.
- Yamamoto, Miyako, Emili Cid, and Fumiichiro Yamamoto. 2012. "Molecular Genetic Basis of the Human Forssman Glycolipid Antigen Negativity." *Scientific Reports*.  
<https://doi.org/10.1038/srep00975>.

- Yu, Haibo, Manfred Amann, Tomas Hansson, Jutta Köhler, Günter Wich, and Wilfred F. Van Gunsteren. 2004. "Effect of Methylation on the Stability and Solvation Free Energy of Amylose and Cellulose Fragments: A Molecular Dynamics Study." *Carbohydrate Research*. <https://doi.org/10.1016/j.carres.2004.05.003>.
- Yu, Xiaofeng, Zhenzhen Du, Xuhong Sun, Chuanqin Shi, Huaixiang Zhang, and Tao Hu. 2015. "Aberrant Cosmc Genes Result in Tn Antigen Expression in Human Colorectal Carcinoma Cell Line HT-29." *International Journal of Clinical and Experimental Pathology*.



# **ABREVIATURAS**

ADN: Ácido desoxirribonucleico

APS: Persulfato amónico

ARN: Ácido ribonucleico

BMGY: Medio tamponado rico en glicerol (*Buffered Glycerol complex Medium*)

BMMY: Medio tamponado rico en metanol (*Buffered Methanol complex Medium*)

CHES: Acido (2-(N-Ciclohexilamina)-etanosulfónico)

CT: Dominio de la cola citoplasmática de mucinas

DTT: Dithiothreitol

EDTA: Acido etilendiaminotetraacético

EGF: Factor de crecimiento epidérmico

Fc: Fragmento constante del anticuerpo

FDA: Agencia para la Administración de Alimentos y Medicamentos

FPLC: Cromatografía líquida de la proteína rápida

GlcNAc: N-Acetilglucosamina

GT: Glicosiltransferasa

HCl: Ácido clorhídrico

HPLC: Cromatografía líquida de alta resolución

IPTG: Isopropil- $\beta$ -D-1-tiogalactopiranosido

LB/LS: Medio de cultivo Luria-Bertani bajo en sal

mAb: Anticuerpo monoclonal

MUC1: Glicoproteína transmembrana mucina 1

NaCl: Cloruro de sodio

PAGE: Electroforesis en gel de poliacrilamida

PBS: Tampón de sal de fosfatos

PDB: Banco de datos de proteínas

PCR: Reacción en cadena de la polimerasa

PDB: Banco de datos de proteínas

PEG: Polietilenglicol

PTS: Dominio de mucina con alta concentración de Prolina, Treonina y Serina

SBA: Soybean agglutinin

ScFv: Fragmento variable de cadena simple

SDS: Dodecil Sulfato Sódico

SOB: Medio súper óptimo (*Super optimal broth*)

SOC: Medio súper óptimo con represión de catabolitos (*Super optimal broth with catabolite repression*)

STD: Diferencia de transferencia de saturación

STn: Antígeno de Thomsen Friedenreich sialilado

T/TF: Antígeno de Thomsen-Friedenreich

TACA: Antígenos de carbohidratos asociados a tumores

TBE: Tampón TRIS/Boro/EDTA

TE: Tampón TRIS/EDTA

TEMED: N, N, N', N'- tetrametiletilenodiamina

Tn: Antígeno de Thomsen-Friedenreich

TR: Regiones de repetición en tándem

TRIS: Tris(hidroximetil)aminometano

UV: Ultravioleta

VNTR: Variable de polimorfismos de repetición en tándem

YPD: Extracto de levadura-peptona-dextrosa

YPDS: Extracto de levadura-peptona-dextrosa-sorbitol

# **INFORMACIÓN GENERAL**

## Códigos de Aminoácidos

Aminoácido	Código de tres letras	Código o una letra	Aminoácido	Código de tres letras	Código o una letra
Alanina	Ala	A	Leucina	Leu	L
Arginina	Arg	R	Lisina	Lys	K
Asparagina	Asn	N	Metionina	Met	M
Aspártico	Asp	D	Fenilalanina	Phe	F
Cisteína	Cys	C	Prolina	Pro	P
Glutámico	Glu	E	Serina	Ser	S
Glutamina	Gln	Q	Treonina	Thr	T
Glicina	Gly	G	Triptófano	Trp	W
Histidina	His	H	Tirosina	Tyr	Y
Isoleucina	Ile	I	Valina	Val	V
Selenocisteína	Sec	U	Pirrolisina	Pyl	O

**Figura 15.1** Código de aminoácidos. Tabla que recoge los nombres completos de los aminoácidos, seguido de los códigos de 3 letras y posteriormente de 1 letra. **Fuente:** <https://es.slideshare.net/pblanco61/aminoacidos-1213>

## Segunda Letra

		U		C		A		G		
Primera Letra	U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	U
		UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys	C
		UUA	Leu	UCA	Ser	UAA	STOP	UGA	STOP	A
		UUG	Leu	UCG	Ser	UAG	STOP	UGG	Try	G
	C	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg	U
		CUC	Leu	CCC	Pro	CAC	His	CGC	Arg	C
		CUA	Leu	CCA	Pro	CAA	Gln	CGA	Arg	A
		CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg	G
	A	AUU	Iso	ACU	Thr	AAU	Asn	AGU	Ser	U
		AUC	Iso	ACC	Thr	AAC	Asn	AGC	Ser	C
		AUA	Iso	ACA	Thr	AAA	Lys	AGA	Arg	A
		AUG	Met	ACG	Thr	AAG	Lys	AGG	Arg	G
	G	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly	U
		GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly	C
		GUA	Val	GCA	Ala	GAA	Glu	GGA	Gly	A
		GUG	Val	GCG	Ala	GAG	Glu	GGG	Gly	G
										Tercera Letra

**Figura 15.2** Código genético. Tabla de correspondencias en la transcripción de los distintos codones a aminoácidos, en humanos. Existen algunas variaciones entre organismos. **Fuente:** <http://www.innovabiologia.com/biodiversidad/diversidad-animal/el-codigo-genetico/>

<https://aminoacidos.top>

## TABLA CON LOS 20 AMINOÁCIDOS

LOS AMINOÁCIDOS SON EL COMPONENTE IMPRESCINDIBLE DE LAS PROTEÍNAS EN ORGANISMOS VIVOS, HAY MÁS DE 500 AMINOÁCIDOS ENCONTRADOS EN LA NATURALEZA, SIN EMBARGO, EN EL SER HUMANO SOLAMENTE HAY 20 DE ELLOS. LOS AMINOÁCIDOS "ESENCIALES" DEBEN OBTENERSE DE LA DIETA, MIENTRAS LOS AMINOÁCIDOS NO ESENCIALES PUEDE SER SINTETIZADO POR EL CUERPO.

● ALIFÁTICOS ● AROMÁTICOS ● ÁCIDOS ● BÁSICOS ● HIDRÓXILOS ● SULFURADOS ● AMINOS ○ NO ESENCIALES ○ ESENCIALES

Estructura Química Código	ALANINA (A) Ala GCT, GCC, GCA, GCG	GLICINA (G) Gly GGT, GGC, GGA, GGG	ISOLEUCINA (I) Ile ATT, ATC, ATA	LEUCINA (L) Leu CTT, CTC, CTA, CTG, TTA, TTG	PROLINA (P) Pro CCT, CCC, CCA, CCG	VALINA (V) Val GTT, GTC, GTA, GTG
FENILAMINA (F) Phe TTC, TTC	TRIPTOFANO (W) Trp TGG	TIROSINA (Y) Tyr TAT, TAC	ACIDO ASPARTICO (D) Asp GAT, GAC	ACIDO GLUTAMINICO (E) Glu GAA, GAG	ARGININA (R) Arg CGT, CGC, CGA, CGG, AGA, AGG	HISTIDINA (H) His CAT, CAC
LISINA (K) Lys AAA, AAG	SERINA (S) Ser TCT, TCC, TCA, TCG, AGT, AGC	TREONINA (T) Thr ACT, ACC, ACA, ACG	CISTEINA (C) Cys TGT, TGC	METIONINA (M) Met ATG, AAC	ASPARAGINA (N) Asn AAT, AAC	GLUTAMINA (Q) Gln GAA, GAG

<https://aminoacidos.top>

**Figura 15.3** Tabla de aminoácidos. Recoge la estructura de cada uno de los 20 aminoácidos **Fuente:** <https://aminoacidos.top/>