

Andrés Manuel González Ramírez

Caracterización biofísica y
estructural de C1GalT1 en la
síntesis del core 1 de los O-
glicanos de tipo mucina

Director/es

Hurtado Guerrero, Ramón

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

© Universidad de Zaragoza
Servicio de Publicaciones

ISSN 2254-7606



Tesis Doctoral

CARACTERIZACIÓN BIOFÍSICA Y ESTRUCTURAL
DE C1GALT1 EN LA SÍNTESIS DEL CORE 1 DE
LOS O-GLICANOS DE TIPO MUCINA

Autor

Andrés Manuel González Ramírez

Director/es

Hurtado Guerrero, Ramón

UNIVERSIDAD DE ZARAGOZA
Escuela de Doctorado

Programa de Doctorado en Bioquímica y Biología Molecular

2023

Caracterización biofísica y estructural de C1GalT1 en la síntesis del *core* 1 de los *O*- glicanos de tipo mucina.

Memoria presentada por **Andrés Manuel González
Ramírez**, graduado en Biotecnología, para optar al
grado de Doctor en Ciencias.

Zaragoza, febrero de 2023



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 **Caracterización biofísica y estructural de C1GalT1 en la síntesis del *core 1* de los *O*-glicanos de tipo mucina** ha sido realizada, bajo su dirección por el graduado ANDRÉS MANUEL GONZÁLEZ RAMÍREZ, 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, febrero de 2023.

La realización de esta tesis doctoral ha sido posible gracias a la financiación del Ministerio de Economía, Industria y Competitividad del Gobierno de España (BFU2016-75633-P y PID2019-105451GB-I00), de la Agencia Aragonesa para la Investigación y Desarrollo (ARAID), de los fondos FEDER (2014-2020) y de la Diputación General de Aragón (DGA, proyectos E34_R17, LMP58_18). Se ha colaborado también con el sincrotrón Diamond Light Source (DLS) de Oxford (línea I24). Durante los cuatro años de realización de esta tesis (desde julio de 2018 hasta noviembre de 2022, ambos incluidos), el doctorando ha sido beneficiario de una subvención del Ministerio de Economía y Competitividad del Gobierno de España destinada a la contratación de personal investigador predoctoral para la formación de doctores (BES-2017-081846) (BOE del 16/09/2017) y cofinanciada por el Fondo Social Europeo.

Fácilmente podría decir que los años de tesis han sido los más complicados de mi vida, pero gracias a la compañía, ayuda y existencia de la gente en la que me he podido apoyar durante este tiempo ha hecho que se convierta en un viaje del que disfrutar y, principalmente, aprender antes de llegar al destino. No solo las personas, sino los libros y las historias junto a sus personajes han sido los que me han ayudado a encontrar la fuerza o el respiro necesario para recomponerme en los momentos de mayor debilidad. Es por ello que de lo que más orgulloso estoy a día de hoy es de mi capacidad de gestionarme mentalmente a nivel profesional y personal, que aun estando lejos de ser perfecta se encuentra a años luz del Andrés que empezó el doctorado por allá en 2018.

Son muchas las cosas que agradecer porque el tiempo y crecimiento experimentado ha sido enorme, así que, aunque espero no enrollarme no prometo nada. Por algún lado tengo que empezar, así que me gustaría en primer lugar felicitar y agradecer a mi director de tesis Ramón Hurtado por aguantarme y por todo lo que hemos hecho juntos estos cuatro años y medio. A pesar de las recurrentes discrepancias que hemos tenido creo que al final conseguimos llegar a un acuerdo, respetándonos mutuamente. Me atrevería incluso a decir que algo de tu ferocidad por la ciencia me has conseguido transmitir y para que negarlo, algo de cariño te he cogido.

Los cimientos de esta andadura los han construido mis compañeros de grupo Javi, Víctor, Ana Mati y Jorge. Los tres últimos me enseñaron a manejar y a navegar por los laboratorios, con sus sugerencias, opiniones y directrices. Gracias por darme el impulso con el que coger inercia y seguir el resto del camino. Camino que inesperadamente crucé

con un viejo conocido de la carrera, de repente apareció Javi para convertirse en mi compañero durante este tiempo. Dicen que el roce hace el cariño, y joder, al final me he acabado alegrando muchísimo de que aparecieses por el BIFI. Has conseguido amenizar las largas horas de purificaciones o de poner placas de cristales charlando del estado actual de los videojuegos o de cómo te conseguías los platinos haciendo trucos. Por cierto, Javi, de nada por hacerte leer a Sanderson. Casi en la recta final, vino Víctor a poner algo de medida al grupo, un poco de calma, aunque de repente empezásemos a hablar más de DnD y de cómo lloraba Javi porque su personaje no estaba roto. Es por esa calma y fortaleza que transmites que te admiro mucho, creo que es algo de lo que deberías sentirte orgulloso. Y, sobre todo, gracias a los cinco por tener tanta paciencia conmigo y aguantar y comprender mis infinitas quejas o errores. Más tarde llegaron otros Ramones: Jesús, David y Billy, que, aunque no he pasado excesivo tiempo con ellos, espero haber hecho todo lo posible para darles el impulso que me dieron a mí al comienzo.

Me gustaría también hacer especial mención al resto de investigadores principales del BIFI por sus ideas, opiniones y consejos; además de hacer posible el uso de ciertas técnicas o equipos que han ayudado a poder mejorar mi trabajo. Quiero pedirle perdón a Sonia por haber pesado SDS en la balanza de precisión en mis comienzos y agradecerle todas las segundas oportunidades y palmeritas que me has dado durante estos años.

Cuando llegué al BIFI me dijeron que lo mejor del centro era claramente el ambiente y el compañerismo que había en él, pero hasta que no lo sientes no eres consciente de que eso significa que son ellos los que sostienen el centro al ayudarse mutuamente y entender por lo que estás

pasando, ya sea de una manera u otra. Siempre se intenta liberar algo de esa frustración, ya sea saliendo a tomar café o simplemente a tomar el aire. Es en realidad en momentos como esos cuando comienzas a entablar amistad con gente que hasta entonces eran compañeros. Así que gracias de corazón a todos los que habéis estado en esos pasillos cuando la purificación no salía, o cuando no estaba el péptido en el cristal, siempre acudíais a la llamada del café. Sois muchos los que os habéis ganado un agradecimiento más que merecido por alentarme y un hueco en mi corazón gracias al trato diario. Gracias a la sangre fresca por reavivar la llama del BIFI un poco más, Sandra, Hefe Carlos, los JGN y los BH habéis traído un soplo de aire fresco que creo necesario y que espero que se mantenga antes de que se caiga el edificio. Los que ya volaron, los BIFIaurios, gracias por vuestros sabios consejos en general y especialmente a María Conde por los infinitos ánimos en las interminables tardes en el laboratorio y por supuesto a Juan y a Matilde por ser referentes a los que seguir a nivel personal y profesional. David, me estoy acordando mientras escribo esto de cuando tuve que ir a la cocina a por papel para secarnos las lágrimas viendo Coco, de cuando vino Nunilo a cultivos a preguntarnos si estábamos escuchando cosas de chinos o de las veces que de repente nos han dado las 5 de la mañana en algún lugar de Zaragoza y no he podido evitar sonreír, gracias por acompañarme a todo, escucharme y enseñarme. Ana J, he aprendido que el truco de no pedir el café muy cargado es para no acabar en la López. Has conseguido convertirme en un verdadero abanderado del Tony Wilson, espero que estés orgullosa, porque yo lo estoy al ver todo lo que has conseguido, eres increíble. Gracias por estar siempre. Y gracias a los dos por haber confiado en mí para apoyaros cuando lo habéis necesitado. Sabéis de sobra que yo lo he hecho una infinidad de veces.

BeaTris, espero que sigas guardando el bolígrafo que te di, porque cuando el día está un poco gris recuerdo que “We are golden” y siempre vuelve algo de color. Polanco, eres de las personas más creativas que conozco y un torreón imposible de derrumbar, que nunca deje de sobarte calle, que no falten los stickers de café y que bailemos más veces al ritmo de Tripaloski. Todo esto parece casi una despedida, y como no lo es voy a parar y a recordaros a todos los que sigáis por el BIFI que nos vemos pronto, que no os libráis de mí.

Gracias a mi familia de Madrid, a la Chupipandi por ser un remanso de paz al que acudir cuando me veía sobrepasado por las circunstancias. Rafa, Foster, Pepelu, Manu, Rocío, Peli, Laura, Irene, Ara, Charlotte gracias por ser aquellos con los que he podido revitalizarme cuando lo necesitaba, de verdad que lo mejor de haber estudiado Biotecnología es haberos conocido a todos vosotros. No puedo irme de Madrid sin mencionar a la Tanga ni a mis compañeros de máster, gracias a Dani, Jose, Sonji, Lucía y Miguel y gracias especialmente por no preguntarme demasiado cuándo es la defensa. A los que me conocen desde hace más de 15 años, a las malditas estiércolas, decirles gracias por parecer que no ha pasado ni un día sin vernos y quererme y tratarme como siempre y hacerme reír como si todo lo demás diese igual. Los Amos del Infierno no han hecho otra cosa más que amenizarme las horas de espera en el laboratorio con las discusiones mañaneras entre Kiel y Slabon, pero quiero mencionar a mi amigo, ahora canario, Scales por ayudarme a evadirme con Sanderson, SoloLeveling o con cualquier otra cosa por tal de alegrarme un ratito.

Hablando de evasión, mi mayor agradecimiento a Brandon Sanderson por sus novelas y todo lo que me han ayudado a crecer en esta etapa.

Muchas gracias a Borkiront y Raener por existir, por dejarme ocupar vuestro lugar los jueves por la tarde. Que eso haya sido posible se debe al trabajo de Ameba y Alonso, que no solo son unos increíbles directores de juego sino unas aún más increíbles personas. Gracias Ameba, por hacerme sentir verdadera curiosidad por conocer el *lore* detrás de la trama, por hacerme buscar conexiones entre canciones infantiles y dioses, por enseñarme que las acciones tienen consecuencias, por repartir adoquín allá donde diriges, pero no te agradezco que me quitases el brazo. Alonso, gracias a ti también por contarme tu historia, por hacer de un personaje al principio en blanco algo que realmente quiera descubrir, por currártelo tanto, por hacer que le haya terminado cogiendo cariño a *Ánima*. Pero, por otra parte, gracias también por ser tú, por saber siempre qué decir, por hacerme sentir orgulloso, por el cariño y la paciencia. El puto rol además me ha dado la oportunidad de conocer a gente increíble con la que pasar el rato como Berto, Elva, Grevol o mis pequeñines David y Noa. Mi pana, verte me hace pensar que las cosas no están tan mal y que la humanidad tiene esperanza si viene gente como tu detrás de mí. Noa, gracias por preguntar si estaba bien a pesar de parecerlo por fuera, por las charlas cotidianas y livianas y en especial por darme el honor dejarme ayudarte cuando lo necesitabas. *Keep Calm and WAKANA*.

Muchas gracias a mi familia riojana y en especial a mis suegros Feliciano y Pili y a mi cuñada Carmen por hacerme sentir como en casa en La Rioja y por mimarme cada vez que voy a Navarrete. Gracias Marta, me caes muy bien. Pero bueno, quería felicitarte por aguantarme tantas mañanas sin café. Y agradecerte con todo mi ser que me hayas acompañado incondicionalmente durante todo este tiempo, por ser mi equipo, por secar mis lágrimas, por ser mi pilar fundamental que ha aguantado todo

por lo que he pasado. Por más conciertos de Bastille juntos, te quiero. Si algo lleva esta tesis es tu ayuda impresa en ella con todo el amor que me has dado.

Y, sobre todo, gracias a mi familia: Raúl y Marta, os quiero muchísimo, gracias por dejarme ayudaros cuando he tenido oportunidad. Espero que lo que tenemos los tres nunca se acabe, que nos apoyemos, aconsejemos y alentemos a seguir persiguiendo lo que queramos como hemos hecho hasta ahora. Gracias papá por apoyarme en todas mis decisiones y por ser un referente del constante esfuerzo, te admiro mucho. Y gracias mamá. Eres la persona más fuerte e inamovible que conozco, gracias por todo, por decirme que un poquito más, por decirme que ya faltaba poco, por darme cariño y amor todos los días a más de 1000 kilómetros. Esto es el fruto de vuestro esfuerzo, así que esta tesis es tan mía como vuestra.

Estamos revelando un mundo nuevo, Jasnah, y el camino abierto ante nosotros estará oscuro hasta que le llevemos luz. Se nos perdonará si tropezamos de vez en cuando en un terreno que no vemos.

El Rimo de la Guerra. Brandon Sanderson.

CONTENIDOS

1. VISIÓN GENERAL	6
1.1. MODIFICACIONES POSTRADUCCIONALES: O-GLICOSILACIÓN.	8
1.2. O-GLICOSILACIÓN DE TIPO GALNAC.....	9
2. INTRODUCCIÓN	12
2.1. LAS MUCINAS	12
2.1.1. Estructura y funciones principales de las mucinas.....	12
2.2. SÍNTESIS E INICIO DE LA O-GLICOSILACIÓN: GALNAC- TRANSFERASAS.....	17
2.3. ELONGACIÓN DE LOS O-GLICANOS.....	24
2.3.1. Síntesis de los núcleos de los O-glicanos.....	24
2.3.2. Extensión de los <i>cores</i> de O-glicanos.....	26
2.4. C1GALT1 Y COSMC.....	28
2.4.1. Historia de C1GalT1.....	29
2.4.2. Historia de <i>Cosmc</i>	31
2.4.3. Estructura y actividad conocidas de C1GalT1.....	33
2.5. GLICOSILACIÓN DISFUNCIONAL Y PATOLOGÍAS: DEFICIENCIAS EN C1GALT1 Y COSMC.	35
2.5.1. El síndrome Tn.....	40
2.5.2. Nefropatía de IgA.	41
2.5.3. Cáncer y mucinas.....	44
3. OBJETIVOS	50
4. METODOLOGÍA.....	52
4.1. CULTIVOS CELULARES.	52
4.1.1. Cultivos celulares de bacterias: <i>Escherichia coli</i> DH5 α y DH10Bac.....	52
4.1.2. Cultivos celulares de mamíferos: HEK293F.	53

4.1.3.	Cultivos celulares de insectos: Sf9 y HighFive TM	54
4.2.	TÉCNICAS DE BIOLOGÍA MOLECULAR.....	55
4.2.1.	Construcción de los plásmidos de C1GalT1.....	55
4.2.1.1.	Plásmidos para expresión en células de mamíferos.	55
4.2.1.2.	Plásmidos para expresión en células de insectos.	56
4.2.1.3.	Mutagénesis de sitio dirigido de los plásmidos para expresión en mamíferos.....	58
4.2.2.	Sistema dirigido de CRISPR/Cas9 para hacer <i>knockouts</i> en HEK293F. 58	
4.2.3.	Creación de <i>knockins</i> de variantes de C1GalT1 en HEK293F ^{Tn}	60
4.2.4.	Transformación de células competentes <i>E. coli</i> DH5a y DH10Bac y selección de colonias recombinantes.	61
4.2.5.	Conservación en glicerol de las colonias transformadas.	62
4.2.6.	El sistema Bac-to-Bac.....	63
4.2.7.	Amplificación y extracción de ADN plasmídico.....	65
4.2.8.	Amplificación y extracción del cromosoma bacteriano recombinante (bácmido).....	66
4.3.	EXPRESIÓN DE PROTEÍNAS.	67
4.3.1.	Expresión transitoria de <i>DmC1GalT1</i> y sus mutantes en HEK293F....	67
4.3.2.	Producción en Sf9 de baculovirus recombinantes que expresen <i>DmC1GalT1</i>	68
4.3.3.	Expresión de <i>DmC1GalT1</i> en Hi5 TM	69
4.3.4.	Electroforesis SDS-PAGE desnaturizante y tinción.....	69
4.3.5.	Electrotransferencia (<i>Western Blot</i>).	70
4.4.	PURIFICACIÓN DE PROTEÍNAS.....	72
4.4.1.	Cromatografía de afinidad con HISTRAP (IMAC).	73
4.4.2.	Intercambio de tampón.	74
4.4.3.	Corte con Endo-H y purificación.	75
4.4.4.	Cromatografía de exclusión molecular.	77
4.4.5.	Concentración, cuantificación y conservación.	78

4.5.	TÉCNICAS DE BIOLOGÍA CELULAR.	79
4.5.1.	Análisis de citometría de flujo.	79
4.5.2.	Inmunocitología.	80
4.6.	GLICOPÉPTIDOS.	81
4.6.1.	Diseño de glicopéptidos.	81
4.6.2.	Síntesis de glicopéptidos.	81
4.6.3.	Preparación de las disoluciones de péptidos y glicopéptidos.	81
4.7.	TÉCNICAS DE ANÁLISIS ESTRUCTURAL	81
4.7.1.	Cristalización de proteínas.	81
4.7.2.	Cocrystalización de <i>DmC1GalT1</i> .	83
4.7.3.	Toma de datos: difracción de rayos X.	84
4.7.4.	Resolución de la estructura.	85
4.7.5.	Resonancia magnética nuclear (RMN).	86
4.7.6.	Resonancia magnética de diferencia de transferencia de saturación (STD-RMN).	87
4.8.	TÉCNICAS BIOFÍSICAS PARA EL ANÁLISIS Y CARACTERIZACIÓN DE PROTEÍNAS.	90
4.8.1.	Absorción UV-visible y cuantificación de proteínas.	90
4.8.2.	Determinación de la constante de unión por ITC.	91
4.8.3.	Análisis de actividad enzimática.	93
4.9.	TÉCNICAS COMPUTACIONALES.	95
4.9.1.	Estudio de homología: Alineamiento de secuencias de <i>C1GalT1</i> .	95
4.9.2.	Predicción de la estructura secundaria de <i>DmC1GalT1</i> .	95
4.9.3.	Dinámicas moleculares (MD).	96
5.	RESULTADOS	98
5.1.	ALINEAMIENTO MÚLTIPLE DE <i>C1GALT1</i> . ELECCIÓN DE ORTÓLOGOS PARA LA EXPRESIÓN, PURIFICACIÓN Y CARACTERIZACIÓN BIOFÍSICA.	98

5.2.	ANÁLISIS COMPUTACIONAL DE LA ESTRUCTURA SECUNDARIA DE <i>DMC1GALT1</i> Y DISEÑO DE CONSTRUCCIONES PARA SU CRISTALIZACIÓN: PRUEBAS DE EXPRESIÓN CON BACULOVIRUS RECOMBINANTES.	100
5.2.1.	Pruebas de expresión de construcciones truncadas de <i>DmC1GalT1</i> . .	102
5.3.	DISEÑO RACIONAL DE GLICOPÉPTIDOS ESPECÍFICOS PARA <i>C1GALT1</i> . 103	
5.4.	CINÉTICAS DE <i>DMC1GALT1</i> FRENTE A DIFERENTES SUSTRATOS. ...	105
5.5.	EXPERIMENTOS DE STD-RMN.....	110
5.6.	<i>DMC1GALT1</i> NO PRESENTA UN MECANISMO CON CARACTERÍSTICAS ALOSTÉRICAS PARA LA UNIÓN A LOS GLICOPÉPTIDOS.	118
5.7.	DETALLES DEL COMPLEJO FORMADO POR <i>DMC1GALT1</i> -UDP- APDT*RP.	123
5.8.	EL SITIO ACTIVO DE <i>DMC1GALT1</i>	132
5.9.	EL RECONOCIMIENTO MOLECULAR DE A-GALNAc-THR ES POSIBLE GRACIAS A LA CONFORMACIÓN DE ALTA ENERGÍA QUE ADOPTA SU ENLACE GLICOSÍDICO.....	133
5.10.	LA TRANSFERENCIA DE LA GALACTOSA SE REALIZA MEDIANTE UN MECANISMO DE INVERSIÓN A TRAVÉS DE <i>DMC1GALT1</i>	137
5.11.	COMPORTAMIENTO Y ACTIVIDAD DE LOS MUTANTES DE <i>DMC1GALT1</i> <i>IN VITRO</i> Y DE LOS MUTANTES DE <i>HSC1GALT1</i> <i>IN VIVO</i>	139
5.12.	SIMULACIONES DE DINÁMICAS MOLECULARES (MD) PARA LA OBTENCIÓN DE COMPLEJOS ESTRUCTURALES PUTATIVOS TRIDIMENSIONALES DE <i>DMC1GALT1</i>	143
6.	DISCUSIÓN Y CONCLUSIONES	157
6.1.	DISCUSIÓN.	157
6.2.	CONCLUSIONES.....	165
7.	BIBLIOGRAFÍA	167

8. APÉNDICES	190
8.1. LISTA DE TABLAS Y FIGURAS.....	190
8.2. LISTA DE ABREVIATURAS.	196
8.3. RECETARIO.....	200
8.4. SECUENCIA Y PROPIEDADES DE LAS PROTEÍNAS RECOMBINANTES. .	202
8.5. INFORMACIÓN GENERAL.	209
8.5.1. El código genético.	209
8.5.2. Los aminoácidos.....	210

1. VISIÓN GENERAL

La diversidad de formas proteicas (el proteoma) es sorprendentemente mayor a lo que se podría predecir por el número de secuencias codificantes de proteínas existentes en el genoma. Esto se debe principalmente a los mecanismos de edición y modificación a nivel de ADN, ARN y proteínas (Walsh et al., 2005) así como a la existencia de diferentes promotores alternativos a nivel genómico (Landry et al., 2003). Así mismo, el proteoma alcanza un nuevo nivel de complejidad a causa de las modificaciones a las proteínas durante o tras su ensamblaje en los ribosomas. Estas transformaciones químicas, denominadas modificaciones postraduccionales (MPTs) (Tabla 1.1), incrementan la variabilidad del proteoma y alteran la localización, estructura, función, actividad e interacción de las proteínas modificadas. Tanto la cadena peptídica, como los grupos funcionales de las cadenas laterales sirven como aceptores químicos para reacciones enzimáticas, no enzimáticas e incluso autocatalizadas que dan lugar a más de 200 MPTs conocidas actualmente (Conibear, 2020).

Tabla 1.1. MPTs más importantes y sus principales funciones. Se muestra en la figura la relación entre la importancia de las MPTs y el buen funcionamiento las mismas para mantener unas condiciones fisiológicas correctas.

MPTs	Principales funciones
Fosforilación	Modulación enzimática e interacción, señalización.
Acetilación	Estabilidad cromatina, interacciones proteína-proteína, control del ciclo celular, metabolismo celular, transporte nuclear.
Ubiquitinación	Señalización, degradación de proteínas intracelulares.
Metilación	Modificación de histonas, interacción proteína-ADN, regulación génica.
Glicosilación <ul style="list-style-type: none">• N-glicosilación• O-glicosilación	Adhesión celular, interacciones célula-matriz, tráfico molecular, activación de receptores, solubilización proteica, plegamiento y transducción proteico, degradación de proteínas, transporte intracelular y secreción.
Acilación (ácidos grasos)	Localización celular, modulación de interacción, señalizadores celulares.
Sulfatación	Modulación de interacciones proteína-proteína y receptor-ligando.
Formación de puentes disulfuro	Estabilidad proteica.
Hidroxilación	Estabilidad e interacción proteína-ligando.

De manera general, las MPTs podrían dividirse en tres grupos distintos: el primero y el segundo grupo engloban tanto la adición química de grupos funcionales como de complejos, respectivamente. Las modificaciones que añaden cadenas polipeptídicas conformarían el tercer grupo. Dejando a un lado esa robusta clasificación, las 10 MPTs más estudiadas son las siguientes: fosforilación, acetilación, ubiquitinación, metilación, glicosilación, SUMOilación, lipidación (palmitoilación, miristoilación y prenilación) y sulfatación (Ramazi & Zahiri, 2021). En adición a las MPTs mencionadas, merece la pena destacar la importancia de la formación de puentes disulfuro entre cisteínas para mejorar la estabilidad estructural proteica (Swaisgood, 2005) (Tabla 1.1).

Entre los cientos de MPTs identificadas, muchas de ellas han sido reconocidas como las responsables del mantenimiento de la homeostasis, siendo la desregulación de esas modificaciones las causas o causantes de muchas patologías (Tabla 1.1) (Ohtsubo & Marth, 2006; Schjoldager & Clausen, 2012). La regulación de todas estas modificaciones está dirigida por un gran número de enzimas entre las que destacan quinasas, proteasas y diferentes enzimas de las que aún no se conoce su mecanismo de acción. Específicamente, a la adición de glicanos o carbohidratos en aminoácidos específicos de la cadena peptídica de las proteínas o en lípidos se le denomina glicosilación. Entre las MPTs, es la más regulada, compleja y abundante en el ser humano, encontrándose al menos una de cada cinco proteínas del proteoma glicosilada (Apweiler et al., 1999; Houry et al., 2011).

La biosíntesis de disacáridos, oligosacáridos y polisacáridos involucra la acción de cientos de diferentes glicosiltransferasas (GTs) (EC 2.4.x.y), enzimas que catalizan la transferencia de grupos azucarados desde moléculas donadoras hasta moléculas receptoras específicas, formando enlaces glicosídicos. Estas proteínas pueden clasificarse según su mecanismo de acción como enzimas de retención o de inversión en función de la estereoquímica de los sustratos y productos de reacción

(Sinnott, 1990). De todas formas, la clasificación de glicosiltransferasas es un tema en constante actualización debido a las numerosas características por las que se pueden agrupar. La base de datos de enzimas que procesan carbohidratos o CAZy, por sus siglas en inglés, se encarga de intentar ordenar o clasificar las glicosiltransferasas en diferentes grupos, siendo un buen referente para saber a qué grupo pertenece la enzima objetivo del estudio (“CAZy Database”; <http://www.cazy.org>) (Drula et al., 2022).

Por otro lado, cuando se menciona la glicosilación, los dos tipos de glicosilación más frecuentes son la *N*-glicosilación sobre residuos de asparagina (Paulson & Colley, 1989) y la *O*-glicosilación en treonina y serina, que puede ocurrir de manera menos frecuente en tirosinas e hidroxilisinas (Stanley, 2011; Steentoft et al., 2013).

1.1. Modificaciones postraduccionales: O-glicosilación.

Todos los tipos de *O*-glicosilación ocurren en proteínas de la vía secretora, exceptuando un tipo de glicosilación que añade GlcNAc en el citosol (Holt et al., 1987). El resto de tipos, que ocurren en proteínas secretadas, son la adición de *O*-*N*-acetilgalactosamina, *O*-xilosa, *O*-manosa, *O*-galactosa, *O*-fucosa y *O*-glucosa (Jentoft, 1990)(Figura 1.1.). La completa síntesis de estas modificaciones puede dividirse en dos pasos principales: la iniciación, en el que las glicosiltransferasas se encargan de reconocer el motivo en el aminoácido aceptor y enlazar el primer monosacárido, y la elongación o procesamiento, en el que numerosas enzimas añaden de manera secuencial los azúcares que son necesarios para la síntesis de los glicanos complejos (Joshi, Narimatsu, Schjoldager et al., 2018).

La complejidad que subyace tras la *O*-glicosilación se debe a la necesidad de sincronizar de manera perfecta y secuencial la regulación de abundantes enzimas (Figura 1.1), que en conjunto generan diferentes tipos de conjugados glicosilados de diferente estructura y función (Joshi et al., 2018; Schjoldager & Clausen, 2012).

etapa de iniciación está controlada por más de 20 isoformas diferentes de las UDP-*N*- α -acetil- α -D-galactosamina *N*-acetilgalactosaminiltransferasas polipeptídicas (o ppGalNAc-Ts) que se encargan de añadir un grupo GalNAc a un residuo de treonina o serina. La redundancia parcial de tal número de isoformas se explica porque, aunque realicen la misma reacción de transferencia, cada una de las isoformas tienen selectividades contextuales y patrones de actividad y expresión diferentes (Bennett et al., 2012; Hurtado-Guerrero, 2016; Schjoldager & Clausen, 2012; Ten Hagen et al., 2003).

La segunda etapa o elongación del α -GalNAc, a diferencia de la iniciación, está típicamente controlada en todas las células por una galactosiltransferasa de inversión de seguimiento (*“follow-up”*), localizada en el aparato de Golgi, denominada C1GalT1 o *core 1 – T – Synthase*. Esta enzima se encarga de añadir un grupo de galactosa al GalNAc unido al residuo de treonina y serina. Tras esta elongación, se comienzan a formar los altamente glicosilados dominios de tipo mucina, que tienen funciones de elevada importancia como la señalización celular y la defensa en diferentes epitelios (Kufe, 2009; Varki et al., 2022).

El mal funcionamiento de C1GalT1 provoca una glicosilación aberrante que puede ocasionar desregularización en la mayoría de procesos fisiológicos, entre ellos la homeostasis tisular y del sistema inmune y la protección e integridad de barreras epiteliales (Wandall et al., 2021; Zeng et al., 2020, 2021). Además, en procesos tumorales agresivos se encuentra muy presente este tipo de glicosilación aberrante (Cummings, 2019; Kudelka et al., 2015).

Debido a la relevante importancia de esta proteína en estos procesos enunciados y a su rol en la elongación de *O*-glicanos, se propone esta tesis doctoral en la que se pretende estudiar estructural y mecanísticamente el funcionamiento de C1GalT1.

2. INTRODUCCIÓN

2.1. Las Mucinas

La mucosa es un componente secretado, viscoso y gelatinoso que cubre la mayoría de superficies epiteliales en diferentes organismos. Se encuentra formado por agua, sales, inmunoglobulinas, proteínas secretadas y mucinas. Estas últimas son proteínas de elevado peso molecular que se caracterizan por su elevado nivel de glicosilación. Tan solo el 25 % del peso total de estas proteínas se debe a su cadena polipeptídica, correspondiendo el 75 % de su peso molecular restante a los glicanos que la forman (Bansil et al., 1995). Descubiertas y descritas por primera vez hace más de 100 años, estas proteínas se encuentran tanto en las secreciones mucosas como en las superficies endoteliales formando barreras protectoras ricas en glicanos, revistiendo la superficie luminal de las células endoteliales para separarlas las condiciones abrasivas químicas y microbiológicas de, por ejemplo, el intestino (Paone & Cani, 2020).

Son el tipo de proteínas que más *O*-glicanos contienen, haciendo que el tipo de glicosilación relacionada con ellas se denomine también glicosilación de tipo mucina, variando la longitud de estos desde monosacáridos hasta cadenas con más de 30 carbohidratos unidos. Entre estos glicanos presentes en las mucinas se incluyen, por ejemplo, los epítomos de los grupos sanguíneos (Varki et al., 2022).

2.1.1. Estructura y funciones principales de las mucinas.

Las mucinas son proteínas con cientos de modificaciones postraduccionales, en la que destacan sobre todo la presencia de glicanos heterogéneos elongados a partir de la *O*-glicosilación de tipo GalNAc unidos a la cadena polipeptídica de la proteína, a la cual se le denomina apomucina. Este tipo de modificación postraduccional comienza en el aparato de Golgi (Röttger et al., 1998), con la transferencia de un grupo GalNAc a un residuo de serina o treonina mediante un enlace *O*-glicosídico.

En menor medida, también pueden encontrarse residuos sulfatados, fosforilados e incluso residuos con *N*-glicanos unidos a ellos (Chugh et al., 2015). Actualmente se han identificado más de 20 genes que codifican proteínas del tipo mucina (MUC). Aunque se clasifiquen en dos tipos en función de si son proteínas secretadas o proteínas transmembrana (Rachagani et al., 2009), ambos tipos se encuentran implicados en la regulación u homeostasis de células epiteliales, actuando como principal barrera de defensa de dichas células (Chugh et al., 2015). Mucinas secretadas serían, por ejemplo, MUC2, MUC5AC, MUC5AB, MUC6, MUC7 y MUC19 (Figura 2.1), las cuales no tienen dominio transmembrana y además tienen en común un dominio en el extremo C-terminal con abundantes residuos de cisteína y un dominio denominado Factor D de von Willebrand (vWF-D), que tienen como función la oligomerización entre ellas, otorgándoles a estas proteínas su característica consistencia gelificada (Bansil et al., 2013; Pérez-Vilar & Hill, 1999).

El otro grupo de mucinas se caracteriza por anclarse a la superficie celular mediante un dominio intermedio transmembrana, de ahí el nombre del grupo al que pertenecen. Forman parte de la clasificación de proteínas ancladas a membrana de tipo I, lo que significa que tienen la región del N-terminal en la parte extracelular y una cola citoplasmática orientada hacia la parte citosólica (Olson et al., 2005) (Figura 2.2). También tienen dominios externos, tanto en la parte extracelular como intracelular, que pueden llegar a extenderse hasta 100 nm más allá de la superficie de la célula, ayudando en su implicación en las vías de señalización celular, en el caso de las colas citosólicas, y potenciando la defensa celular en los extremos extracelulares. A este grupo pertenecen proteínas como MUC1, MUC3, MUC4, MUC12, MUC16 y MUC17. La heterogeneidad y complejidad estructural de estas proteínas se explica gracias a las diferentes modificaciones postraduccionales que sufren, aunque principalmente la mayoritaria es la *O*-glicosilación (Hollingsworth & Swanson, 2004).

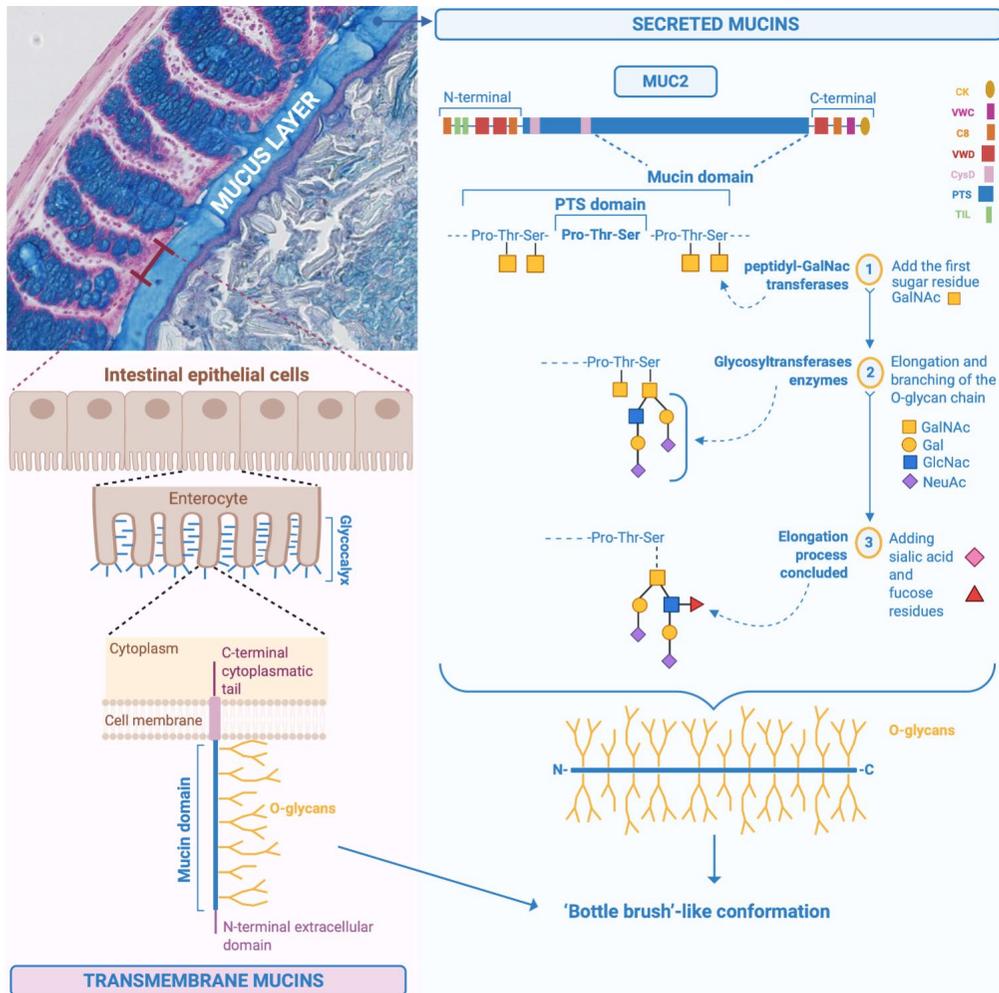


Figura 2.1 Mucinas de transmembrana y secretadas. Representación gráfica y localización de mucinas secretadas y de transmembrana en el epitelio intestinal. Se encuentran representada las diferentes zonas comunes en las mucinas transmembrana (abajo izquierda) y específicamente la estructura de MUC2, una mucina secretada (arriba derecha). Aparecen representadas el inicio y elongación de la glicosilación específica en los dominios PTS de la apomucina de MUC2. Fuente (Paone & Cani, 2020).

La zona extracelular de las mucinas se compone principalmente de tres dominios bien definidos por investigadores como Hollingsworth y Swanson (2004) o Kufe (2009).

El primero, formado por repeticiones en tándem de número variable (VNTR), es el responsable de la alta glicosilación de estas proteínas, y es el rasgo distintivo de

esta familia de proteínas. ya que contienen una elevada cantidad de residuos de prolina, treonina y serina, siendo los dos últimos aminoácidos muy susceptibles a la glicosilación de tipo GalNAc. A estas regiones se les denomina **dominios PTS** por la abundancia de los residuos mencionados (Hanisch, 2001; Moniaux et al., 2001).

Dominios implicados en interacciones ligando-receptor de los cuales se desconoce su función específica en las mucinas, aunque una glicosilación errónea en ellos está asociada con determinados tipos de cáncer. A estos dominios se les denomina tipo EGF por compartir características derivadas del factor de crecimiento epidérmico (Hollingsworth & Swanson, 2004; Mukhopadhyay et al., 2013).

Por último, también se encuentran en estas zonas extracelulares unos dominios con elevada glicosilación de una longitud de al menos 120 residuos aminoacídicos. Estos dominios reciben el nombre de SEA por haberse descubierto por primera vez en la proteína espermática, la enteroquinasa y la agrina. Hay estudios que sugieren que su función está relacionada con la protección celular frente a tensiones mecánicas (Pelaseyed et al., 2013). Muchos de estos dominios tipo SEA poseen capacidad de autoproteólisis; la glicosilación de los mismos previene su degradación y reafirma la importancia de la hiperglicosilación de estas proteínas (Pei & Grishin, 2017).

Por otro lado, la región citosólica de las mucinas se encarga, al igual que los dominios tipo EGF, de procesos de señalización celular y de interacción con proteínas asociadas al citoesqueleto gracias a la presencia de diferentes sitios de fosforilación (Kufe, 2009; Senapati et al., 2008).

La función principal de las mucinas, gracias a las funciones específicas de los dominios mencionados, es la de proteger a las células de daños físicos y químicos, así como de agentes microbianos infecciosos. Es por ello que se encuentran a través de todas las superficies epiteliales del cuerpo, desde las vías respiratorias, pasando

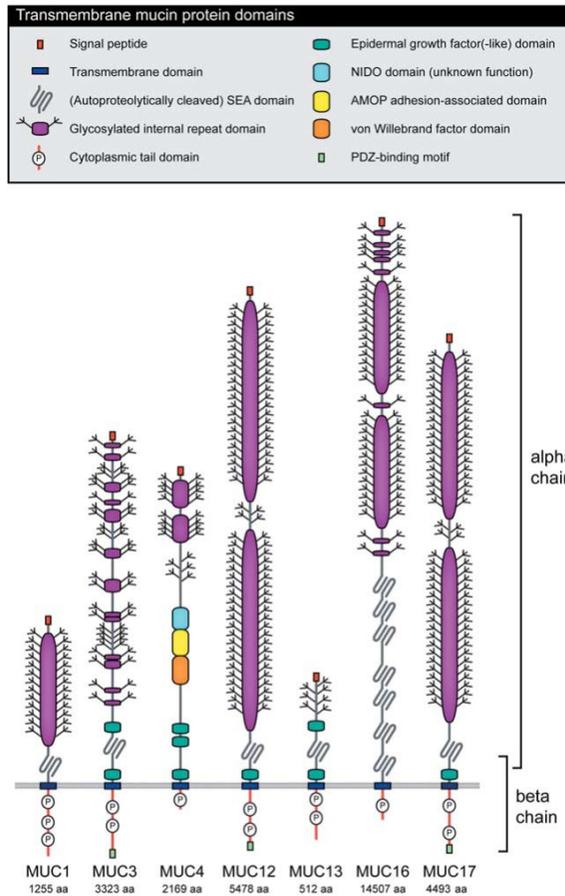


Figura 2.2. Familia de mucinas de transmembrana. Estructura de los dominios de las mucinas transmembrana MUC1, MUC3, MUC4, MUC12, MUC13, MUC16 y MUC17. Los dominios extracelulares se componen normalmente de los dominios PTS hiperglicosilados. Todas las mucinas transmembranas excepto MUC4 contienen dominios SEA donde pueden cortarse autoproteolíticamente. Los extremos N-terminal y C-terminal (cadena α y β en la imagen respectivamente) permanecen no asociados covalentemente. Varias de estas proteínas transmembrana también pueden contener dominios tipo EGF. Los dominios en la cola citoplasmática pueden ser fosforilados y funcionar como vías de señalización celular. Fuente (van Putten & Strijbis, 2017).

por los tractos gastrointestinales e incluso en los epitelios genitourinarios. Adicionalmente, dependiendo de la mucina, pueden funcionar como repelentes de interacciones con la superficie celular o justo lo contrario, promover esas interacciones, ayudando a la adhesión celular mediante la unión a proteínas de unión a carbohidratos capaces de reconocer los *O*-glicanos presentes en las

mucinas. Estas interacciones otorgan la capacidad de transmisión de señales, ganando relevancia en la proliferación, supervivencia y diferenciación celular.

En los siguientes apartados también se tratarán cómo las deficiencias en la glicosilación correcta de los dominios de las mucinas conllevan a la pérdida de ciertas propiedades, que pueden acabar en diferentes patologías.

2.2. Síntesis e inicio de la *O*-glicosilación: GalNAc-Transferasas.

La síntesis de los *O*-glicanos tipo mucina es, como se ha mencionado previamente, una modificación postraduccional de especial importancia en organismos pertenecientes al reino de los metazoos, estando ausente en plantas, levaduras y bacterias. Más del 80% de las proteínas que circulan por la vía secretora celular son glicosiladas por la maquinaria enzimática presente en dicha vía (Bennett et al., 2012; Gill et al., 2011). Aun así, tan solo un 1% del genoma se encarga de codificar genes implicados en la *O*-glicosilación, la cual se encuentra altamente conservada en el curso evolutivo de muchas especies (Apweiler et al., 1999; Ho et al., 2016; Lowe & Marth, 2003). El inicio de esta modificación ocurre con la transferencia de *N*-acetilgalatosamina o GalNAc al grupo hidroxilo de la cadena lateral de un residuo de serina o treonina de la proteína que está sufriendo la glicosilación, en condiciones fisiológicas normales en el aparato de Golgi (Gill et al., 2011; Hang & Bertozzi, 2005; Röttger et al., 1998; K. G. ten Hagen et al., 2003b). Esta reacción es catalizada por una de las isoformas de la familia de las enzimas conocidas ppGalNAc-Ts, las cuales se encuentran englobadas en el grupo GT27 de la base de datos de enzimas que procesan carbohidratos (“CAZy Database”; <http://www.cazy.org>) (Drula et al., 2022). El producto de la reacción está formado por un residuo de GalNAc unido mediante un enlace α -1 a serina o treonina, conocido también como antígeno Tn (GalNAc- α 1-*O*-Ser/Thr) (Chia et al., 2016; Hang & Bertozzi, 2005; Varki et al., 2022). A partir de este producto, los glicanos se extienden para formar estructuras más complejas que se detallarán en los

siguientes apartados.

En los seres humanos pueden encontrarse hasta 20 isoformas diferentes de esta familia de proteínas encargada de realizar la iniciación de la glicosilación tipo GalNAc (Bennett et al., 2012; K. G. ten Hagen et al., 2003b). Aunque el número de isoformas pueda variar entre especies, los genomas de organismos complejos contienen familias de secuencias de elevada homología, demostrando que la función y la estructura de la familia de proteínas de las ppGalNAc-Ts se encuentra altamente conservada evolutivamente (Schwientek et al., 2002; K. G. ten Hagen et al., 2003b).

A pesar de tener un número tan elevado de isoformas, se ha comprobado que su expresión varía en función de la célula y del tejido, siendo importante resaltar su desregulación a nivel de expresión en procesos tumorales (Bennett et al., 2012; Brooks et al., 2007). Como ejemplo, ha sido descrito que los genes codificantes para las GalNAc-T1 y T2 se expresan de manera más extendida en el organismo, mientras que hay otras isoformas que sólo lo hacen bajo unas condiciones específicas, como en la diferenciación celular (Mandel et al., 1999; Raman et al., 2012).

Estas enzimas comparten entre ellas la estructura típica de proteínas ancladas a transmembrana de tipo II (Figura 2.3) presentes en el aparato de Golgi, estando compuestas por una pequeña cola en el extremo N-terminal orientada al citoplasma, un pequeño dominio hidrofóbico de membrana, una región tipo tallo de longitud variable entre isoformas y un dominio catalítico en el C-terminal que se encuentra en el lumen del aparato de Golgi (Paulson & Colley, 1989; Wildt & Gerngross, 2005). La principal diferencia que presentan con respecto a otras glicosiltransferasas eucariotas es el dominio de lectina de tipo ricina que tienen en el C-terminal, formado por casi 120 aminoácidos (Imberty et al., 1997).

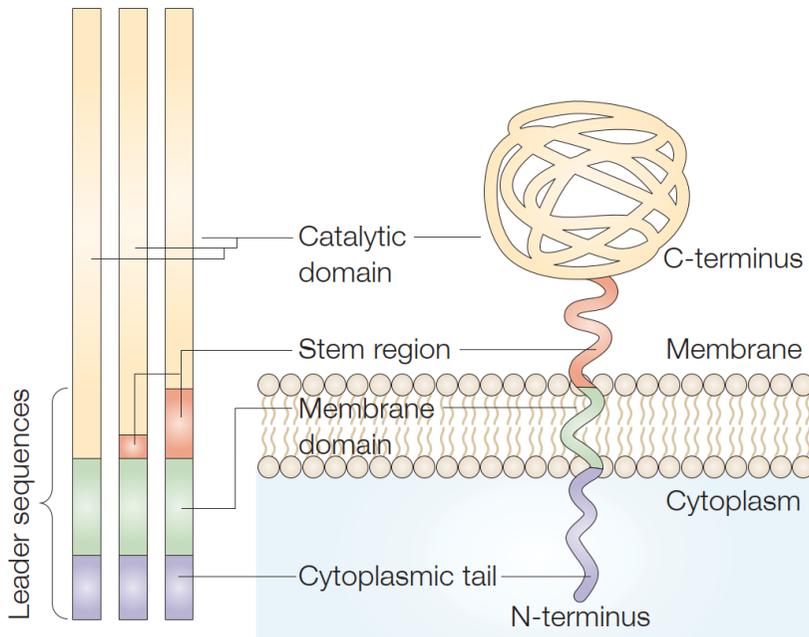


Figura 2.3. Proteínas transmembrana del tipo II. La mayoría de glicosiltransferasas se anclan a la membrana a través de un dominio de anclaje de tipo II. En este tipo de proteínas, el dominio catalítico se encuentra en el extremo C-terminal, anclado a la membrana por un dominio transmembrana, seguido de una cola citoplasmática, en el extremo N-terminal. Fuente (Wildt & Gerngross, 2005).

El dominio catalítico de las GalNAc-Ts destaca por contener un motivo estructural de tipo GT-A, caracterizado por dos plegamientos estrechamente unidos β - α - β de tipo Rossmann (Rao & Rossmann, 1973), los cuales contienen residuos capaces de unirse en este caso a UDP-GalNAc (Hanukoglu, 2015). Adicionalmente presentan un motivo conservado en todas las isoformas humanas de unión al ión Mn^{2+} definido por la secuencia DxH, siendo x cualquier aminoácido, que además interacciona con la región difosfato del UDP de la molécula donadora (Bennett et al., 2012). Es interesante mencionar que la mayoría de glicosiltransferasas que usan UDP azucarados y se unen a metal suelen tener el motivo DxH y no el DxH visto en las GalNAc-Ts (F. K. Hagen et al., 1999).

Los residuos que se unen al UDP-GalNAc en estas enzimas suelen formar parte de

un asa flexible, o *flexible loop* por su traducción en inglés, que suele sufrir grandes cambios conformacionales una vez se une al UDP-GalNAc (Hurtado-Guerrero, 2016; Lira-Navarrete et al., 2014). Además, hay un *flexible linker* que conecta el dominio catalítico de las GalNAc-Ts con su particular dominio de lectina de tipo ricina (Hazes, 1996), el cual suele contener tres potenciales sitios de unión a azúcares, aunque no todos son funcionales (Gerken et al., 2013; J. W. Pedersen et al., 2011; Revoredo et al., 2016). Existen estudios que demuestran que la función de este dominio de lectina es la modulación y mejora de la eficiencia catalítica de las GalNAc-Ts frente a sustratos con una elevada *O*-glicosilación, como las mucinas (Bennett et al., 2012; J. W. Pedersen et al., 2011; K. G. ten Hagen et al., 2003b; Wandall et al., 2007).

El mecanismo de acción de las glicosiltransferasas puede clasificarse en dos tipos en función de la estereoquímica del enlace glicosídico del carbono anomérico del sustrato donador, respecto al producto final. Es decir, en caso de que el enlace formado, por ejemplo, con la proteína, mantenga la misma conformación que en el nucleótido azucarado, será una enzima de retención; mientras que, si el enlace glicosídico final cambia con respecto al presente en el nucleótido azucarado, se trataría de una enzima de inversión. Las glicosiltransferasas de inversión suelen seguir un mecanismo de sustitución simple en el que el aceptor realiza un ataque nucleofílico al carbono 1 del azúcar donador. Las de retención, por otro lado, no parecen operar mediante un mecanismo de dos pasos que involucre la formación de un intermediario glicosiltransferasa-azúcar, sino que se propuso un mecanismo llamado S_Ni , en el que el grupo saliente y el ataque nucleofílico ocurren de forma coordinada y tienen lugar en la misma cara del glicano (Lairson et al., 2008). En el caso específico de las GalNAc-Ts, se tratan de enzimas de retención pues el producto de la transferencia mantiene la misma estereoquímica que la presente en UDP-GalNAc, la cual es α (Chan et al., 2012; Hurtado-Guerrero & Davies, 2012; Lairson et al., 2008; Wagner & Pesnot, 2010). En concreto, el mecanismo de acción

de la GalNAc-T2 fue demostrado anteriormente en el grupo de investigación en el que se ha realizado esta tesis doctoral, siendo el mismo un mecanismo tipo S_{Ni} como se propuso anteriormente para otras glicosiltransferasas (G. J. Davies, 2001; Hurtado-Guerrero, 2016; Lira-Navarrete et al., 2014).

Debido a que las GalNAc-Ts inician y, en consecuencia, determinan los sitios de *O*-glicosilación en proteínas altamente *O*-glicosiladas, como las mucinas, deben tener una serie de características que les permitan glicosilar sus dianas de manera correcta (Hollingsworth & Swanson, 2004; Kufe, 2009). El gran número de isoformas de estas enzimas, las cuales son semi-redundantes entre algunas de ellas, se explica debido a sus preferencias de sustrato a la hora de realizar la transferencia del grupo GalNAc. Gracias a estas preferencias, la familia de las GalNAc-Ts puede ser clasificada en (Revoredo et al., 2016) (Figura 2.4):

1. Isoformas con selectividad hacia péptidos y glicopéptidos con un GalNAc previo situado lejos del nuevo sitio de glicosilación (GalNAc-T1 y GalNAc-T2).
2. Isoformas con especificidad mixta entre péptidos y glicopéptidos (GalNAc-T4).
3. Isoformas con especificidad estricta hacia glicopéptidos (GalNAc-T7 y GalNAc-T10).

Esta distinción, sin embargo, se basa en la actividad frente a sustratos que carecen o contienen uno o más motivos GalNAc-*O*-Ser/Thr, permitiendo clasificarlas en glicosiltransferasas tempranas, intermedias o tardías. Así, por ejemplo, GalNAc-T2 sería temprana, GalNAc-T4 intermedia y GalNAc-T10 tardía (Bennett et al., 2012; Pratt et al., 2004; Revoredo et al., 2016).

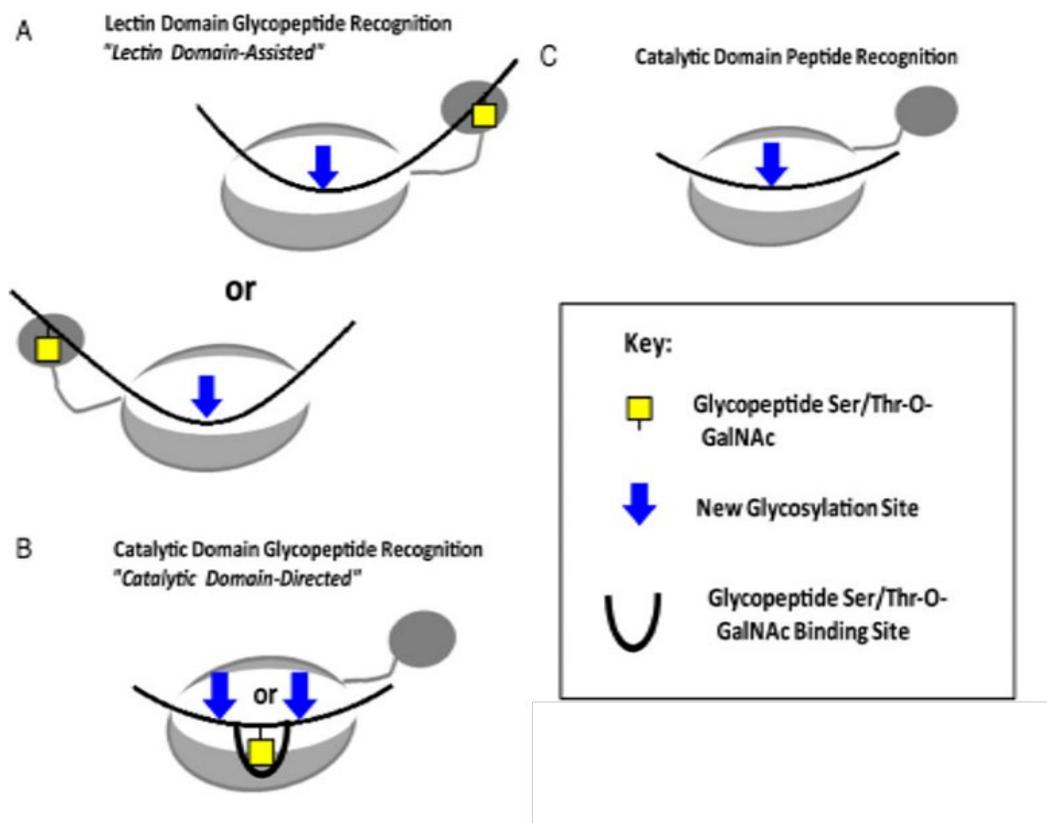


Figura 2.4. Representación gráfica de la asistencia de los dominios catalíticos y de lectina de las GalNAc-Ts sobre péptidos o glicopéptidos. A. El reconocimiento del sustrato se lleva a cabo por el dominio de lectina. B. Reconocimiento mediado por el dominio catalítico. C. El reconocimiento del péptido es mediado por el dominio catalítico. Los óvalos grandes representan el dominio catalítico y los pequeños, tanto a la izquierda como a la derecha, el dominio de lectina. Las líneas gruesas representan el péptido o glicopéptido. Fuente: adaptado de (Revoredo et al., 2016).

La actividad de las GalNAc-Ts sobre los glicopéptidos ha sido también clasificada en otras dos clases más, basándose en su capacidad de glicosilación de corto alcance (o vecina) y de largo alcance (o remota). El primer tipo tiene como preferencia sustratos en los que el residuo de azúcar se encuentre unido al dominio catalítico, implicando una glicosilación entre 1 y 3 aminoácidos de distancia con respecto al sitio previamente glicosilado. Por otro lado, la remota prefiere sustratos en los que el azúcar permita la unión al dominio de lectina, lo que hace que se dirija el sitio aceptor distante (entre 6 y 17 residuos de distancias) al dominio catalítico para que se produzca la glicosilación (Bennett et al., 2012; de las Rivas et al., 2017). Tanto

las isoenzimas de corto como de largo alcance pueden realizar la transferencia en una orientación específica, ya sea en orientación N-terminal o C-terminal. Incluso hay algunas isoenzimas que pueden realizarlo en ambas direcciones (de las Rivas et al., 2019).

Aunque la mayoría de isoenzimas de esta familia sea capaz de glicosilar sustratos aceptores comunes, sugiriendo cierta redundancia entre ellas (Gerken et al., 2011; Revoredo et al., 2016), de manera paradójica se ha demostrado como varias isoenzimas de las GalNAc-Ts pueden ser muy específicas hacia algunos sustratos proteicos. Este hecho ha podido ser identificado gracias a la aproximación diseñada por el grupo del Dr. Henrik Clausen denominada “Simple Cell” (SC) (Schjoldager et al., 2012, 2015; Steentoft et al., 2011, 2013; Yang et al., 2014). Esta técnica se basa en la generación de líneas celulares en las que el gen codificante para *Cosmc* se encuentra silenciado, impidiendo la elongación del antígeno Tn (Apartado 2.3) al no poder C1GalT1 realizar un correcto plegamiento. De esta forma, pueden analizarse la glicosilación realizada por las diferentes isoformas de GalNAc-Ts presentes en cada línea celular (Schjoldager et al., 2015; Yang et al., 2014). Gracias a esta técnica se identificaron sustratos específicos tanto para GalNAc-T2 (ApoC-III) (Schjoldager et al., 2012) como para GalNAc-T11 (repeticiones de clase A de la familia de receptores de LDL)(N. B. Pedersen et al., 2014; S. Wang et al., 2018).

Aunque gracias a todos estos avances se ha podido arrojar algo de luz al funcionamiento y al balance entre promiscuidad y especificidad de las GalNAc-Ts, la regulación de dicho balance queda aún sin respuesta. Por tanto, es necesario seguir ahondando en la investigación en estas enzimas para desentrañar este tipo de cuestiones.

2.3. Elongación de los *O*-glicanos.

2.3.1. Síntesis de los núcleos de los *O*-glicanos.

En condiciones, normales, una vez las GalNAc-Ts han realizado su función correctamente, los residuos de serina o treonina glicosilados con GalNAc, o antígeno Tn, son elongados para producir hasta ocho estructuras centrales o “*cores*” (Figura 2.5). Los *cores* 1, 2, 3 y 4 son los más frecuentes y los que suelen sufrir una extensión de forma escalonada por un gran número de glicosiltransferasas. Esta elongación da lugar a una gran heterogeneidad topológica, surgiendo cadenas de glicanos que van desde estructuras lineales simples, hasta estructuras extensamente ramificadas y complejas (Ho et al., 2016).

A partir del antígeno Tn se producen las diferentes ramificaciones mediante las enzimas mencionadas a continuación. La T-sintasa, *core* 1 β 1,3-galactosiltransferasa o C1GalT1 transfiere una molécula de galactosa, desde el donador UDP-galactosa, al antígeno Tn, formando el *core* 1 (Gal β 1-3GalNAc, antígeno T) (Cao et al., 1996), siendo los *core* 1 *O*-glicanos las estructuras azucaradas más comunes y detectadas en la mayoría de células (Bergstrom et al., 2017; Ju, Cummings, et al., 2002; Ju & Cummings, 2002). El *core* 2 se forma por la adición y ramificación al antígeno T de una molécula de GlcNAc, llevado a cabo principalmente por la enzima GCNT-3, aunque también pueden realizar esta transferencia las enzimas GCNT-1 y 4, o *core* 2 β 1,6-*N*-acetilglucosaminiltransferasa (Schwientek et al., 1999), el cual es más específico de algunos tipo celulares y su expresión está altamente regulada durante la activación de linfocitos, la estimulación de citoquinas y el desarrollo embrionario (Brockhausen et al., 2022).

A partir del antígeno Tn se producen las diferentes ramificaciones mediante las enzimas mencionadas a continuación. La T-sintasa, *core* 1 β 1,3-

galactosiltransferasa o C1GalT1 transfiere una molécula de galactosa, desde el donador UDP-galactosa, al antígeno Tn, formando el *core 1* (Gal β 1-3GalNAc, antígeno T) (Cao et al., 1996), siendo los *core 1* O-glicanos las estructuras azucaradas más comunes y detectadas en la mayoría de células (Bergstrom et al., 2017; Ju, Cummings, et al., 2002; Ju & Cummings, 2002). El *core 2* se forma por la adición y ramificación al antígeno T de una molécula de GlcNAc, llevado a cabo principalmente por la enzima GCNT-3, aunque también pueden realizar esta transferencia las enzimas GCNT-1 y 4, o *core 2* β 1,6-N-acetilglucosaminiltransferasa (Schwientek et al., 1999), el cual es más específico de algunos tipo celulares y su expresión está altamente regulada durante la activación de linfocitos, la estimulación de citoquinas y el desarrollo embrionario (Brockhausen et al., 2022).

De manera alternativa a la formación del *core 1*, puede ocurrir la transferencia de una molécula de GlcNAc en vez de galactosa mediante un enlace β -1,3, formándose la estructura denominada *core 3* mediante la enzima B3GNT6 (Iwai et al., 2002), la cual parece tener solo actividad en epitelios mucosos como los tractos gastrointestinal y respiratorio y las glándulas salivares. En células de colon, por ejemplo, esta enzima se encuentra sobre expresada, lo que produce que las mucinas en el colon se encuentren altamente glicosiladas con O-glicanos que contienen el *core 3*, a diferencia de otras líneas celulares en las que las estructuras más abundantes son la del *core 1* (Capon et al., 2001). Esta estructura sirve como sustrato aceptor a la enzima encargada de realizar la síntesis del *core 4*, GCNT3, la cual añade GlcNAc mediante un enlace β -1,6 al residuo de GalNAc de igual manera que lo hace en el *core 1*. De esta forma, se pueden diferenciar las tres N-acetilglucosaminiltransferasas (GCNT) en dos grupos, dependiendo de si son capaces de formar sólo el *core 2* – GCNT1 y GCNT4 o GCNT de tipo L – o si además de formar el *core 2* puede también transferir el azúcar al *core 3*, formando el *core 4* – GCNT3 o GCNT de tipo M – (Brockhausen et al., 2022).

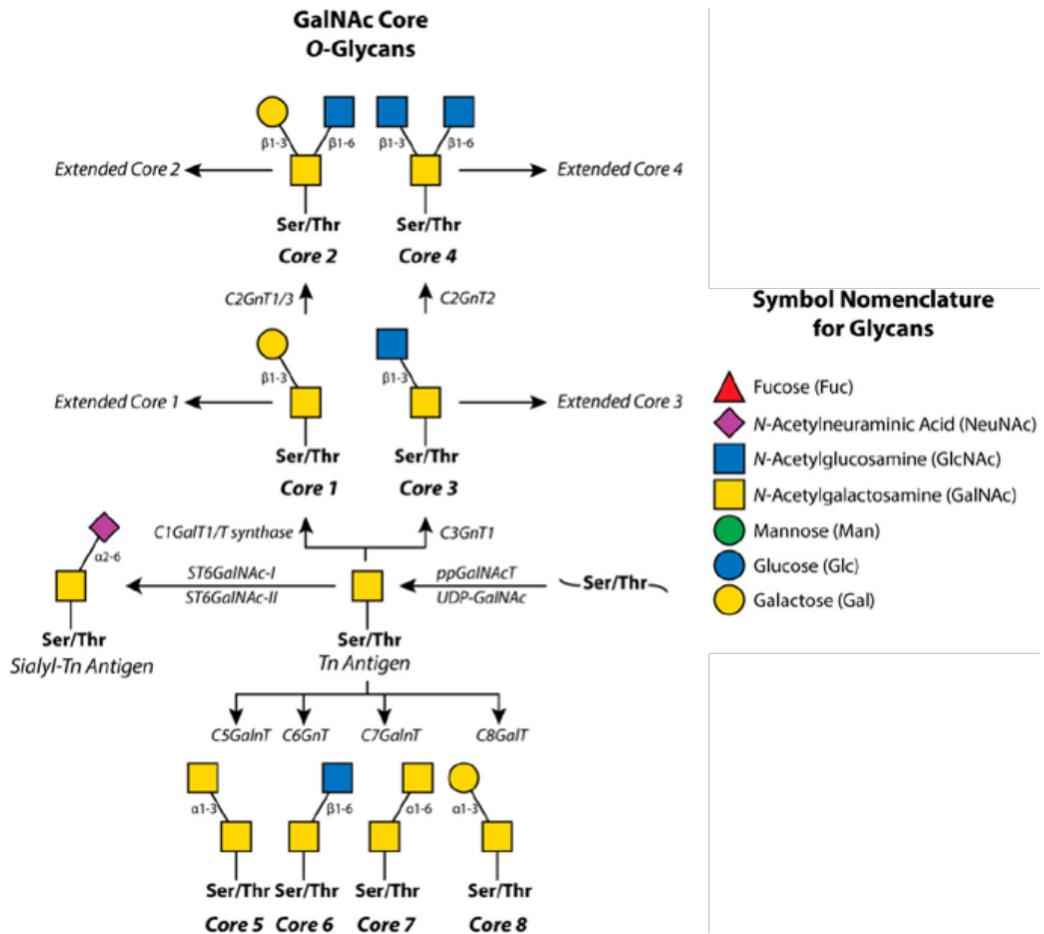


Figura 2.5. Estructuras y vías de síntesis de los O-glicanos. Los símbolos usados para los glicanos vienen determinados por la SNFG (Symbol Nomenclature for Glycans) (Neelamegham et al., 2019; Varki et al., 2015). Figura modificada a partir de (Wilkinson & Saldova, 2020).

Existen además otros núcleos, denominados *cores* 5, 6, 7 y 8, cuya frecuencia de aparición es extremadamente escasa, no habiéndose encontrado el *core* 7 en humanos a día de hoy (Figura 2.5) (Varki et al., 2022).

2.3.2. Extensión de los *cores* de O-glicanos.

Regulado por la acción de diferentes glicosiltransferasas, los mencionados *cores* se elongan para formar un amplio abanico de glicanos complejos, que incluyen los epítomos de grupos sanguíneos ABO y de Lewis, ácidos polisialilados, el antígeno

i linear (Gal β 1-4GlcNAc β 1-3Gal), y los antígenos I de GlcNAc β 1-6-ramificados. Las extensiones de glicanos del tipo 1 (Gal β 1-3GlcNAc) o del tipo 2 Gal β 1-4GlcNAc pueden repetirse y hacer de andamio para la unión de azúcares o grupos funcionales adicionales. Los sitios terminales de los *O*-glicanos pueden contener enlaces tipo α con fucosas o ácido siálicos, galactosa, GalNAc y GlcNAc tanto mediante enlaces tipo α como tipo β , y sulfatos (Varki et al., 2022). Muchos de estos grupos terminales son reconocidos por proteínas como son las selectinas (Cummings, Chiffolleau, et al., 2022), en el caso de los antígenos de Lewis sialilados y sulfatados, o por galectinas en caso de que las estructuras terminen en galactosa (Cummings, Liu, et al., 2022). De cualquier forma, todas estas estructuras juegan un papel importantísimo en la inmunidad celular y la homeostasis de la misma (Ellies et al., 1998; Homeister et al., 2001; Yeh et al., 2001), interacciones célula-célula (Homeister et al., 2001; Leppänen et al., 1999; McEver et al., 1995; Yeh et al., 2001), biología vascular (Tenno et al., 2007), angiogénesis (L. Xia et al., 2004), linfangiogénesis (J. Fu et al., 2008) y génesis y funcionamiento de las plaquetas (Y. Wang et al., 2012). Los *cores* y epítomos antigénicos más relevantes se recogen en la tabla 2.1.

Tras atisbar la profundidad, variedad y complejidad de todo este tipo de modificaciones que hacen posible la formación de estas estructuras de *O*-glicanos tan complejas, es totalmente comprensible observar un abismo conforme las estructuras se van ramificando. Es por ello que es necesario enfocarse en pasos específicos de esta modificación postraducciona l e ir intentando descubrir poco a poco el funcionamiento y regulación de todos estos pasos.

Tabla 2.1. Núcleos o *cores* de los glicanos de tipo GalNAc y epítomos antigénicos de las mucinas.
Adaptado de (Brockhausen et al., 2022).

Core	Estructura
Antígeno Tn	GalNAc- α -Ser/Thr
Antígeno STn	Sia- α 2,6-GalNAc- α -Ser/Thr
Core 1 o antígeno T	Gal- β 1,3-GalNAc- α -Ser/Thr
Antígeno ST	Sia- α 2,3-Gal β 1-3GalNAc α 1-O-Ser/Thr
Core 2	GlcNAc- β 1,6(Gal- β 1,3)-GalNAc- α -Ser/Thr
Core 3	GlcNAc- β 1,3-GalNAc- α -Ser/Thr
Core 4	GlcNAc- β 1,6(GlcNAc- β 1,3)-GalNAc- α -Ser/Thr
Epítomo	Estructura
Grupo sanguíneo H	Fuc- α 1,2-Gal-
Grupo sanguíneo A	GalNAc- α 1,3(Fuc- α 1,2)-Gal-
Grupo sanguíneo B	Gal- α 1,3(Fuc- α 1,2)-Gal-
Grupo sanguíneo linear B	Gal- α 1,3-Gal-
Grupo sanguíneo i	Gal- β 1,4-GlcNAc- β 1,3-Gal-
Grupo sanguíneo I	Gal- β 1,4-GlcNAc- β 1,6(Gal- β 1,4-GlcNAc- β 1,3)-Gal-
Grupo sanguíneo Sd(a), Cad	GalNAc- β 1,4(Sia- α 2,3)-Gal-
Grupo sanguíneo Lewis a	Gal- β 1,3(Fuc- α 1,4)GlcNAc-
Grupo sanguíneo Lewis x	Gal- β 1,4(Fuc- α 1,3)GlcNAc-
Grupo sanguíneo S Lewis x	Sia- α 2,3-Gal- β 1,4(Fuc- α 1,3)GlcNAc-
Grupo sanguíneo Lewis y	Fuc- α 1,2-Gal- β 1,4(Fuc- α 1,3)GlcNAc-

2.4. C1GalT1 y *COSMC*

La *core* 1 β 3galactosiltransferasa, *core* 1 β 3GalT, C1GalT1 o T-sintasa (EC 2.4.1.122) es la principal protagonista de esta tesis doctoral y la única encargada de realizar la síntesis del *core* 1, también conocido como antígeno T, en la ruta de glicosilación de tipo mucina. Esta estructura, a la que se referirá a partir de ahora como *core* 1 o antígeno T, es el precursor más común de una extensa variedad de estructuras complejas de *O*-glicanos de tipo mucina en mucinas, glicoproteínas de membrana y en general proteínas que circulan por la vía secretoria (Ju et al., 2011). Convirtiéndose, esta enzima, en una de las glicosiltransferasas más importantes en la biosíntesis y biología de los *O*-glicanos de tipo mucina.

Curiosamente, para que la T-sintasa en vertebrados, incluyendo humanos, sea activa necesita la presencia de una chaperona molecular específica llamada Cosmc (chaperona molecular específica de la *core 1* β 3galactosiltransferasa) o C1GalT1C1 (chaperona 1 de la *core 1* β 3galactosiltransferasa), la cual hace que su plegamiento *in vivo* se realice correctamente, permitiendo que pueda realizar su función. Varios estudios han demostrado que la regulación de la biosíntesis de *O*-glicanos de tipo mucina está principalmente regulada por la relación entre C1GalT1 y Cosmc (Aryal et al., 2010; Ju et al., 2008; Ju & Cummings, 2002, 2005; Y. Wang et al., 2010).

2.4.1. Historia de C1GalT1.

El antígeno T y producto de la reacción de transferencia de la C1GalT1, más tarde designado como *core 1*, fue descubierto antes que la actividad de la propia enzima. La primera persona en detectar este antígeno fue Thomsen, en una muestra de sangre contaminada por bacterias, que en colaboración con Friedenreich en 1930 por primera vez publicaron un estudio definiendo dicha estructura, haciendo que se le denominara como antígeno TF (Thomsen-Friedenreich) o T para acortar (Friedenreich, 1931). Tres décadas más tarde, la estructura del antígeno T fue resuelta (Klenk & Uhlenbruck, 1960). Su actividad enzimática fue caracterizada por primera vez a partir de la glándula submaxilar porcina en 1971 (Schachter et al., 1971) usando mucinas de la glándula submaxilar ovina tratada con sialidasa como aceptor y UDP-Gal como donador, aunque su actividad se ha observado en muchos otros tejidos, como en hígado de rata (Brockhausen et al., 1992), y líneas celulares tales como células de riñón de hámster (BHK) (Stojanovic et al., 1984).

Durante muchos años se realizaron intentos para purificar esta enzima desde diferentes fuentes naturales como hígados de rata (Brockhausen et al., 1992) o embriones de pollo (Furukawa & Roth, 1985), pero no fue hasta 2002 que se consiguió purificar homogéneamente por Tongzhong Ju y Richard Cummings (Ju

& Cummings, 2002) a partir de microsomas de hígado de rata. Basándose en la secuencia N-terminal de la T-sintasa de rata, se clonó el ADN codificante de la T-sintasa humana (Ju, Brewer, et al., 2002) y se descubrió el tamaño de su secuencia genética (1,8 kb) así como el número de aminoácidos que componen la enzima, 363 concretamente, la cual se trata de una proteína de transmembrana del tipo II. La secuencia se compone de un dominio citoplásmico en el extremo N-terminal (6 aas), una región hidrofóbica transmembrana (26 aas), seguido de una región tipo tallo y el dominio catalítico de la enzima. Además, análisis de *Northern blot* revelaron transcritos de C1GalT1 en todos los tejidos examinados, indicando que esta enzima se expresa de forma general en todas las células del cuerpo, pero con un rango de variación en su nivel de expresión dependiendo del tejido (Ju, Brewer, et al., 2002).

También fueron clonados o identificados los ortólogos de C1GalT1 de varias especies vertebradas: chimpancé, vaca, perro, rata, ratón, pájaro, rana y pez cebra; así como de invertebrados: mosca y *C. elegans* (Ju, Brewer, et al., 2002; Ju et al., 2006). Los alineamientos múltiples de secuencia aminoacídica muestran una elevada homología entre especies, especialmente entre las especies de mamíferos. Es curioso que esta enzima sea la única que comparte muy pocas similitudes con motivos conservados en otras β 3 galactosiltransferasas. Además, cabe destacar la existencia en todos los ortólogos de dos dominios presentes en todas las especies: un motivo DDD y otro CCSD (Ju et al., 2006).

Aunque en especies inferiores como *Drosophila melanogaster* existen al menos tres putativos genes homólogos que codifican la C1GalT1 – C1GalTA (CG9520) está confirmado que codifica una C1GalT1 altamente activa – (Ju, Brewer, et al., 2002; R. Müller et al., 2005), en organismos superiores, y específicamente en mamíferos, existe un único gen funcional codificante de C1GalT1, denominado *C1GALT1* (Ju, Brewer, et al., 2002; L. Xia et al., 2004). A diferencia de la mayoría de

glicosiltransferasas, los ortólogos de la T-sintasa de mamíferos no presentan secuencias de *N*-glicosilación (-Asn-X-Ser/Thr-), sin embargo, la de organismos inferiores como *C. elegans*, por ejemplo, tiene 4 sitios potenciales de *N*-glicosilación. Es curioso también destacar que la *CeC1GalT1* fue expresada de manera recombinante en células de insecto, teniendo cierta actividad galactosiltransferasa, la cual disminuyó muchísimo al expresarse en células de mamífero. La razón por la diferencia de esta actividad en diferentes células no es conocida (Ju et al., 2006).

2.4.2. Historia de *Cosmc*

Desde hace tiempo es conocido que la línea de células T-linfoblastoides Jurkat carecen de la actividad de C1GalT1 y principalmente por ello sintetizan *O*-glicanos truncados como el antígeno Tn (Piller et al., 1990). Tras una serie de experimentos para entender qué factores causaban la represión de la actividad de C1GalT1 en estas células, en los que se incluyeron la transfección de un plásmido que expresase la C1GalT1 humana recombinante y el uso de inhibidores de proteosomas, se descubrió que la proteína se encontraba presente en las células mediante *Western blot*, pero seguía sin tener actividad enzimática. Por ello, se llegó a la conclusión de que estas células debían de carecer de algún o algunos factores críticos para el plegamiento o estabilidad de C1GalT1. Se acabó descubriendo que este factor crítico era la chaperona molecular específica de C1GalT1, más tarde denominada *Cosmc*, cuyo ADNc codifica para una supuesta proteína transmembrana de tipo II sin actividad enzimática. Esta chaperona se encontraba mutada en la línea celular Jurkat, pudiéndose corregir la actividad de C1GalT1 y la estructura de los *O*-glicanos, transfectando la *Cosmc* silvestre (Ju & Cummings, 2002). La expresión de C1GalT1 y de *Cosmc* se encuentra ubicuamente coordinada, indicando la cercana relación biológica entre ellas (Ju et al., 2008). Es importante, además, recalcar la falta de actividad propia de *Cosmc* y su localización celular en diferentes

lugares que C1GalT1, encontrándose esta en el aparato de Golgi mientras que Cosmc reside principalmente en el retículo endoplasmático (ER) (Ju et al., 2008; Sun et al., 2011). Esto es posible gracias al dominio transmembrana de Cosmc, haciendo que se localice en el retículo endoplasmático y que forme un homodímero, mediante un puente disulfuro, siendo ambas características críticas para que ocurra el correcto plegamiento y funcionamiento de C1GalT1 (Sun et al., 2011). Una de las cosas más interesantes de esta chaperona es que no se encuentra presente en organismos invertebrados, apareciendo solo en organismos vertebrados con una elevada conservación en su secuencia. Además, solo se ha detectado un gen que codifique para esta proteína (Ju et al., 2008).

Debido a que el mecanismo de funcionamiento mediante el cual Cosmc ayudaba al plegamiento de C1GalT1 no era conocido, los investigadores Ju y Cummings propusieron un modelo que se ilustra en la figura 2.6. Según este modelo, Cosmc, se localiza en el retículo endoplasmático y se une a C1GalT1 recién sintetizada, asistiendo al plegamiento de la misma y evitando que se formen agregados que en consecuencia se degradarían por la vía proteosómica (Ju et al., 2008; Ju & Cummings, 2002). Esta hipótesis fue respaldada más tarde y mejor descrita, indicando que en la ausencia de Cosmc, C1GalT1 mal plegada formaba agregados, acumulándose y translocándose del retículo endoplasmático al citoplasma donde se degradaba mediante el proceso de proteólisis dependiente de ubiquitinación (Aryal et al., 2012). Aun así, la incorporación de Cosmc era capaz de recuperar la actividad de la C1GalT1 desnaturalizada (Aryal et al., 2010).

Aunque el origen de Cosmc no está claro, existen evidencias que sugieren que Cosmc podría haber evolucionado a partir de C1GalT1 o que al menos ambas tienen el mismo origen. Incluso hubo un momento en el que se llegó a la conclusión de que se trataba de otra C1GalT1 (C1GalT2) debido a que comparten un 20% de identidad entre ellas (Ju & Cummings, 2002; Kudot et al., 2002). Esto fue corregido

más tarde en la literatura, reconociéndola como la chaperona de C1GalT1 (Kudot et al., 2006).

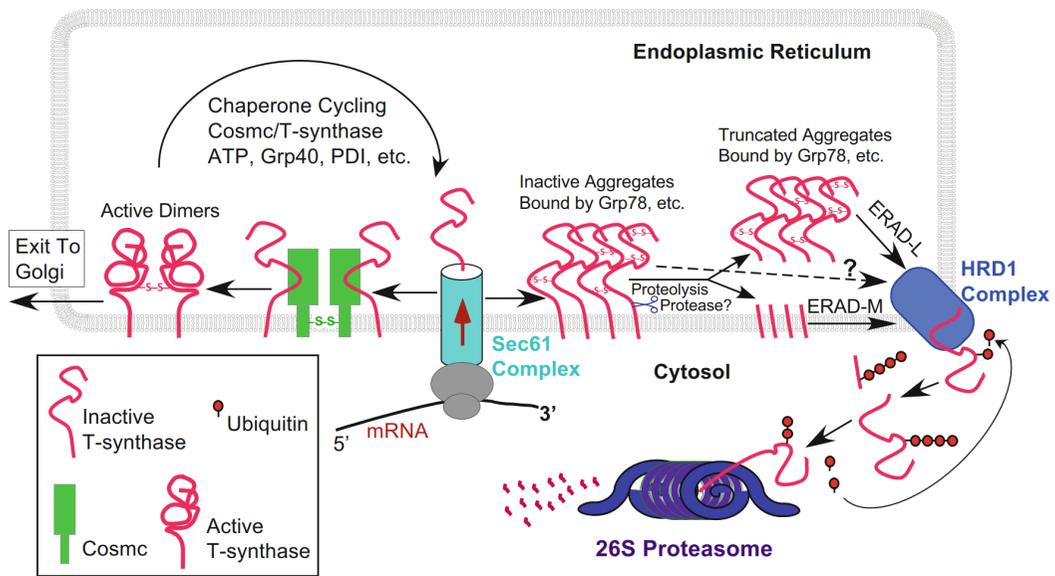


Figura 2.6. Modelo del funcionamiento de Cosmc como chaperona molecular en el retículo endoplasmático para asistir al plegamiento de C1GalT1. La Cosmc de organismos superiores (verde) funciona de manera homodimérica predominantemente en el retículo endoplasmático, interactuando con la recién sintetizada cadena aminoacídica de C1GalT1 (rojo) ayudando a su plegamiento. La T-sintasa o C1GalT1 nativa, se especula que, para funcionar correctamente necesita formar un homodímero que se desplace hasta el aparato de Golgi donde realiza la función de sintetizar el antígeno T. En ocasiones en las que *Cosmc* se encuentra mutada, disfuncional, o su expresión silenciada, las cadenas polipeptídicas de C1GalT1 forman agregados inactivos que son cortados proteolíticamente por la región del tallo de la proteína primero, siendo más tarde poliubiquitinada para ser degradada en el proteosoma. Fuente: (Ju & Cummings, 2014).

2.4.3. Estructura y actividad conocidas de C1GalT1.

La forma nativa de C1GalT1 fue purificada de hígado de rata por Ju y colaboradores en 2002, confirmando que se trataba de una proteína asociada a membrana, siendo un homodímero unido mediante un puente disulfuro, con un peso aproximado de entre 84 y 86 kDa. Según las predicciones computacionales, todos los ortólogos de C1GalT1 parecen ser proteínas transmembranas del tipo II. En humanos, rata y ratón, la proteína tiene el mismo número de aminoácidos, con más de un 90 % de identidad entre las tres. Es muy interesante destacar que en vertebrados no existe

ninguna secuencia canónica propensa a *N*-glicosilarse, a diferencia de en vertebrados inferiores, en las que la secuencias de C1GalT1 es algo más larga que en mamíferos y en algunos de ellos contienen algún que otro sitio capaz de ser *N*-glicosilado. (Ju, Brewer, et al., 2002; Ju, Cummings, et al., 2002) En invertebrados, los ortólogos son bastante más grandes en tamaño y suelen tener múltiples sitios de *N*-glicosilación (Ju, Brewer, et al., 2002; Yoshida et al., 2008). Cosmc, por otro lado, no se encuentra presente en organismos invertebrados, como *Drosophila melanogaster* o *Caenorhabditis elegans* (Ju et al., 2006), a diferencia de en organismos vertebrados, en la que es esencial su presencia en el retículo endoplasmático para el correcto plegamiento de C1GalT1 (Cummings, 2019). Además, cabe volver a resaltar que Cosmc muestra similitud de secuencia con C1GalT1 – siendo también una proteína transmembrana de tipo II que forma también un homodímero (Sun et al., 2011) – a excepción de que carece del motivo catalítico DxD, explicando su falta de actividad (Hanes et al., 2017; Y. Wang et al., 2010).

La expresión recombinante de C1GalT1 de humanos se ha llevado a cabo en diferentes líneas celulares de mamíferos (293T, CHOK1, y CHO-Lec1), obteniéndose la proteína funcional gracias a que estas líneas celulares son capaces de expresar Cosmc de manera basal, permitiendo el correcto plegamiento de la proteína recombinante (Ju, Brewer, et al., 2002; Ju et al., 2008). También se pueden expresar de manera recombinante tanto Cosmc como la T-sintasa humana en el sistema de expresión de baculovirus en células de insectos (Sf9 y Hi5). Sin embargo, a no ser que se realice una co-expresión de las dos, la C1GalT1 se produce de manera inactiva y mal plegada por no tener la actividad necesaria de su chaperona específica (Ju & Cummings, 2002, 2005, 2010). A nivel cinético y de preferencias de sustrato, aunque las GalNAc-Ts tienen claras preferencias por las secuencias peptídicas de los sustratos aceptores (de las Rivas, Coelho, et al., 2018; de las Rivas et al., 2019, 2020), no se sabe claramente hasta qué punto el primer

paso de elongación de los glicanos se encuentra mediado o no por la secuencia peptídica alrededor del residuo de GalNAc y/o incluso por las posiciones o agrupación de otros residuos de GalNAc. En cualquier caso, dos estudios diferentes usando la C1GalT1 humana y de rata con una serie de glicopéptidos señalaron que la secuencia en torno al sitio de glicosilación puede modular ligeramente los parámetros cinéticos de esta enzima (Granovsky et al., 1994; Perrine et al., 2009). Otro estudio realizado con la C1GalT1 recombinante de *D. melanogaster* expresada mediante el sistema de expresión de baculovirus fue capaz de demostrar que esta enzima era activa en diferentes glicopéptidos, aunque no se realizasen experimentos cinéticos completos (R. Müller et al., 2005). No obstante, el *O*-glicoproteoma de tipo GalNAc es increíblemente vasto y con una variación de secuencias enorme entre sitios de glicosilación. Es por ello por lo que C1GalT1 tiene que ser capaz de transferir eficientemente el residuo de galactosa a todos residuos unidos a GalNAc (Tn) indiscriminadamente e independientemente de la secuencia peptídica subyacentes y de las agrupaciones de *O*-glicanos alrededor del sitio de glicosilación (Steentoft et al., 2013).

2.5. Glicosilación disfuncional y patologías: deficiencias en C1GalT1 y Cosmc.

Aunque se han mencionado levemente, especialmente los antígenos Tn y T, es este el apartado en el que deben describirse junto a los antígenos STn y ST, además de analizar y desarrollar su relevancia en patologías. Estos antígenos son estructuras azucaradas intermedias que se producen en los procesos de formación de los *O*-glicanos. El antígeno Tn es el producto de la reacción de las GalNAc-Ts, siendo su estructura GalNAc- α 1-O-Ser/Thr, como se ha mencionado anteriormente. De igual forma, el antígeno T es la estructura formada tras la transferencia realizada por la C1GalT1, o *core* 1 (Gal β 1-3GalNAc α 1-O-Ser/Thr). Por último, la adición de ácido siálico tanto al antígeno Tn como al antígeno T hacen que se formen sus análogos sialilados, los antígenos STn y ST. El antígeno STn, se forma por la adición de

ácido siálico por parte de la enzima ST6GalNAcI mediante una ruta alternativa a la formación del antígeno T, que adiciona una molécula de ácido siálico al antígeno Tn, formándose la estructura Sia- α 2,6-GalNAc- α -Ser/Thr. Por otra parte, alternativamente a la formación del *core* 2, se forma el antígeno ST, por la acción de la ST3GalI, cuya estructura es Sia- α 2,3-Gal- β 1,3-GalNAc- α 1-O-Ser/Thr (C. Fu et al., 2016). Aunque se han detallado anteriormente, los antígenos más importantes se vuelven a mostrar en la figura 2.7.

Los antígenos mencionados son glicanos de tipo O-GalNAc que pertenecen a la familia de los antígenos Thomsen-Friedenreich. Estos antígenos se encuentran entre los glicanos antigénicos específicos de células tumorales o TACAs (por sus siglas en inglés *Tumor-associated carbohydrate antigens*) (Chia et al., 2016; Hakomori, 2001; Sletmoen et al., 2018). Los antígenos STn, T y, en especial, el antígeno Tn (Taujale et al., 2021) se encuentran en la superficie de más del 80% de células tumorales y en numerosas glicoproteínas, especialmente en las mucinas. Además de haberse descrito su papel en la patogénesis del cáncer, también se han encontrados estas glicosilaciones en ciertas enfermedades como la enfermedad inflamatoria del intestino, la nefropatía IgA o el síndrome Tn (Chugh et al., 2015; Ju et al., 2013; Taujale et al., 2021; van Putten & Strijbis, 2017).

La formación de estos atípicos glicanos puede deberse a glicosilaciones incompletas que originen estas estructuras truncadas o la expresión aberrante de cadenas de glicanos (sLE_x, SLE_a, los cuales son definidos también como TACAs) (Padler-Karavani, 2014) (Figura 2.8). La adición del primer carbohidrato, la molécula de GalNAc, se lleva a cabo en el aparato de Golgi en condiciones

normales (Gill et al., 2011; Hang & Bertozzi, 2005; Röttger et al., 1998; K. G. ten Hagen et al., 2003b), pero en condiciones tumorales, esta se ha propuesto que pueda ocurrir en el retículo endoplasmático por la familia de las GalNAc-Ts como se ha demostrado en varios estudios (Chia et al., 2016; Gill et al., 2011).

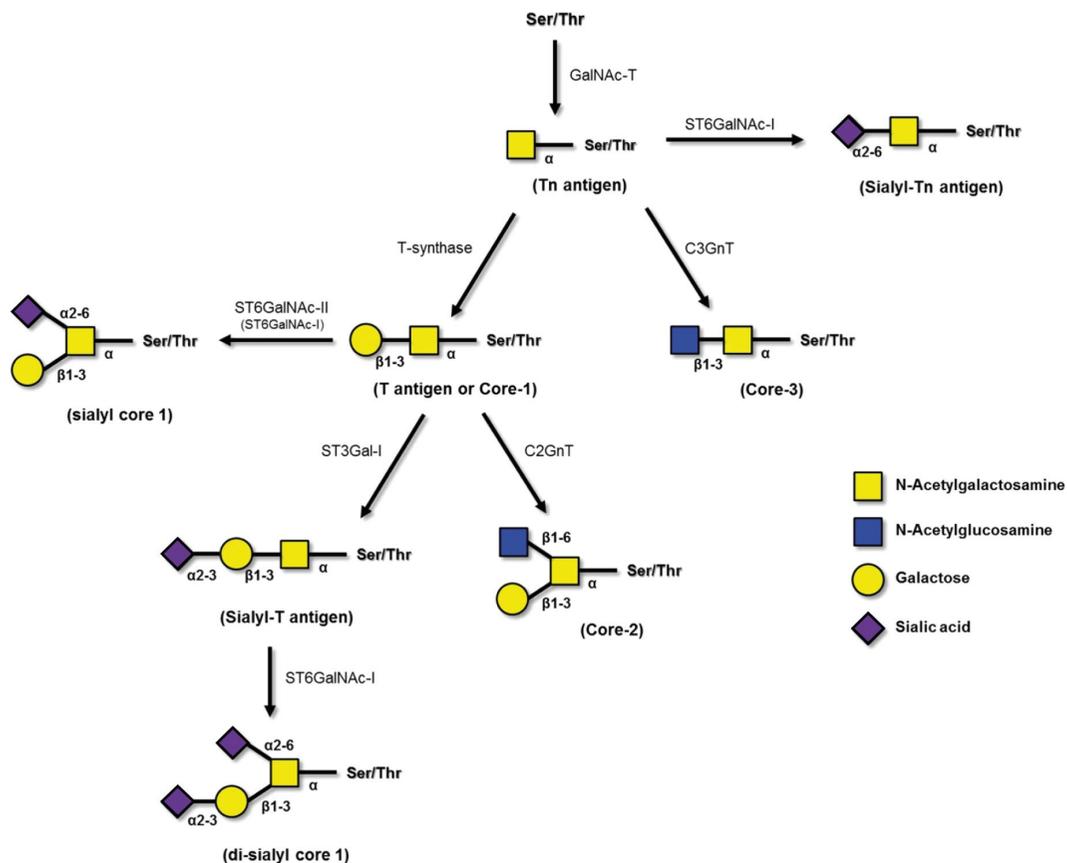


Figura 2.7. Biosíntesis de los precursores de las cadenas de O-glicanos. La síntesis de las tres vías principales de formación de los antígenos STn, T y del core 3 llevan a la extensión de los mismos para formar complejas cadenas de O-glicanos. Los monosacáridos están representados de acuerdo a lo establecido por la SNFG (Symbol Nomenclature for Glycans) (Varki et al., 2015). Adaptado de (Loureiro et al., 2018).

Cuando tiene lugar la re-localización de las GalNAc-Ts en el retículo endoplasmático desde el aparato de Golgi, se favorece el acceso a sitios secundarios de glicosilación u ocultos de manera normal en el aparato de Golgi, incrementándose la actividad de estas enzimas y, en consecuencia, la presencia de antígenos Tn y STn (Chia et al., 2016). No obstante, en la mayoría de los casos, esta desregulación fisiológica se debe al mal funcionamiento de *Cosmc* debido a la hipermetilación de su promotor transcripcional, lo que desemboca en el silenciamiento del mismo (Radhakrishnan et al., 2014). Al no poder funcionar correctamente *Cosmc*, ocurre la agregación y el mal funcionamiento de C1GalT1,

lo que elimina que se pueda añadir un residuo de galactosa al antígeno Tn, generando un desequilibrio y aumentando el número de los antígenos Tn oncogénicos.

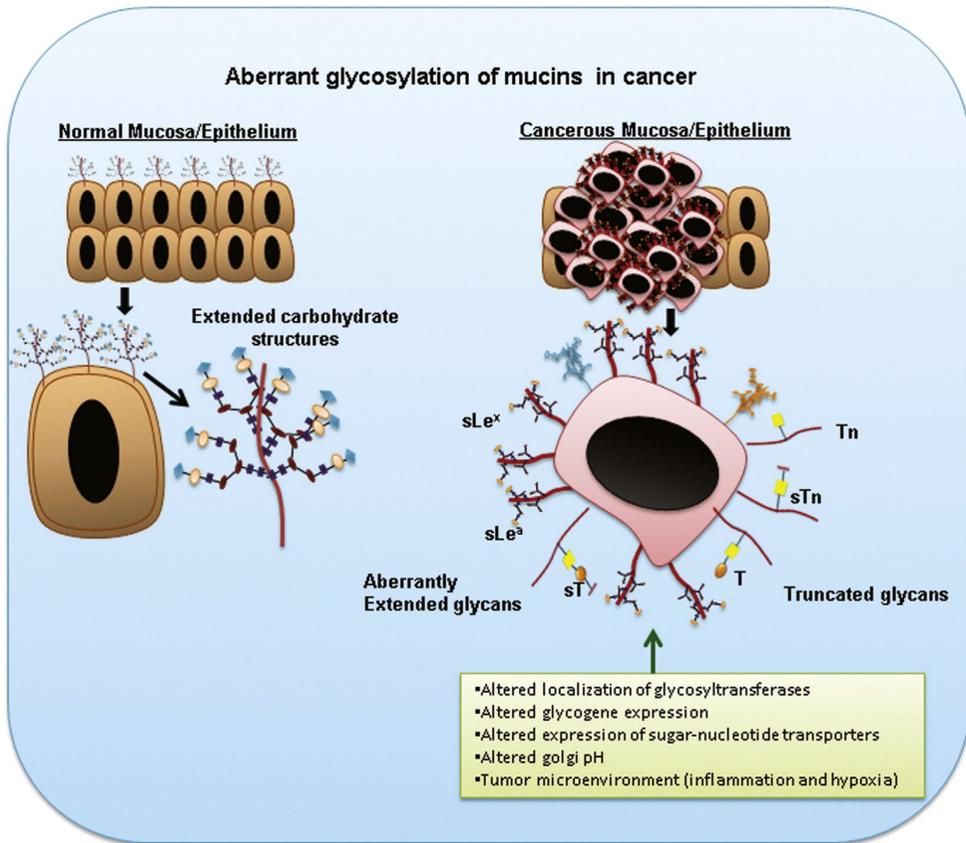


Figura 2.8. Glicosilaciones aberrantes de las mucinas en tumores y condiciones cancerosas. En tumores malignos, diferentes tipos de glicosilaciones pueden ocurrir, resultando en las estructuras truncadas como los antígenos Tn, STn y T, así como estructuras extendidas del *core 2* como son los epítomos sLe^x y SLEa. Todos estos glicanos son conocidos como glicanos antigénicos específicos de células tumorales. Esta glicosilación aberrante es el resultado de diversos factores como la desregulación de la expresión de glicoproteínas y de transportadores de nucleótidos azucarados, la deslocalización de glicosiltransferasas, pH anómalo en el aparato de Golgi y el microambiente tumoral. Fuente: (Chugh et al., 2015).

Simultáneamente al desarrollo de estos antígenos tumorales, la glicosilación implica desregulaciones importantes en el sistema inmune, lo que suele desembocar en enfermedades autoinmunes (Lübbbers et al., 2018; Mahajan & Pillai, 2016; Marth & Grewal, 2008; Maverakis et al., 2015; Reily et al., 2019). Estas glicosilaciones

aberrantes se han encontrado en diversas enfermedades, como en la nefropatía IgA (IgAN), por ejemplo. Específicamente, las interacciones entre las inmunoglobulinas de tipo lectina que se unen a ácido siálico (Siglecs) y los ácidos siálicos está muy relacionado con la autoinmunidad ya que se encargan de regular la señalización de los receptores de células B y los de tipo Toll (Zeng et al., 2021), siendo el primero el gran regulador en el desarrollo, diferenciación, supervivencia y tolerancia de las células B (Niiri & Clark, 2002; Rajewsky, 1996; Rawlings et al., 2017). Los receptores Toll además juegan un papel fundamental en la mediación de la patogénesis en enfermedades autoinmunes en respuesta tanto a señales ambientales como de desarrollo. En el contexto de la patogénesis de las enfermedades autoinmunes, ganan importancia los glicanos truncados o aberrantes porque pueden provocar cambios en la mediación del inicio y progresión de la autoinmunidad, aunque todo esto sea aún un terreno bastante inexplorado. Lo que sí queda claro es que una de las MPTs más abundante en la superficie celular de las células B es la *O*-glicosilación de tipo GalNAc, por lo que una deficiencia en la misma debe producir drásticos cambios en la función y fisiología del sistema inmune. La generación de *O*-glicanos, que se encuentran abundantemente expresados en leucocitos, requieren de la expresión de tanto *Cosmc* como de C1GalT1 (Aryal et al., 2010, 2012; Ju et al., 2008; Y. Wang et al., 2010; L. Xia & McEver, 2006). Recientemente, se ha demostrado que la eliminación del gen de *Cosmc* en células B murinas, que resulta en una sobreexpresión del antígeno Tn, reduce enormemente la migración de estas células a los nódulos linfáticos (Zeng et al., 2020). Además, se ha demostrado que en células B de ratones *knockout* de *Cosmc* ocurren indicios como la esplenomegalia e hipergammaglobulinemia que indican un potencial desarrollo de enfermedades autoinmunes (Zeng et al., 2021).

Cosmc es un interesante gen candidato a estudiar, por su capacidad de permitir la función correcta de C1GalT1, la cual es a su vez es una proteína muy interesante por su implicación en varias enfermedades autoinmunes e inflamatorias como la

enfermedad de bowel (Kudelka et al., 2020), la nefropatía IgA (Sun et al., 2015), el síndrome Tn (Lee-Sundlov et al., 2020), cáncer (Ju et al., 2014) y Alzheimer (Gollamudi et al., 2020).

El mal funcionamiento de C1GalT1, como se ha mencionado con anterioridad, se debe a la necesidad de la presencia de su chaperona específica Cosmc. En todos los casos en los que Cosmc no se encuentra presente, ya sea por el silenciamiento de su promotor transcripcional o por cualquier otra razón, C1GalT1 no va a poder realizar su plegamiento de manera correcta y por ende tampoco su función.

Además, *knockouts* homocigóticos de C1GalT1 en ratones y en *D. melanogaster* presentan letalidad embrionaria, con una angiogénesis defectuosa y hemorragia fetal en el embrión, y un fenotipo dominante del sistema central nervioso en *D. melanogaster*. Todo esto indica que la presencia de C1GalT1 es esencial para el desarrollo normal y la angiogénesis (L. Xia et al., 2004; Yoshida et al., 2008).

En humanos, varias enfermedades pueden relacionarse directamente con el mal funcionamiento de C1GalT1 y la sobreexpresión del antígeno Tn. En esta tesis se detallarán ligeramente el síndrome Tn, la nefropatía IgA y el cáncer.

2.5.1. El síndrome Tn

El síndrome Tn se trata de un trastorno hematológico poco frecuente que se caracteriza por la expresión del antígeno Tn en las subpoblaciones de células sanguíneas de diversos orígenes. Fue descubierta por primera vez en un paciente con problemas de poli-aglutinación en los eritrocitos en 1957 (Moreau et al., 1957). Clínicamente, los pacientes que padecen esta enfermedad no presentan síntomas aparentes y no suelen requerir tratamiento, aunque análisis sanguíneos pueden desvelar anemia hemolítica moderada y un número reducido de trombocitos y leucocitos (E. G. Berger, 1999). Los mecanismos que hacen que aparezca este fenotipo parecen ser multifactoriales y no se comprenden demasiado bien.

Diferentes estudios han demostrado que las glicoproteínas de los eritrocitos y leucocitos de pacientes con este síndrome tienen bajas cantidades de galactosa y de ácido siálico, lo que conlleva a una sobreexpresión del antígeno Tn que se relaciona con la deficiencia en la actividad de C1GalT1. Parece ser que esta deficiencia se debe a cambios genéticos adquiridos a través de mutaciones somáticas en *Cosmc*, causando un cambio en el marco abierto de lectura y/o en el codón de parada o incluso que la chaperona no pueda ser transcrita, resultando en una casi completa o total pérdida de la actividad de la chaperona (Crew et al., 2008; Ju & Cummings, 2005). Todas las mutaciones de *Cosmc* identificadas en la sangre de pacientes con el síndrome Tn han sido descritas por Ju y colaboradores (Ju et al., 2011), concluyendo que las células sanguíneas de pacientes con este síndrome son genéticamente mosaicas; ocurriendo en algunas líneas inmortalizadas con este síndrome la recuperación de la actividad de C1GalT1 (Thurnher et al., 1992). Mediante la creación de ratones *knockouts* de *Cosmc* en células enterocromafines y hematopoyéticas (EHC) presentes en el epitelio que cubre el tracto gastrointestinal se observó que estos animales desarrollaban megatrombocitopenia y trastornos sanguíneos, demostrando que la trombocitopenia y hemorragias en pacientes con el síndrome Tn se deben principalmente al deterioro en la biogénesis y función de las plaquetas causado por la posesión de mutaciones somáticas en el gen que codifica *Cosmc* (Y. Wang et al., 2012). En definitiva, la ausencia de una glicosilación funcional y fisiológicamente controlada conlleva a una gran variedad de trastornos en las células sanguíneas como se ha mencionado en este apartado y como se ha resumido en la revisión realizada por la Dr. Lee-Sundlov y colaboradores (Lee-Sundlov et al., 2020).

2.5.2. Nefropatía de IgA.

En segundo lugar, la nefropatía de Inmunoglobulina A (IgA), también denominada enfermedad de Berger en honor a su descubridor el Dr. Jean Berger en 1968 (J.

Berger & Hinglais, 1968), es la glomerulonefritis, o inflamación de los glomérulos de los riñones (Levy & Berger, 1988; Mestecky et al., 2013), desembocando en el fallo renal terminal en los pacientes mayores de 25 años en el 20-40 % de los casos (D'Amico, 2000). La mayoría de casos suelen ser esporádicos y no se ha conseguido identificar ningún gen causante de los mismos (Beerman et al., 2007), sin embargo, diversos factores ambientales y fisiológicos parecen ser necesarios para la aparición de esta enfermedad. La nefropatía IgA se caracteriza por el depósito de IgA1 en el mesangio renal (tejido conjuntivo localizado entre glomérulos), basándose su diagnóstico actualmente en los síntomas clínicos de la glomerulonefritis, la hematuria y proteinuria. La diagnosis final ha de realizarse mediante biopsias renales y pruebas histológicas de depósitos de IgA1 (Julian & Novak, 2004), depósitos que son los verdaderos causantes de la inflamación glomerular que conlleva al daño renal. La relación de esta enfermedad con la *O*-glicosilación de tipo mucina se observa en la deficiencia de galactosa y la simultanea aparición de los antígenos Tn y STn en la región bisagra (HR – *Hinge Region*) de la IgA1 en pacientes con esta enfermedad, lo cual puede ser el motivo del inicio de la patogénesis (Hiki, 2009; Mestecky et al., 2008). La IgA1 humana es una inmunoglobulina, secretada por las células secretoras de anticuerpos (ASC – *Antibody Secreting Cells*) (Nguyen et al., 2019), y *O*-glicosilada en nueve potenciales sitios de glicosilación en la HR, de los cuales cinco suelen estar ocupados por glicanos elongados desde el *core* 1, mono o disialilados (Mattu et al., 1998). En la nefropatía IgA, esta glicosilación se intercambia por la expresión de antígenos Tn y STn, posiblemente debido a la reducción de la actividad de C1GalT1 restringida por las células-B (Allen et al., 1997). Que *Cosmc* y C1GalT1 tengan un papel importante en la patogénesis de la nefropatía IgA es un tema polémico debido a que varios estudios sugieren que los niveles de transcripción de *Cosmc* y/o *C1GalT1* en células B son bastante bajos en pacientes con esta enfermedad (Inoue et al., 2010; Qin et al., 2005, 2008; Xie et al., 2010; Yamada et al., 2010). Un estudio

achaca (Suzuki et al., 2008) estas glicosilaciones aberrantes en células que secretan IgA1 no solo al bajo nivel de transcripción en *Cosmc* y *C1GalT1* sino también a un aumento en los niveles de transcripción de *ST6GalNAcII*. Otros enlazan esta enfermedad a polimorfismos en las propias secuencias de *Cosmc* y *C1GalT1* (Li et al., 2007; Pirulli et al., 2009), aunque otro artículo concluya que no existen mutaciones de *Cosmc* en pacientes con nefropatía IgA en países europeos (Malycha et al., 2009).

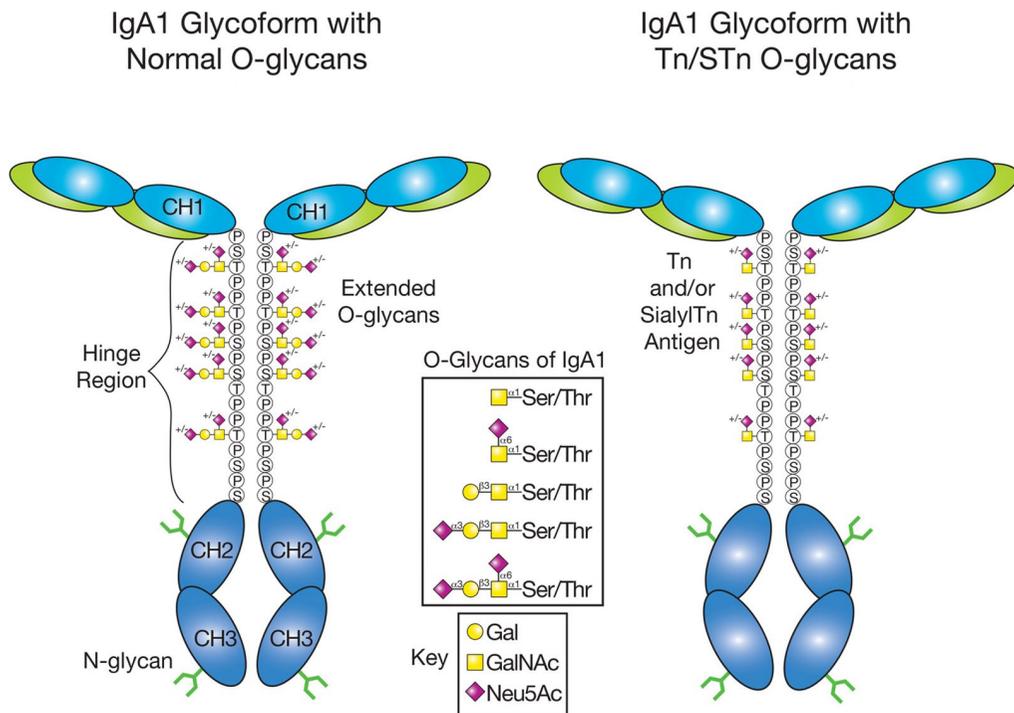


Figura 2.9. Las dos glicofomas de la inmunoglobulina A1 en el plasma humano. IgA1 es la principal glicoproteína del plasma humano y existen dos glicofomas distintas de esta proteína. La primera glicofoma es la mayoritaria (80-85%) con *O*-glicanos normales formados mayoritariamente por estructuras elongadas del *core* 1 mono y/o disialiladas. La forma minoritaria (10-15%) se encuentra glicosilada únicamente por los antígenos Tn y/o STn. Adaptado de (Lehoux et al., 2014).

Definitivamente, en 2014, Lehoux y colaboradores (Lehoux et al., 2014) describieron mediante una aproximación proteómica que en esta enfermedad hay

dos poblaciones diferentes de IgA1 (Figura 2.9), una glicosilada con *O*-glicanos elongados a partir del *core* 1, mientras que la otra presentaba exclusivamente estructuras Tn y STn; siendo la cantidad de IgA1 en pacientes con IgAN algo mayor que en los pacientes control. Mediante la adición de C1GalT1 en ensayos *in vitro* se consiguió elongar la población glicosilada con el antígeno Tn. Con todo esto, se llegó a la conclusión que la presencia de estos antígenos no es exclusiva de la nefropatía IgA, y sugieren que el mal funcionamiento *in vivo* de C1GalT1 hacia la IgA1 en subpoblaciones de células B (Zeng et al., 2021) o del plasma sanguíneo, así como un aumento en los niveles IgA1, pueden contribuir a la patogénesis de la nefropatía de IgA.

2.5.3. Cáncer y mucinas.

Ha sido observado de manera directa la presencia de proteínas glicosiladas con glicanos truncados y otras glicosilaciones aberrantes en mucinas presentes en células cancerígenas, resultando en una hipoglicosilación asociada a tumores documentada de manera extensa en la literatura (Chugh et al., 2015; Deng et al., 2013; Häuselmann & Borsig, 2014; Kufe, 2009; Taylor-Papadimitriou et al., 1999).

El principal papel de estos glicanos aberrantes, o glicanos antigénicos asociados a tumores, en la proliferación celular ha sido descrito y parece estar relacionado principalmente con la adhesión de las células cancerígenas (Häuselmann & Borsig, 2014), la motilidad de ellas y su capacidad de interaccionar de forma anómala con otras células e invadir otros tejidos, resultando todo esto en la promoción de la metástasis. De igual manera, son capaces de evadir la respuesta inmune (Madsen et al., 2013) y de mantener la renovación de células madre cancerígenas, aumentando su capacidad de expansión (Figura 2.10) (Ponnusamy et al., 2011).

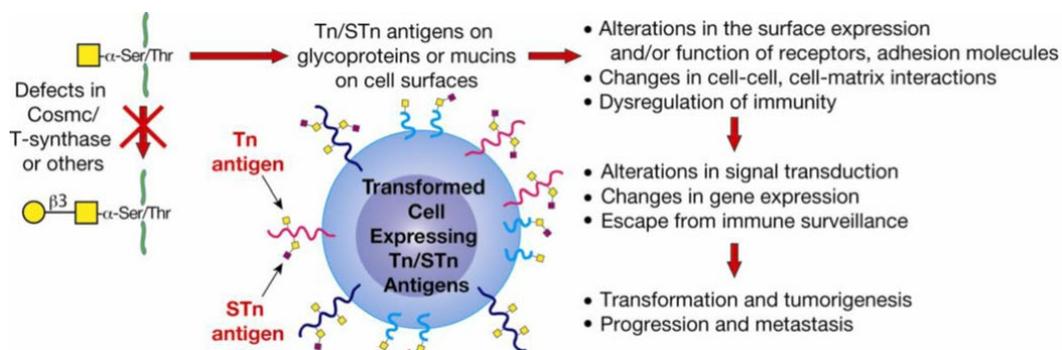


Figura 2.10. Antígenos Tn y STn como TACAs en células tumorales. Estas estructuras no se expresan en células sanas ni en tejidos no tumorales. El resultado de la sobreexpresión del antígeno Tn/STn puede deberse a defectos en *Cosmc* o en *C1GalT1*, aunque hasta la fecha parece ser que solo defectos en la transcripción de *Cosmc* o mutaciones en la misma son los principales responsables de la presencia de ellos en tumores humanos o líneas celulares cancerígenas. Esta glicosilación aberrante se encuentra en la mayoría de carcinomas y su expresión se correlaciona con la progresión de la enfermedad. Aunque el papel que juegan en la oncogénesis no se conoce exactamente, sí que se conoce que la glicosilación aberrante provoca, como se ha mencionado anteriormente, la desregulación de receptores de membrana, cambios en la señalización célula-célula o célula-matriz y/o en la regulación del sistema inmune. Es por esa razón que podrían producirse subsecuentes cambios en la expresión génica, la transducción de señales y/o las interacciones fisicoquímicas que podrían facilitar el inicio, progreso y metástasis del tumor. Fuente: (Ju et al., 2013).

Las mucinas transmembrana juegan un papel importantísimo en la adquisición de estas características mencionadas por las células cancerígenas. En células sanas se encargan de mantener la barrera de mucosa funcional y de defender las células de invasiones patógenas en las superficies mucosas (Figura 2.11a, b). De manera específica y con respecto a lo mencionado, van Putten y Strijbis (2017) describieron y respaldaron cómo la glicosilación anormal de la mucosa epitelial de colon desemboca, no solo en una pérdida de sus funciones defensivas, sino que también contribuyen al aumento de procesos inflamatorios, fomentando vías de señalización celular, que pueden acabar en la enfermedad inflamatoria del intestino (Figura 2.11c). La asociación directa de la sobreexpresión de mucinas transmembrana asociada a adenocarcinomas diversos puede derivar en la pérdida de la polarización de la membrana y la desregularización de la adhesión celular promoviendo la metástasis (Figura 2.11d) (van Putten & Strijbis, 2017).

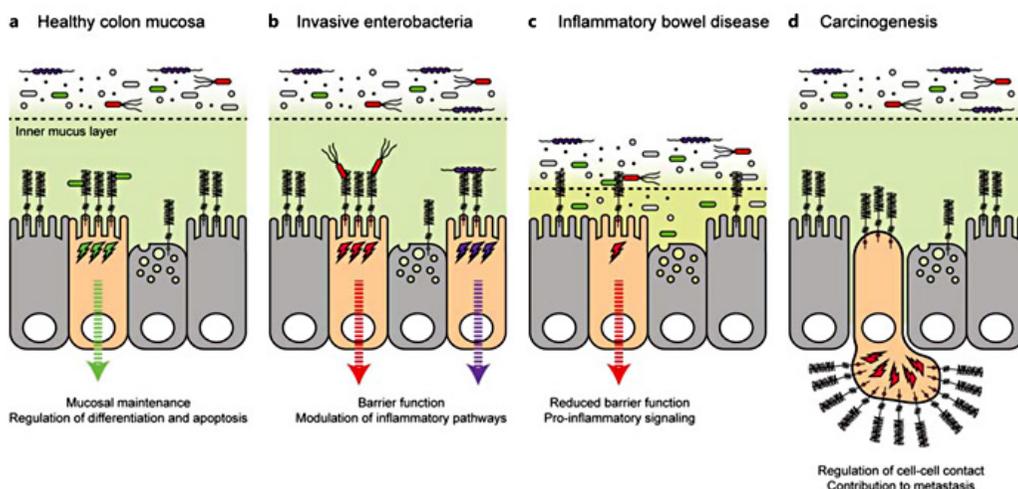


Figura 2.11. Funciones de las mucinas transmembrana en las células de colon. **a.** En condiciones en las que las células de colon se encuentran sanas, la capa de mucosa está formada por una capa interior y otra exterior, siendo la segunda la que se relaciona con la microbiota. La capa interior se considera un ambiente con condiciones estériles y capaz de contribuir a la homeostasis de la diferenciación y regulación de células epiteliales. **b.** La alta glicosilación de los dominios de las mucinas de la capa exterior son los encargados de limitar las posibles invasiones de agentes patógenos. En caso de que ciertos patógenos consigan alcanzar las mucinas transmembrana de la capa interior, se inicia una cascada de señalización a través de sus dominios citosólicos que promueve la respuesta inflamatoria. **c.** La falta de glicosilación en las mucinas transmembrana provoca una elevada inflamación celular debido a la disminución de la barrera interior, siendo los patógenos capaces de llegar a la superficie celular de manera más fácil. **d.** La sobreexpresión aberrante de mucinas transmembrana es una característica muy destacable de los adenocarcinomas tales como el de colon, de páncreas, de ovario o de mama. La sobreexpresión de estas conlleva la pérdida de la polaridad celular y se vincula directamente con una peor prognosis en los procesos tumorales. Debido a que estas mucinas regulan un elevado número de interacciones célula-células, esta sobreexpresión desemboca en una contribución elevada a la metástasis. Fuente (**van Putten & Strijbis, 2017**).

En tumores mamarios, se puede relacionar directamente la agresividad tumoral con MUC1, que aumenta la patogénesis del cáncer fomentando la activación de varias vías de señalización que desembocan en interacciones con receptores tirosinaquinasa como el EGFR y el ERB2 (receptor de estrógeno B2) y la estabilización de β -catenina y el ER α (receptor de estrógeno alfa). MUC1 puede también regular la apoptosis mediante la inhibición de vías apoptóticas como JNK y NF- κ B además de otras vías no intrínsecas (Agata et al., 2008; Bafna et al., 2010; Kufe, 2013; Wei et al., 2005). De igual forma, se ha podido relacionar la sobreexpresión de MUC1 en

patologías gastrointestinales como, por ejemplo, el cáncer de páncreas (Hinoda et al., 2003; R. Q. Wang & Fang, 2003) y en otras no relacionadas con el tracto gastrointestinal como el cáncer ovárico (Deng et al., 2013).

Existen numerosos ejemplos de mucinas implicadas en la patología tumoral de diferentes tipos de cáncer: colon, pecho, pulmón, páncreas y útero entre otros (Irimura et al., 1999; Matsukita et al., 2003). Dependiendo de la mucina existen evidencias de cómo las células de cáncer de ovario son protegidas de la respuesta de las células NK del sistema inmune por MUC16 (Gubbels et al., 2010; Thériault et al., 2011) o como MUC4 puede promover la metástasis en cáncer de pecho o de páncreas (Chaturvedi et al., 2008; Ponnusamy et al., 2011). Estos abrumadores ejemplos hacen que ahondar en el conocimiento mecanístico y estructural de las enzimas implicadas en la *O*-glicosilación sea de especial importancia para entender las vías de actuación fisiológicas y poder desarrollar tratamientos específicos que permitan controlar la actuación descontrolada y disminuir sus efectos patogénicos.

Para finalizar, parece ser que la presencia de los antígenos Tn y STn tienen enorme importancia en la capacidad de realizar interacciones carbohidrato-carbohidrato (CCI) (Sletmoen et al., 2018), postulándose como promotores de segundo nivel en la carcinogénesis a causa de su agregación y activación de receptores celulares que reconocen glicanos altamente *O*-glicosilados (Haugstad et al., 2012, 2016). Además de todo lo mencionado, C1GalT1 tiene un papel importante en todos los procesos tumorales debido a que interacciona con diferentes factores reguladores que intervienen directamente en la migración, invasión y adhesión celular. Por ejemplo, la sobreexpresión de C1GalT1 en carcinoma hepatocelular produce una modificación en la *O*-glicosilación de la Integrina $\beta 1$, la cual regula las interacciones entre las células cancerígenas y la matriz extracelular en microambientes tumorales, contribuyendo a la metástasis de las células cancerígenas. Esto implicaría que C1GalT1 influye en el comportamiento agresivo

de las células de carcinoma hepatocelular modificando la vía de señalización en la que participa la Integrina $\beta 1$ (Liu et al., 2014).

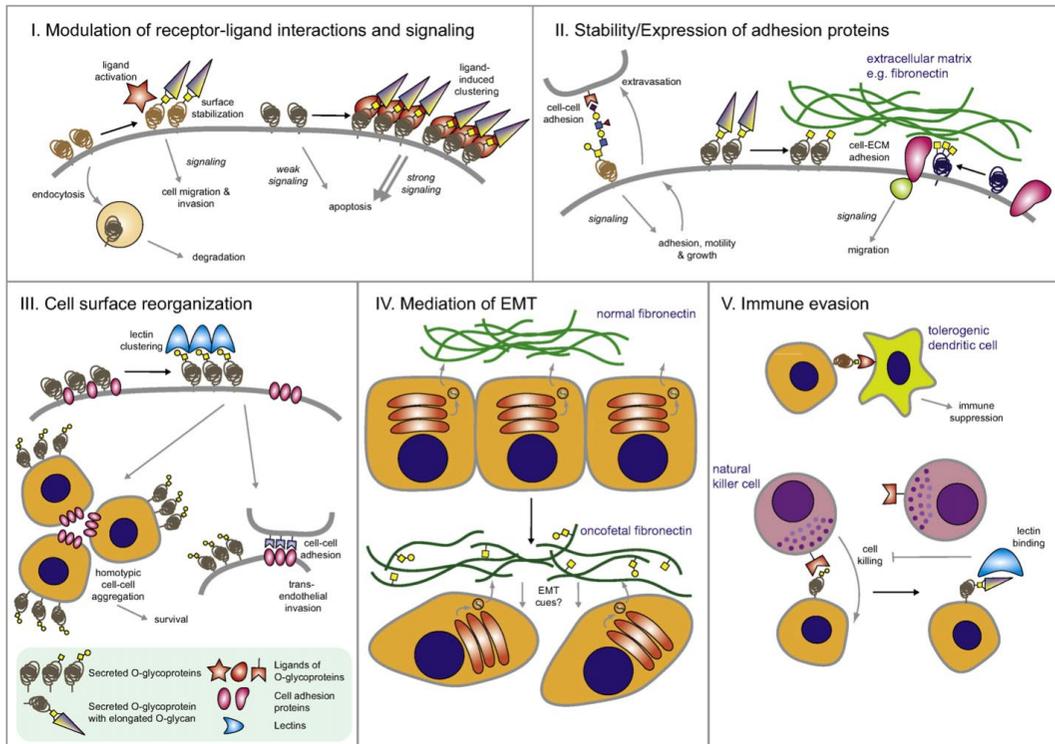


Figura 2.12. Papel de los glicanos de tipo *O*-GalNAc en la invasión de células cancerígenas. En las diferentes imágenes, las flechas grises indican vías de señalización u otros procesos celulares, mientras que las flechas negras indican un cambio anormal en la *O*-glicosilación. **I.** Estabilización de la expresión de receptores de la superficie celular mediante la reducción de la endocitosis o promoviendo la formación de complejos. **II.** Modificación de las vías de señalización relacionadas con la adhesión celular mediante cambios en la *O*-glicosilación y, en consecuencia, propiedades de determinadas proteínas esenciales en la adhesión célula-célula. **III.** Reorganización de la membrana, favoreciendo la exposición de proteínas esenciales en la adhesión célula-célula. **IV.** En los casos en los que ocurre una sobreactivación de las GalNAc-Ts, se puede inducir una transición celular del epitelio a la mesénquima, alterando la morfología celular y promoviendo la oncogénesis. **V.** Gracias a la alteración en los *O*-glicanos, puede ocurrir la evasión del sistema inmune mediante la unión de la lectina enmascarando las células frente a las NK o a través de células T tolerogénicas que reconozcan el dominio de lectina. Fuente: (Chia et al., 2016).

Como se muestra en los diferentes modelos de la figura 2.12, son muchas las implicaciones que pueden tener la existencia de glicosilaciones aberrantes en tumores, sin embargo, a nivel molecular no se saben cuáles son específicamente las dianas proteicas implicadas ni qué mecanismos regulan dicha glicosilación. Sobre

todo, la falta de información estructural sobre *Cosmc* y C1GalT1 ha impedido obtener perspectivas mecanísticas de la reacción de transferencia del residuo de galactosa al antígeno Tn o el entendimiento de las bases moleculares necesarias para el reconocimiento de la molécula de GalNAc y de los aminoácidos circundantes. A nivel fisiológico, sí que es sabido que el mal funcionamiento de C1GalT1 se debe, en la mayoría de casos, a la regulación de su chaperona específica; y en la mayoría de procesos tumorales se debe específicamente al silenciamiento de su promotor (Radhakrishnan et al., 2014).

En esta tesis, se propone un acercamiento multidisciplinar para entender las bases moleculares de una de las principales y más importantes protagonistas del proceso de la *O*-glicosilación de tipo GalNAc, C1GalT1, mediante la comprensión de la catálisis que realiza y el reconocimiento de sus sustratos donadores y aceptores. De esta manera, se podrá tener un mayor conocimiento de su implicación tanto en las enfermedades mencionadas como en cáncer.

3. OBJETIVOS

El objetivo principal de esta tesis doctoral es comprender las bases moleculares del funcionamiento de C1GalT1. Para ello, se fijó un objetivo principal que comenzó con la expresión de esta enzima de diferentes especies y que terminó con comprender cómo C1GalT1 reconoce y glicosila proteínas previamente glicosiladas por las GalNAc-Ts. Es por ello que los objetivos tratados en esta tesis son los siguientes:

- Elección de ortólogos y diseño de vectores de las construcciones de C1GalT1 para su expresión recombinante.
- Expresión recombinante de C1GalT1 en sistemas de expresión eucariotas: expresión transitoria en células de mamíferos y expresión mediante el sistema de expresión de baculovirus en células de insecto.
- Purificación de las proteínas recombinantes expresadas.
- Diseño racional y síntesis de glicopéptidos específicos para C1GalT1.
- Caracterización cinética frente a glicopéptidos específicos.
- Caracterización de las interacciones proteína ligando mediante ITC y STD-RMN.
- Cristalización de C1GalT1 en complejo con glicopéptidos y UDP-Mn⁺² para entender las bases moleculares de como reconoce a sus ligandos
- Compresión y estudio de los aminoácidos involucrados en la interacción con los péptidos o en su catálisis mediante mutagénesis dirigida.
- Bases moleculares de cómo C1GalT1 reconoce a α -GalNAc-Thr frente a α -GalNAc-Ser
- Simulaciones de estructura tridimensional mediante dinámicas moleculares.

4. METODOLOGÍA

En esta tesis doctoral, el trabajo realizado y desarrollado tiene una aproximación multidisciplinar donde se ha integrado un elevado número de técnicas experimentales y computacionales con el fin de obtener respuestas con respecto a las bases estructurales y moleculares de C1GalT1. Debido a esta heterogeneidad, ha sido necesaria la implicación de diferentes grupos de investigación, sin los cuales no hubiese sido posible realizar todos los experimentos detallados en este capítulo. Es por ello de especial importancia remarcar cuáles son las técnicas que no se han realizado en el grupo de investigación al que pertenece el doctorado. De la misma forma se realizará una pequeña introducción sobre estas técnicas realizadas por otros grupos de investigación para remarcar la importancia dentro del trabajo completo, añadiendo detalles sobre los métodos usados.

4.1. Cultivos celulares.

Para la expresión de proteínas recombinantes se han usado líneas celulares eucariotas para conseguir un correcto plegamiento de las proteínas expresadas debido a la existencia de modificaciones postraduccionales los puentes disulfuro. La obtención del ADN y generación de báculos recombinantes se realizó en dos cepas diferentes de *E. coli*.

4.1.1. Cultivos celulares de bacterias: *Escherichia coli* DH5 α y DH10Bac.

Las cepas de bacterias se usaron para la amplificación, obtención y preparación del ADN para su uso en la expresión de proteínas en cultivos de mamíferos e insectos. Dependiendo del sistema de expresión se usó una cepa diferente de *E. coli*:

- La cepa DH5 α se usó para la amplificación de ADN plasmídico para su posterior transfección en células de mamíferos HEK293F.
- La cepa DH10Bac se usó para la obtención del báculo recombinante para la creación de baculovirus recombinantes mediante la transfección en

células de insectos Sf9.

De forma general, las condiciones de cultivo de *E. coli* son 37°C y, si el cultivo es en suspensión, una agitación entre 180 y 220 rpm dependiendo del volumen de cultivo. Para el crecimiento de estas dos cepas de bacterias se usa medio LB.

4.1.2. Cultivos celulares de mamíferos: HEK293F.

Según la Colección Americana de Cultivos Tipo (*American Type Culture Collection*), las células HEK293 (CRL-1573™) son células embrionarias de riñón humano (*Human Kidney Embryo* en inglés) con morfología epitelial usadas globalmente en multitud de ensayos de toxicidad y de biotecnología industrial (<https://www.atcc.org/products/crl-1573>). Existen dos tipos principales de cultivos de células de organismos superiores: en adhesión y en suspensión, y no todas las líneas celulares están adaptadas para crecer en suspensión, teniendo como beneficio principal un aumento de la densidad celular de los cultivos y una mayor facilidad de escalado con respecto a los cultivos en adhesión.

Todas las variantes glicomodificadas de HEK293, que se usaron para comprobar la actividad en células de los mutantes de C1GalT1, se cultivaron en matraces para adhesión celular (*T-Flasks*) en medio DMEM (Sigma-Aldrich) con 10% de suero fetal bovino (FBS) inactivado (Sigma-Aldrich) y 2 mM de GlutaMAX (Gibco) en un incubador humidificado a 37°C con 5% de CO₂.

Las células usadas para cultivos en suspensión y expresión de proteínas fueron las HEK293F (ThermoFisher), una variante comercial de las HEK293 modificadas y adaptadas para el crecimiento en suspensión de alta densidad celular sin suero, permitiendo de esa manera una mayor expresión transitoria de proteínas. Estas células se cultivan en suspensión, en matraces especiales cuyo tapón dispone de una membrana para intercambio continuo de gas (Corning), en medio FreeStyle F17 (Thermo Fisher) suplementado con 2% de GlutaMAX y 0,1% de Kolliphor P188

(Sigma Aldrich) para evitar agregados celulares. Los cultivos se mantienen en agitación constante (125 rpm) en un incubador a 37°C con 8% de CO₂ y 80% de humedad. Las células se mantuvieron siempre a una densidad celular entre 0,4 x 10⁶ y 4 x 10⁶ células/ml.

4.1.3. Cultivos celulares de insectos: Sf9 y HighFive™.

Tanto las células Sf9 como las HighFive™ (Gibco) se cultivaron en suspensión en medio libre de proteínas Insect XPRESS (LONZA) a 27°C y 130 rpm. La densidad celular se mantuvo entre 0,4 x 10⁶ y 4 x 10⁶ células/ml. Las células Sf9 provienen de células aisladas del ovario de pupas de *Spodoptera frugiperda* (Figura 4.1a) (comúnmente conocida como “Oruga Militar Tardía”). De igual manera, las HighFive™ (Hi5 para abreviar) provienen de células aisladas del ovario de *Trichoplusia ni* (Fig. 4.1b) (o “Gusano Falso Medidor”). El uso de dos líneas celulares diferentes se debe a que la eficiencia de transfección en Sf9 es mucho mayor, por lo que son mucho más adecuadas para la generación de baculovirus recombinantes, mientras que se ha demostrado que las Hi5 son capaces de alcanzar niveles más altos de expresión de proteínas recombinantes (Wilde et al., 2014).

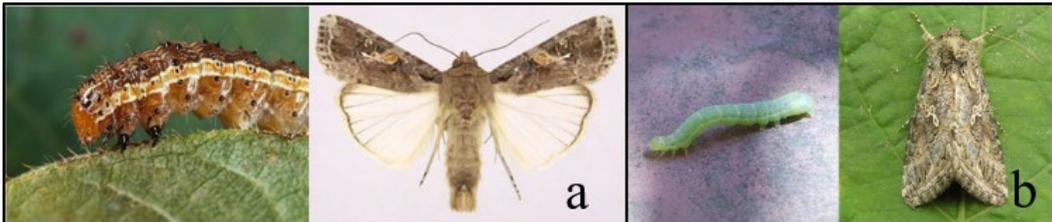


Figura 4.1. Estados larvarios y adultos de *Spodoptera frugiperda* (a) y *Trichoplusia ni* (b).

una secuencia codificante para una etiqueta de 6 residuos de histidina en el extremo 3' seguida de un codón de parada de traducción. Mediante esta clonación realizada también por GenScript se obtuvo el vector pHL-Sec-DmC1GalT1-6xHis.

Para los ensayos en células y *knockins* dirigidos a AAVS1, Genewiz USA realizó la síntesis de la secuencia humana (*Homo sapiens*) completa de C1GalT1, optimizando la frecuencia de codones, con una etiqueta Myc en el carboxilo terminal y realizó el subclonaje en el vector EPB71 (Fig. 4.3), obteniendo el vector EPB71-HsC1GalT1-Myc. (Addgene ID 90018).



Figura 4.3. Mapa del vector EPB71. Usado por Genewiz como base para la creación de los diferentes plásmidos usados en la creación de *knockins*. Fuente: <https://www.addgene.org/90018/>.

4.2.1.2. Plásmidos para expresión en células de insectos.

La secuencia de ADN codificante para los residuos T43 al Q388 junto con la

secuencia de secreción “*mellitin honey bee*” (Tessier et al., 1991), al igual que en el caso de la secuencia codificante para expresión en mamíferos, fue sintetizada y optimizada su frecuencia de codones para la expresión en células de insecto por GenScript. El constructo contenía una secuencia de reconocimiento para BamHI en el extremo 5' y en el extremo 3' una secuencia codificante para 6 residuos de histidina, un codón de parada y una secuencia de reconocimiento para EcoRI.

Mediante esos sitios de reconocimiento de enzimas de restricción, la secuencia codificante fue clonada en el vector pFastBac1 (Fig. 4.4) obteniendo el vector pFastBac1-mellitin-DmC1GalT1(T43-Q388)-6xHis. Se sintetizaron mediante



Figura 4.4. Mapa del vector pFastBac™1. Usado por GenScript como base para la creación de los diferentes plásmidos usados para la creación de los bécmidos y baculovirus recombinantes para la expresión de proteínas en células de insecto.

mutagénesis del vector arriba mencionado 5 construcciones adicionales de *DmC1GalT1*: H55-Q388, S73-Q388, T43-P343, T43-P354 y T43-S373.

4.2.1.3. Mutagénesis de sitio dirigido de los plásmidos para expresión en mamíferos.

Todos los mutantes de *DmC1GalT1* (R152A, Y213A, Y218A, D255A, W300A y Y325A) fueron generados mediante un protocolo estándar de mutagénesis de sitio dirigido por GenScript usando como base el vector pHL-Sec-*DmC1GalT1*-6xHis.

De igual forma, GenScript realizó la mutagénesis de los residuos indicados usando el vector EPB71-*HsC1GalT1*-Myc.

4.2.2. Sistema dirigido de CRISPR/Cas9 para hacer *knockouts* en HEK293F.

El sistema CRISPR/Cas es el proceso adaptativo del sistema inmune de muchas bacterias y archaeas que les permite adquirir inmunidad frente a plásmidos y virus exógenos (Horvath & Barrangou, 2010). El estudio y modificación de este sistema ha permitido usarlo como herramienta de edición del genoma denominada CRISPR/Cas9. La base sobre la que funciona este sistema es el uso de un fragmento de ARN “guía” (gARN) y la endonucleasa Cas9, que se une a un ADN objetivo y corta en los sitios especificados por el gARN (Cui et al., 2018) (Fig. 4.5).

El gARN está formado principalmente por dos componentes: una secuencia constante a la que se une Cas9 y una secuencia variable que se une, siendo complementaria, al ADN objetivo (Cong & Zhang, 2015). El ADN objetivo, por otro lado, debe tener dos regiones para que Cas9 pueda realizar su función endonucleasa y cortar la doble cadena de ADN. La primera es la región variable complementaria al gARN que se decide usar (o “*protospacer*”) y una segunda región o PAM (“*protospacer adjacent motif*”). Una vez Cas9 se une al *protospacer* y reconoce el motivo PAM, realizando el corte de la doble hebra, puede dar lugar dos escenarios distintos. El primero es la reparación del ADN por la maquinaria

celular endógena, produciendo la eliminación de la secuencia seleccionada. El segundo escenario sería la reparación del corte por la maquinaria con la adición de una secuencia exógena, ocupando el hueco creado por el corte (Wyman & Kanaar, 2006).

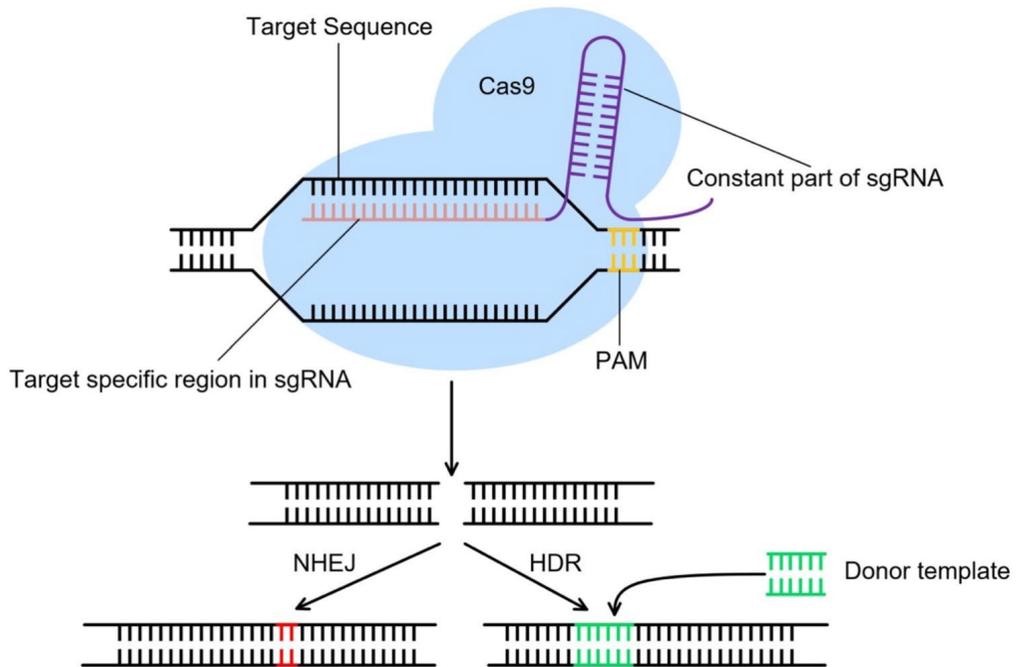


Figura 4.5. Sistema de edición genómica CRISPR/Cas9. Fuente: (Cui et al., 2018).

Tras explicar de manera breve el funcionamiento de CRISPR/Cas9, se detallará el procedimiento seguido por el Dr. Yang Zhang de la Universidad de Copenhague.

Los *knockouts* mediante CRISPR/Cas9 se realizaron usando la GlycoCRISPR, una librería de gARNs para la selección de todas las glicosiltransferasas humanas (Narimatsu et al., 2018). Se usaron células previamente modificadas capaces de producir *O*-glicanos hasta el *core* 1, denominadas HEK293^{core1} (KO *GCNT1/ST3GAL1/2/ST6GALNAC2/3/4*). Estas células expresan de manera estable GFP-MUC1 al exterior celular, pudiéndose usar esta proteína como marcador para comprobar la *O*-glicosilación proteica o proteína reportera, denominada en inglés *reporter* (Nason et al., 2021). Las HEK293^{core1} fueron crecidas en placas de seis

pocillos (NUNC) hasta llegar aproximadamente a un 70% de confluencia. En ese momento fueron transfectadas con 1 μ g de gARN, capaz de reconocer el gen *C1GALT1*, y 1 μ g de Cas9-PBKS etiquetada con RFP (proteína roja fluorescente) usando lipofectamina 3000 (ThermoFisher) como agente de transfección siguiendo el protocolo del fabricante. 24 h después de la transfección, las células con expresión de RFP fueron seleccionadas mediante un tipo de citometría de flujo denominada clasificación celular activada por fluorescencia (FACS sorter – SONY SH800) (Adan et al., 2017). Tras una semana de cultivo, las células seleccionadas previamente fueron clasificadas de manera aislada en placas de 96 pocillos. Los clones sin el gen *C1GALT1* se cribaron mediante PCR usando la detección Indel por análisis de amplicones (IDAA) (Yang, Steentoft, et al., 2015), amplificando el sitio de unión del gARN. Los cebadores flanqueantes usados para dicha amplificación fueron:

- *Forward*: 5'-CCTGCTGTGGGACTGAAAAC-3'.
- *Reverse*: 5'-TGCATCTCCCCAGTGCTAAG-3'.

Posteriormente los clones seleccionados fueron verificados mediante secuenciación de Sanger (Shendure et al., 2017).

4.2.3. Creación de *knockins* de variantes de C1GalT1 en HEK293F^{Tn}.

Para la creación de *knockins* se usó un ObLiGaRe (“*Obligated Ligation-Gated Recombination*” o recombinación controlada por ligación obligada en castellano) (Maresca et al., 2013) modificado para dirigirse al sitio de integración 1 asociado a adenovirus (AAVS1) usando dos nucleasas de dedos de Zinc invertidas (ZFNs) que flanqueasen el gen de *C1GALT1* mutado en los plásmidos donadores (Yang, Wang, et al., 2015). Este protocolo se realizó de la manera descrita con anterioridad para *knockouts* dirigidos usando 1 μ g de ZFN etiquetada con GFP/E2-Crimson y 2 μ g del plásmido donador. 48 horas tras la transfección se realizó una selección del 10-

15 % de células que más expresión presentaban tanto de GFP como de E2-Crimson mediante FACS y se cultivaron durante una semana más. En ese momento, los *knockins* dirigidos se cribaron de manera individual mediante PCR usando dos cebadores específicos para el área de unión entre el plásmido donador y el locus humano de AAVS1, y otros dos cebadores que flanqueasen el locus del gen introducido. Se realizó también una PCR específica para comprobar el número de copias del gen insertado:

- *Forward*: 5'- CCTTACCTCTCTAGTCTGTGCTAG-3'.
- *Reverse*: 5'- CGTAAGCAAACCTTAGAGGTTCTGG-3'.

4.2.4. Transformación de células competentes *E. coli* DH5 α y DH10Bac y selección de colonias recombinantes.

La transformación de células competentes consiste en la introducción en bacterias silvestres de un plásmido exógeno que contiene el gen de interés. Las células competentes se tratan de bacterias tratadas químicamente para que el ADN plasmídico pueda introducirse de manera más eficiente. Dependiendo del fin de la transformación, el plásmido tendrá unas características diferentes. En estos ensayos, todas las transformaciones bacterianas realizadas tenían como fin la amplificación de ADN (DH5 α) o la generación del b α cmido recombinante (DH10Bac), para expresar las proteínas en células de mamíferos o de insectos, respectivamente. El protocolo de transformación es casi idéntico en ambos casos, por lo que se enunciarán las diferencias cuando sea necesario. Todos los pasos se realizaron en esterilidad.

1. Adición de 200 ng de ADN plasmídico sobre una alícuota de 250 μ l células competentes e incubar la mezcla en hielo durante 30 minutos.
2. Realización de choque térmico a 42°C durante 1 minuto.
3. Incubación en hielo durante 2 minutos.
4. Adición de 500 μ l de SOC frío (Apéndice 8.3).

5. Incubación a 37°C y agitación de 180-220 rpm durante 1,5 horas en el caso de las células DH5 α o 4 horas en el caso de las células DH10Bac.

Transcurrido ese tiempo, se recuperan las células y se siembran en placas de LB/Agar (Apéndice 8.3) con los antibióticos correspondientes para su correcta selección. En el caso de las DH5 α solo es necesario añadir ampicilina 100 $\mu\text{g/ml}$ como antibiótico para comprobar el éxito de la transformación. Las placas se incuban durante la noche (O/N) a 37°C.

Las células DH10Bac contienen en su cromosoma el genoma de baculovirus (bácmido). La transposición entre el vector insertado y el bácmido ocurre a través del método de transposición Tn7, generando el bácmido recombinante con nuestro gen de interés. Para confirmar el correcto inserto del gen de interés en el bácmido las células se crecieron durante al menos 48 horas en placas de LB/Agar con kanamicina 50 $\mu\text{g/ml}$, gentamicina 7 $\mu\text{g/mL}$, tetraciclina 10 $\mu\text{g/mL}$, 100 $\mu\text{g/mL}$ Bluo-gal y 40 $\mu\text{g/mL}$ IPTG. Los clones positivos fueron aquellas colonias que presentaban color blanco gracias a la disrupción del gen *lacZ* integrado en el sitio de transposición. Las colonias blancas se seleccionaron y volvieron a crecer en placas como las anteriores durante al menos 48 horas para confirmar el inserto de nuestro gen de interés.

4.2.5. Conservación en glicerol de las colonias transformadas.

Para evitar transformar cada vez que sea necesario, las células transformadas se almacenan a -80°C. Para almacenarlas así, es necesario seleccionar una colonia aislada y crecerla O/N en 10 ml de cultivo LB (Apéndice 8.3) a 37°C con agitación (180-220 rpm), usando los antibióticos que hemos usados en las placas de LB/Agar (en el caso de las DH10Bac no es necesario usar IPTG ni Bluo-Gal ya que la colonia ha sido validada previamente).

Tras el crecimiento del cultivo O/N, manteniendo en todo momento la esterilidad,

se añaden en un vial estéril de 1,5 ml: 300 µl de glicerol estéril al 80% y 700 µl del cultivo crecido. Tras etiquetarlo apropiadamente, el vial se introduce en nitrógeno líquido para una rápida congelación y se almacena a -80°C.

4.2.6. El sistema Bac-to-Bac.

Para producir los baculovirus recombinantes de manera rápida y eficiente, se usó el sistema de expresión Bac-to-Bac® (Invitrogen) (Fig. 4.6). Este método fue desarrollado por investigadores de Monsanto y se basa en la transposición de sitio específica de un casete de expresión en un vector lanzadera de baculovirus (bácmido) que se propaga en *E. coli* (Luckow et al., 1993). Este sistema usa las propiedades del transposón Tn7 para simplificar y mejorar la generación del ADN bacmídico recombinante. Este sistema tiene dos componentes principales:

- El vector donador con el gen (o genes) de interés clonado, denominado pFastBac. Dependiendo del vector usado, la expresión del gen está controlada por un promotor u otro. En el presente caso, se usó el pFastBac1, que contiene el promotor de polihedrina (PH) del virus de polihedrosis multinuclear de *Autographa californica* (AcMNPV – “*Autographa californica* multiple nuclear polyhedrosis virus”). Este casete de expresión se encuentra flanqueado por los brazos izquierdo y derecho de Tn7 y además contiene un gen de resistencia a gentamicina y una señal de poliadenilación SV40 para formar un mini – Tn7.
- El segundo componente del sistema se trata del huésped para el vector pFastBac, la cepa de *E. coli* DH10Bac. Esta cepa contiene:
 - Un “vector lanzadera de baculovirus” o bácmido (bMON14272; 136 kb) con un replicón de bajo número de copias mini – F, un gen de resistencia a kanamicina y la secuencia codificante del péptido LacZα en la que se ha insertado el sitio de unión del transposón bacteriano (mini – attTn7). La inserción del sitio de transposición

- mini – attTn7 no interrumpe el marco de lectura del péptido LacZ α .
- Un plásmido ayudante (o “*helper*”) (pMON7124; 13.2 kb) encargado de codificar la transposasa, otorgando la función de transposición de Tn7 *in trans* y confiriendo resistencia a tetraciclina (Barry, 1988).

Una vez se realiza la transformación del plásmido pFastBac1 en las células DH10Bac, ocurre la transposición entre el elemento mini – Tn7 del vector pFastBac1 y el mini – attTn7 objetivo del báculo, generando el báculo recombinante gracias a la presencia de proteínas de transposición codificadas en el plásmido ayudante.

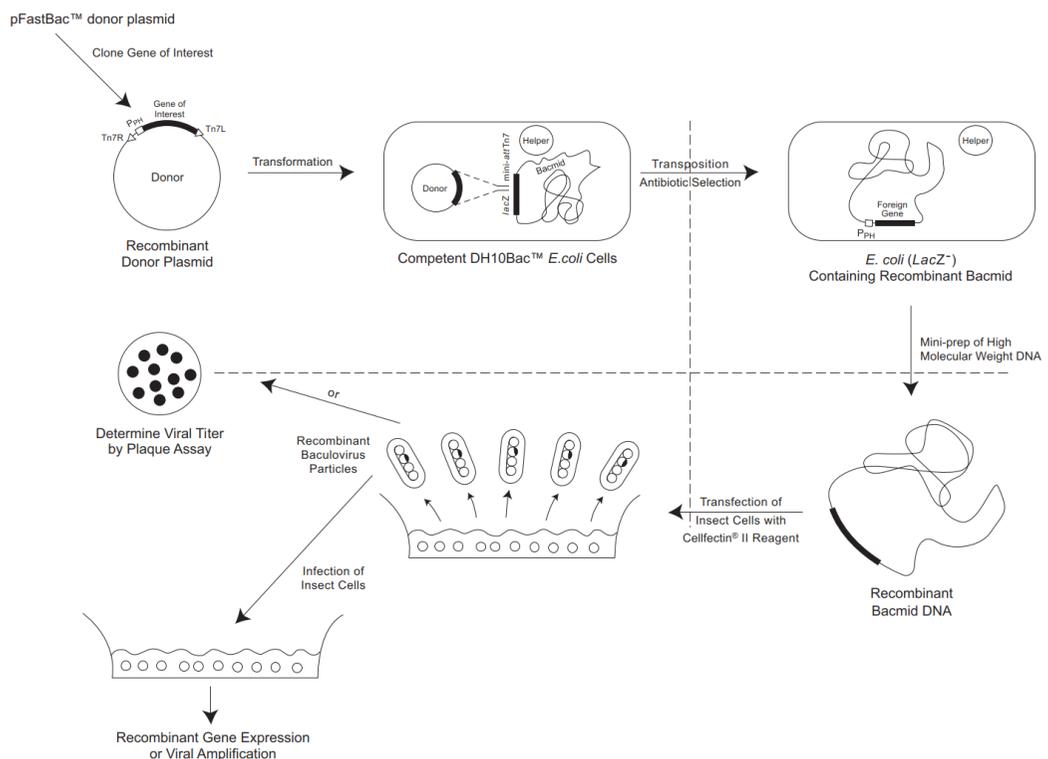


Figura 4.6. Esquema simplificado del sistema de expresión de baculovirus. En la imagen se muestra desde la generación del báculo recombinante hasta la generación y amplificación de virus recombinantes. Fuente: (Invitrogen, 2015).

El báculo recombinante es capaz de propagarse en *E. coli* DH10Bac, confiriendo

la resistencia a kanamicina y produciendo la disrupción del marco abierto de lectura del gen codificante de *lacZ*, presente en el cromosoma para formar colonias azules (Lac^+) en presencia de un sustrato cromogénico como Bluo-gal o X-gal y el inductor, IPTG. Las colonias que contengan el b́acmido recombinante presentan un color blanco mientras que las que presentan el b́acmido sin alterar seŕan de color azul debido al sustrato cromogénico. Las colonias blancas tienden a ser algo ḿas grandes, por lo que para evitar falsos positivos es necesario elegir las colonias ḿas grandes y aisladas para analizarlas en las placas de LB/Agar preparadas para su selecci3n.

4.2.7. Amplificaci3n y extracci3n de ADN plasmídico.

Para la transfecci3n en mamíferos es necesaria una gran cantidad de ADN plasmídico (3 μg ADN/ml c3lulas) como se ha mencionado anteriormente. Es por ello que es necesaria la amplificaci3n y extracci3n del ADN plasmídico, que se lleva a cabo en primer lugar creciendo c3lulas DH5 α transformadas con el plásmido objetivo en un matraz de 250 ml con 100 ml de LB y ampicilina O/N a 37°C y agitaci3n (180-220 rpm). Al d́a siguiente, los cultivos crecidos se inoculan en matraces de 5 l con 2 l de LB y ampicilina y se dejan creciendo hasta el d́a siguiente a 37°C y agitaci3n de 180 rpm. El tercer d́a se recoge el pellet bacteriano mediante centrifugaci3n a 10000 rpm a 4°C durante 10 min y se congela.

Una vez se tiene el pellet congelado, se procede a realizar el aislamiento del ADN plasmídico mediante un kit de gigaprep Purelink Expi Endotoxin-Free Giga Plasmid Purification Kit (Invitrogen), siguiendo las instrucciones del fabricante (Fig 4.7).

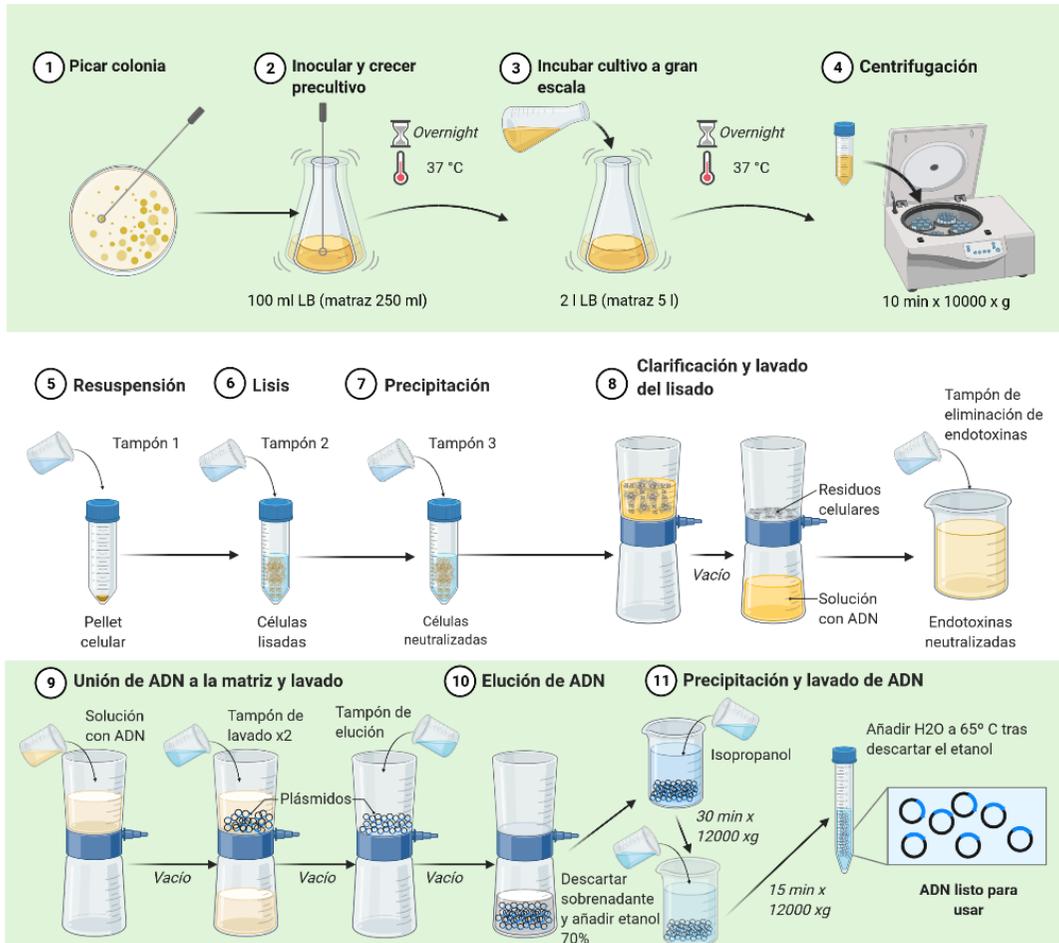


Figura 4.7. Protocolo de aislamiento de ADN plasmídico. Mediante el uso del kit Purelink Expi Endotoxin-Free Giga Plasmid Purification Kit (Invitrogen). Realizado por el doctorando mediante Biorender.

El ADN obtenido de la purificación se cuantificó mediante absorción a 260 nm mediante un nanodrop espectrofotómetro/fluorímetro DS-11 Series (DeNovix) para obtener la concentración a usar en la transfección.

4.2.8. Amplificación y extracción del cromosoma bacteriano recombinante (bácmido).

Una vez se ha realizado la reacción de transposición, es necesario aislar el ADN recombinante de alto peso molecular (bácmido) para usarlo en la transfección de células de insecto y poder generar los baculovirus recombinantes.

Para ello, se crecieron DH10Bac transformadas, y reestriadas para comprobar que eran clones recombinantes positivos, en 50 mL de LB con kanamicina, gentamicina y tetraciclina *overnight* a 37°C con agitación (180-220 rpm). Los siguientes pasos fueron exactamente igual que para realizar la amplificación de ADN plasmídico, pero inoculando cultivos de 500 ml de LB con los antibióticos usados en el precultivo. El aislamiento del báculo recombinante se llevó a cabo usando el kit de extracción NucleoBond BAC 100 (MACHEREY-NAGEL) siguiendo las instrucciones del fabricante. Como aclaración, para poder extraer correctamente un ADN de tan gran tamaño es crucial dejar que la clarificación del lisado ocurra por gravedad y que tanto el isopropanol como el etanol usado sean de una buena calidad. La cuantificación del ADN extraído se realizó de igual forma que la del ADN plasmídico.

4.3. Expresión de proteínas.

4.3.1. Expresión transitoria de *DmC1GalT1* y sus mutantes en HEK293F.

La expresión transitoria de C1GalT1 y de sus mutantes en células de mamíferos se llevó a cabo mediante un protocolo de transfección optimizado en el grupo de investigación.

El protocolo de transfección se indica a continuación:

1. Crecer el volumen deseado de células a una densidad de $2-3 \times 10^6$ células/ml.
2. Centrifugar los cultivos a 200 x g durante 10 minutos, desechar el medio viejo y resuspender las células en medio fresco FreeStyle F17 (ThermoFisher) suplementado (2% GlutaMAX y 0,1% Kolliphor P188).
3. Añadir al cultivo 3 µg de ADN/ml de cultivo celular mezclado con 0,9 µl de NaCl 1,5 M/ml de cultivo celular e incubar durante 5 min en el incubador.
4. Tras la incubación, es necesario añadir 9 µg/ml de cultivo celular de PEI-MAX 40 K (Polietilenimina) mezclado con 0,9 µl de NaCl 1,5 M/ml de

cultivo celular, obteniendo una concentración final de NaCl de 3,75 mM en el cultivo.

5. Duplicar el volumen de cultivo a las 24 horas con medio suplementado. Además, se añade ácido valproico (VPA) (Sigma Aldrich) a una concentración final de 2,2 mM.
6. Los cultivos se recogen 6 días tras la transfección realizando dos pasos de centrifugación: 5 minutos a 300 g y 15 minutos a 4000 g para eliminar todas las células, obteniendo así el sobrenadante con la proteína en cuestión secretada.

4.3.2. Producción en Sf9 de baculovirus recombinantes que expresen DmC1GalT1.

Para la generación de baculovirus recombinantes fue necesario realizar primero una transfección del bÁcmido recombinante obtenido con anterioridad (P0) y luego realizar dos amplificaciones para aumentar la cantidad de virus (P1 y P2). Para la obtención del P0 se realizó la transfección del bÁcmido recombinante en células Sf9 con bajo número de pases. Las células se diluyeron en medio fresco Insect XPRESS (LONZA) a una densidad de $0,8 \times 10^6$ células/ml 3-4 horas antes de la transfección. Se usó 1 µg de bÁcmido/ml de células, diluido en 100 µl de PBS preatemperado a 37°C, y se mezcló vigorosamente con PEI-MAX 40K a un *ratio* de 1:4 ADN:PEI (4 µg PEI/ml de células). La formación del complejo entre el ADN y PEI se permitió durante 20-30 minutos a temperatura ambiente y tras ello se añadió la mezcla al cultivo celular (Scholz & Suppmann, 2017). El cultivo transfectado se incubó durante 7 días a 27°C con agitación de 130 rpm y luego se recogió mediante centrifugación de las células durante 10 minutos a 4000 g. El sobrenadante resultante que contenía el P0 se almacenó en oscuridad a 4°C tras añadir 10% de FBS inactivado.

Para la amplificación tanto del P1 como del P2 se diluyeron células Sf9 a $1,5 \times 10^6$

células/ml en medio fresco Insect XPRESS antes de añadir 0,25% del virus anterior (P0 para el P1, P1 para el P2), aumentando así la cantidad de virus en el sobrenadante y en consecuencia su capacidad infectiva. Para el P1, los cultivos fueron incubados durante 7 días mientras que para el P2 la incubación fue de 5 días. La incubación, recogida y almacenaje de los P1 y P2 se llevó a cabo de la misma manera que se ha descrito en el caso del P0.

4.3.3. Expresión de *DmC1GalT1* en Hi5™.

Las células Hi5 se diluyeron a $1,5 \times 10^6$ células/ml en medio Insect XPRESS. En el momento de la infección se añadió al cultivo Kifunensina-Bio-X (CarboSynth) (usando una concentración final de 5 μ M para permitir el corte de los N-glicanos durante la purificación (Harrison & Jarvis, 2006) seguido de un 3% del P2 de baculovirus recombinante. El sobrenadante celular se recogió a las 48 horas tras la infección mediante centrifugación a 300 x g durante 5 min seguida de una segunda centrifugación a 8000 x g durante 10 min.

4.3.4. Electroforesis SDS-PAGE desnaturalizante y tinción.

Para poder comprobar tanto la expresión como la pureza o estado de las proteínas es necesario el uso de una técnica sencilla y rápida con la que poder comprobar cómo se encuentran. La electroforesis desnaturalizante es una técnica con las características mencionadas en la que se realiza la separación en geles de poliacrilamida de muestras proteicas, cargadas negativamente, en función del peso molecular mientras se aplique una diferencia de potencial eléctrico, permitiendo su posterior visualización mediante un proceso de tinción (Brunelle & Green, 2014).

En primer lugar, es necesario desnaturalizar térmicamente las muestras durante 5 minutos a 100°C en presencia de un tampón de carga Laemmli (Apéndice 8.3) que contiene entre sus componentes dodecil sulfato sódico (SDS), un detergente capaz de unirse a las proteínas inespecíficamente confiriéndoles una relación

tamaño/carga constante.

Dependiendo del rango de tamaño de proteínas que necesitemos visualizar y separar usaremos un diferente porcentaje de acrilamida en la receta del gel de poliacrilamida. Además, el gel consta generalmente de dos fases, una fase de apilado o “*stacking*”, con un bajo porcentaje de acrilamida (4%), en la que las proteínas se “apilan” al inicio de la segunda fase, llamada “*running*” o de separado en la que las proteínas “corren” y se separan dependiendo del porcentaje de acrilamida en esta fase (8-20%).

Con las muestras ya cargadas en el gel, se aplica con una fuente de alimentación un campo eléctrico con voltaje constante de 130 V durante 80 – 100 minutos en función del porcentaje de acrilamida. De esta manera, las muestras cargadas migran por el gel a una velocidad proporcionalmente inversa al logaritmo de su peso molecular, pudiéndose comparar así con un patrón estándar de peso molecular comercial (marcador de peso molecular, en nuestro caso Page Ruler Protein Ladder, (Thermo Fisher)).

Se han empleado geles de acrilamida convencionales con un 10 – 12% de la misma para comprobar las diferentes purificaciones o cortes con glicosidasas realizadas en la tesis. Para poder visualizar la separación y el tamaño de las muestras proteicas es necesario realizar una tinción con Azul de Coomassie y una destinción, tras la primera, con una disolución de metanol y ácido acético. Una vez se realiza la destinción, se puede observar a simple vista las bandas correspondientes a las proteínas cargadas en el gel.

4.3.5. Electrotransferencia (*Western Blot*).

Hay veces en las que la electroforesis desnaturizante no es suficiente para poder visualizar la expresión de proteínas debido a que están muy diluidas (sobre todo en sistemas de expresión de secreción) o que la muestra se encuentra con muchas

impurezas. Por esa razón, hay que hacer uso de otras técnicas más sensibles como es la electrotransferencia o “Western Blot”, con la que podemos visualizar proteínas gracias a la unión de anticuerpos específicos mediante quimioluminiscencia.

Debido a que todas las construcciones que suelen usarse en el grupo de investigación suelen llevar una cola de histidinas para facilitar su purificación, es lógico pensar en usar esa misma etiqueta para detectar la expresión de proteínas. Es necesario remarcar que esta técnica es cualitativa y solo permite cuantificar la expresión siempre que haya una muestra de expresión conocida en el ensayo.

El primer paso para realizar esta técnica es realizar una electroforesis desnaturizante y una vez el gel con las muestras ha terminado de correr, se empieza la electrotransferencia. La transferencia electroforética es el método más usado en la realización de western blots, en el que un campo eléctrico se usa para eluir las proteínas del gel y transferirlas a las membranas – las cuales pueden ser de nitrocelulosa o de PVDF (difluoruro de polivinilo). Durante este proceso, tanto el gel como la membrana se encuentran juntos con papel de filtro húmedo entre dos electrodos. El voltaje aplicado entre electrodos hace que las proteínas migren a la membrana siguiendo la corriente generada por dicho voltaje entre electrodos. En el centro de investigación se usó un aparato de electrotransferencia semihúmeda Trans-Blot Turbo – Transfer System (BIO-RAD). Este método se categoriza dentro de los sistemas de rápida transferencia y el esquema de montaje se ilustra en la figura 4.8. En las electrotransferencias realizadas se usaron membranas PVDF y los papeles de filtro y tampón indicados por el fabricante.

Una vez se ha completado la migración de las proteínas a la membrana (ha sido exitosa si el marcador de peso molecular se ha transferido) se realiza la saturación de la membrana con 5 ml de una solución de BSA 0,05 g/ml en PBST al 0,1 % (Tween 20) durante 1 hora en agitación a temperatura ambiente. Tras ello se añaden 0,4 µl del anticuerpo THE™ His Tag Antibody [HRP] (mAb – ratón) (GenScript)

a una concentración de 0,5 mg/ml. Se realiza la incubación o bien a 4°C O/N o durante una hora a temperatura ambiente. No es necesaria la realización de una incubación con un anticuerpo secundario ya que este anticuerpo está conjugado con la peroxidasa de rábano (HRP), lo que permite revelar la membrana tras realizar tres lavados de 5 minutos con PBST al 0,1 %. Tras los lavados, se añaden los reactivos del kit Supersignal™ West Pico PLUS Chemiluminiscent Substrate (ThermoFisher) en una proporción 1:1, se incuba durante 5 minutos en oscuridad y se realiza la revelación de la membrana mediante un visualizador Chemidoc MP – Imaging System para detectar la quimioluminiscencia.

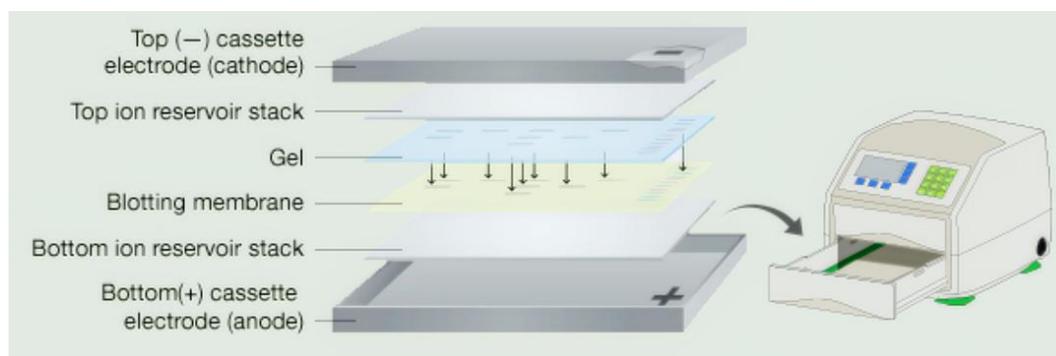


Figura 4.8. Esquema de la correcta colocación de todos los componentes para la electrotransferencia mediante el Trans-Blot Turbo. Imagen adaptada de (BioRAD, 2012).

4.4. Purificación de proteínas.

Al expresar la proteína mediante secreción celular, ya sea en Hi5 o en HEK293F, el primer paso de purificación es realizar una diálisis para cambiarle al sobrenadante el tampón. En este caso, todas las proteínas secretadas fueron dializadas contra el tampón 25 mM TRIS pH 7,5, 300 mM NaCl, 10 mM imidazol O/N realizando tres cambios de diálisis con agitación mantenida a 4°C para conseguir un intercambio de tampón eficaz. Tras los tres cambios de diálisis se realizó la filtración, con filtros de 0,45 µm (Scharlau), del sobrenadante.

4.4.1. Cromatografía de afinidad con HISTRAP (IMAC).

La principal operación de separación realizada en la purificación de proteínas es la cromatografía de afinidad, ya que permite la interacción específica entre ciertos residuos de la proteína y la matriz de la columna de cromatografía, consiguiendo eliminar la mayoría de impurezas que no se deseen mantener en la muestra.

Un tipo especial de cromatografía de afinidad es la denominada IMAC (por sus siglas en inglés) o cromatografía de afinidad por metales inmovilizados. En ella, las interacciones ocurren entre residuos de histidina que hemos añadido a nuestras construcciones y el metal inmovilizado en la matriz, siendo el níquel el usado durante la presente tesis, aunque también es posible el uso de cobalto para este fin. Todas las proteínas purificadas durante este periodo predoctoral han sido purificadas mediante IMAC de níquel. Además, como todas han sido secretadas al medio extracelular, este protocolo es aplicable a todas las proteínas purificadas.

Una vez se ha realizado la filtración del sobrenadante, éste se carga en una columna (o dos si la proteína se expresa muy bien) HisTrap (GE Healthcare – Cytiva) equilibrada previamente con tampón 25 mM TRIS pH 7,5, 500 mM NaCl, 10 mM imidazol. El aumento de NaCl con respecto al tampón de diálisis supone una ventaja gracias a que el aumento en la concentración de NaCl puede eliminar interacciones inespecíficas. Después que la muestra haya sido cargada, se lava la columna hasta que la absorbancia a 280 nm se estabilice. Tras el lavado, se realiza un gradiente de imidazol de 10 a 400 mM con tampón 25 mM TRIS pH 7,5, 300 mM NaCl, 400 mM imidazol para eluir la proteína y separarla de otras impurezas que hayan podido unirse a la columna. Para el almacenamiento y mantenimiento de la columna de níquel – y de la mayoría de columnas usadas en purificación de proteína – es necesario limpiarla con cinco volúmenes de H₂O Milli-Q seguidos de otros cinco volúmenes de etanol al 20% filtrado con filtros de tamaño de poro de 0,22 µm (Scharlab).

Para asegurar que se ha realizado una purificación correcta de nuestra proteína de interés, y conocer si se encuentra correctamente se realiza una electroforesis desnaturante – SDS PAGE.

4.4.2. Intercambio de tampón.

Tras el primer paso de purificación, es necesario realizar un intercambio de tampón para eliminar el imidazol con el que ha eluido la proteína de interés y, en el caso de que sea necesario cortar los glicanos, para cambiar el buffer al óptimo para el corte con EndoH

Aunque el intercambio puede realizarse mediante diálisis o centrífuga (apartado 4.4.5), el uso de columnas para intercambiar tampónes cuando el volumen de la muestra oscila entre 10 y 50 ml es lo más habitual. La columna que se usó fue la HiPrep 26/10 Desalting (Cytiva), que gracias a la matriz porosa que contiene permite separar mediante exclusión molecular los componentes en solución. Esto permite que las moléculas grandes – como son las proteínas – puedan estar menos tiempo retenidas en la columna que las pequeñas – componentes del tampón – y puedan separarse.

Para usar esta columna, es necesario equilibrarla con al menos tres volúmenes de columna (53 ml) de H₂O Milli-Q, para eliminar la disolución en la que se almacena la columna (etanol al 20%), y otros tres volúmenes del tampón al que se va a intercambiar la muestra. Una vez realizado esto, se inyectan como máximo 15 ml de muestra (el cual es el volumen muerto de la columna) que deben comenzar a recogerse tras el volumen muerto. Tras recoger la proteína en el nuevo tampón, se apreciará un cambio en la conectividad del cromatograma que corresponde con el tampón en el que se encontraba antes la proteína. Si fuese necesario inyectar más muestra es necesario esperar a que la columna se vuelva a equilibrar en el tampón que deseamos, volviendo la conductividad al valor inicial. Una vez se ha finalizado

el proceso, se obtiene una solución con la proteína disuelta en el tampón deseado y se debe proceder a lavar la columna, para almacenarla a 4°C, con al menos 3 volúmenes de H₂O Milli-Q seguidos de 3 volúmenes de etanol al 20% filtrado por 0,22 μm.

4.4.3. Corte con Endo-H y purificación.

En el caso de la construcción usada para la cristalización de *DmC1GalT1*, el intercambio de tampón se realizó a 25 mM MES pH 6,2 y 150 mM NaCl, pH óptimo para el correcto funcionamiento de la endoglicosidasa-H (Endo-H) (Fig. 4.9). La Endo-H usada en esta tesis se trata de una proteína de fusión expresada y purificada en el laboratorio, la cual se encuentra unida a la proteína de unión a maltosa (MBP) con una cola de histidinas.

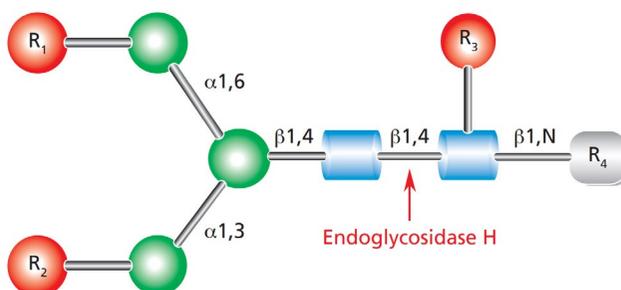


Figura 4.9. Sitio de corte y requerimientos estructurales para que se produzca el corte de los glicanos por la endoglicosidasa-H. Los cilindros azules representan moléculas de GlcNAc y las esferas verdes residuos de manosa. R₁ representa una cadena de oligomanosas; R₂ puede ser cualquier mono u oligosacárido unido al carbono 2 o al 4 o un H; R₃ debe ser un H o una molécula de fucosa unida mediante un enlace α-1-6 para que Endo-H pueda realizar el corte; R₄ representa el aminoácido glicosilado, que es un residuo de asparagina. Imagen adaptada de (Sigma-Aldrich, 2014).

Una vez realizado el intercambio de tampón, se cuantificó mediante Bradford (Biorad) la cantidad de proteína en solución y se añadió Endo-H en un *ratio* 3:250 (Endo-H:Proteína) para cortar los N-glicanos y tener una población homogénea de proteína para la cristalización. Tras incubar la reacción a 18°C durante 20 horas, el corte fue comprobado mediante SDS-PAGE (Figura 4.10). Fue necesario separar una muestra de la solución antes del corte para poder comprobarlo posteriormente.

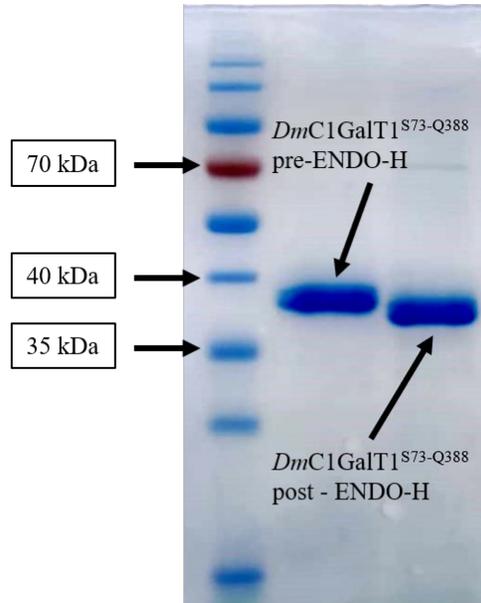


Figura 4.10. Gel SDS-PAGE que muestra la pureza de *DmC1GalT1*^{S73-Q388} y el cambio de peso molecular en el gel tras el tratamiento de la proteína (*overnight*) con Endo-H. El gel de SDS-PAGE se realizó en diversas ocasiones ($n > 6$), confirmando la eliminación de los *N*-glicanos tras el tratamiento con Endo-H.

La MBP-Endo-H – gracias a que contiene la proteína de unión a maltosa (MBP), fusionada a Endo-H – se eliminó de la solución, tras comprobar el corte de los glicanos mediante SDS-PAGE, usando una columna MBP-Trap (Cytiva) equilibrada en tampón 25 mM TRIS pH 7,5 y 150 mM NaCl. Para eluir la Endo-H unida a la columna se usó el mismo tampón en el que fue equilibrada con 10 mM maltosa. Al no unirse nuestra proteína objetivo a la columna, se recoge el líquido que pasa a través de la columna o *flow through*, realizando lo que se denomina una purificación negativa.

Para el buen mantenimiento de la columna MBP-Trap es necesario realizar la regeneración de la matriz, equilibrando la columna con al menos cinco volúmenes (25 ml) de SDS al 0,1% filtrado (0,45 μ m) para eliminar cualquier impureza retenida, tras haberla equilibrado con cinco volúmenes de H₂O Milli-Q. Tras este paso se realiza la limpieza de la columna como se ha mencionado en el apartado 4.4.1. La pureza de la muestra se puede comprobar mediante un gel SDS-PAGE.

4.4.4. Cromatografía de exclusión molecular.

Las columnas de cromatografía de exclusión molecular se componen de una matriz porosa y canales de tamaño selectivo, permitiendo una separación molecular por tamaño. El tiempo de retención de las muestras es inversamente proporcional al tamaño de las mismas, indicando que los agregados proteicos tendrán un menor tiempo de retención que las unidades aisladas. Además, permite definir a través del tiempo de retención el estado de oligomerización que presentan las proteínas o incluso separar los oligómeros en caso de que estén formados por diferentes números de monómeros. Este paso se realiza con el fin de eliminar cualquier tipo de agregados o impurezas o separar los posibles oligómeros de la muestra para facilitar luego la cristalización de las proteínas (van der Laan et al., 1989).

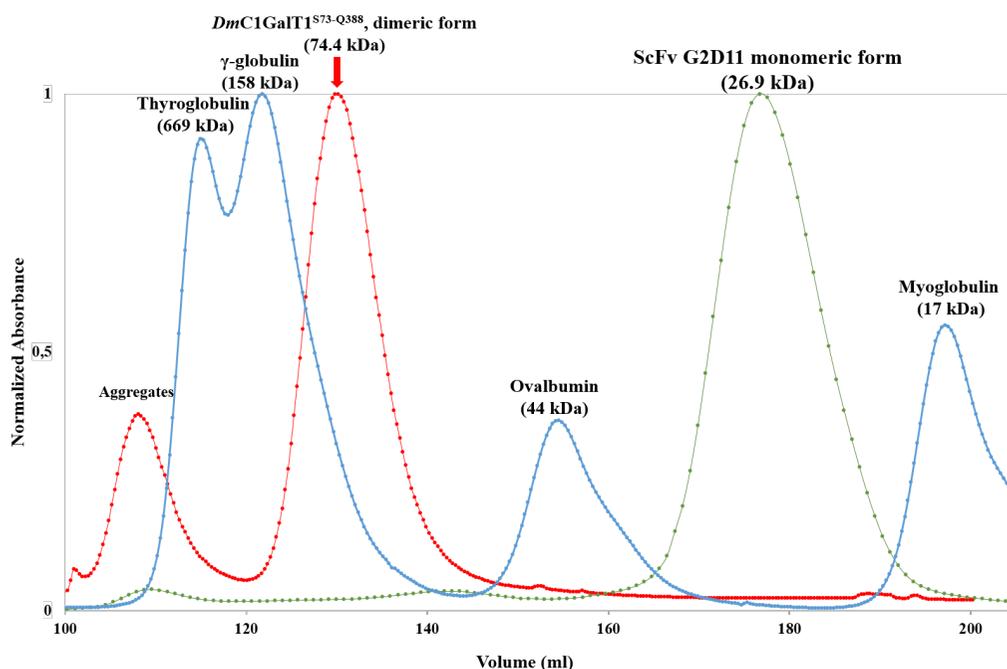


Figura 4.11. Análisis de *DmC1GalT1*^{S73-Q388} mediante cromatografía de exclusión molecular en la columna HiLoad 26/60 Superdex. El cromatograma muestra diferentes proteínas como marcadores de peso molecular y la forma predominante de *DmC1GalT1*^{S73-Q388}, que aparece como dímero (una pequeña parte aparece agregada).

Para realizar la exclusión molecular se usó la HiLoad 26/60 Superdex 75 (Cytiva),

cuyo rango de separación óptimo se encuentra entre los 70 y 3 kDa. El volumen de la columna es 330 ml y el volumen muerto 100 ml. Para su equilibrado es necesario usar un volumen de columna de H₂O Milli-Q y un volumen de columna del tampón en el que se vaya a realizar la exclusión. En todos los casos, se ha usado tampón 25 mM TRIS pH 7,5 y 150 mM NaCl para realizar esta operación. Para llevar la cromatografía a cabo, es necesario reducir el volumen de la muestra para poder inyectar en la columna menos de 5 ml (apartado 4.4.5) a un flujo de 2 ml/min una vez equilibrada en el tampón deseado. Se debe esperar el volumen muerto de la misma (100 ml) y entonces comenzar a recoger fracciones para separar por tamaño las diferentes moléculas en disolución en la muestra inyectada (Figura 4.11). Tras acabar el proceso se realiza una electroforesis en gel de poliacrilamida para comprobar cada una de las fracciones y verificar nuestra proteína de interés.

4.4.5. Concentración, cuantificación y conservación.

Tras el intercambio de tampón o la cromatografía de exclusión molecular es necesario concentrar la proteína bien para realizar la exclusión molecular o para su almacenaje o uso en ensayos, respectivamente. Para ello se realiza la ultrafiltración – mediante centrifugación a 4000 rpm y 4°C – con centricones de 15 ml (Milipore) con un tamaño de poro mayor al peso molecular de la proteína a concentrar (100, 30, 15 o 3 kDa). En esta tesis doctoral, todas las proteínas purificadas tenían un tamaño superior a 30 kDa, por lo que se usaron esos centricones con *cut off* de 30 kDa.

Para llevar a cabo este proceso es necesario equilibrar la membrana del centricón con 15 ml de H₂O Milli-Q seguido de un equilibrado en el tampón que se ha usado en el intercambio de tampón o en la exclusión molecular. Para este proceso se usó una centrífuga de mesa Beckman Coulter Allegra X-30-R.

Existe la posibilidad de realizar en estos centricones el intercambio de tampón;

concentrando y diluyendo el concentrado al menos tres veces en el tampón final deseado. Esta técnica suele realizarse para eliminar la sal presente en el tampón antes de realizar los cribados de cristalización, pero en el caso de *DmC1GalT1*^{S73-Q388} no se llevó a cabo porque la proteína precipitaba al eliminar la sal del tampón.

Una vez las proteínas están concentradas, es necesario cuantificarlas midiendo su absorbancia a 280 nm, usando los coeficientes de extinción teóricos (obtenidos introduciendo en <https://web.expasy.org/protparam/> la secuencia aminoacídica) en un nanodrop espectrofotómetro/fluorímetro DS-11 Series (DeNovix)(Apartado 4.8.1.).

Tras la cuantificación, se almacenan alícuotas de las proteínas purificadas a -80°C, congelándolas en nitrógeno líquido antes de guardarlas, o a 4°C si se van a usar inmediatamente.

4.5. Técnicas de biología celular.

Los experimentos en células fueron realizados en colaboración con los Drs. Yang Zhang y Yoshiki Narimatsu, del departamento de medicina molecular y celular del *Centre for Glycomics*, de la Universidad de Copenhague, Dinamarca.

4.5.1. Análisis de citometría de flujo.

La citometría de flujo es una técnica que permite medir, mientras la célula fluye en suspensión a través de un detector, múltiples características físicas de unidades celulares de manera simultánea, como el tamaño o la granularidad. Su funcionamiento se basa en la dispersión de luz o fluorescencia de las células a investigar, gracias a la adición de colorantes o de anticuerpos monoclonales dirigidos tanto a moléculas extracelulares ubicadas en la superficie celular como a moléculas en el interior celular. De esta forma, la citometría de flujo se convierte en la herramienta perfecta para el análisis detallado de poblaciones celulares rápida y eficazmente (Adan et al., 2017).

Mediante citometría de flujo se pudo determinar la cantidad de *core 1* en la superficie celular, usando un anticuerpo monoclonal de producción propia en la Universidad de Copenhage (Steentoft et al., 2019), el cual está específicamente dirigido a la glicosilación del *core 1* (mAb 3C9).

Para llevar a cabo esto, las células se incubaron en hielo con el mAb 3C9 (se usó el sobrenadante sin diluir del hibridoma en el que se produjo el anticuerpo, lo que corresponde a una dilución 1:1) durante 30 minutos, seguido de un lavado e incubación con el anticuerpo secundario conjugado Alexa Fluor 647 de cabra anti-IgM de ratón (1 ug/ml) (Invitrogen, catálogo: A21235). Los lavados y diluciones se realizaron con PBS (tampón fosfato salino) con 1% de BSA. Las células tratadas se resuspendieron para su análisis por citometría de flujo (SONY SA3800). La intensidad de fluorescencia media de la unión de las poblaciones celulares a mAb 3C9 se cuantificó mediante el software FlowJo (FlowJo LLC).

4.5.2. Inmunocitología.

Esta técnica consiste en la fijación celular en una matriz y su posterior tinción o tratamiento con anticuerpos para observar mediante colorimetría o fluorescencia antígenos o regiones celulares específicas (Skoog & Tani, 2011).

En este caso, las células se fijaron con acetona fría durante 10 minutos y fueron incubadas O/N a 4°C con los anticuerpos anti Myc (mAb 9E10; ATCC, catálogo: CRL-1729) y mAb 3C9, en ambos casos usando sobrenadante no diluido del hibridoma, lo cual es equivalente a una dilución 1:1. Tras ello se realizó la incubación con el anticuerpo secundario conjugado Alexa Fluor 594 de cabra anti-IgM de ratón (1 ug/ml) (Invitrogen, catálogo: A21044). Todas las muestras fueron fotografiadas, utilizando un Zeiss Axioskop 2 plus con AxioCam MR3, y analizadas con ImageJ (NIH).

4.6. Glicopéptidos.

4.6.1. Diseño de glicopéptidos.

El diseño de glicopéptidos específicos para C1GalT1 se ha realizado, en el grupo en el que se ha llevado a cabo la tesis doctoral, de manera racional gracias al estudio previo publicado en 2009 por el grupo del Dr. Thomas Gerken en el que se determinan las preferencias de sustrato de *HsC1GalT1* (Perrine et al., 2009) y teniendo en consideración las secuencias de repetición en tándem de las mucinas.

4.6.2. Síntesis de glicopéptidos.

Por otro lado, la síntesis orgánica y purificación de los péptidos y glicopéptidos diseñados racionalmente se llevó a cabo por el grupo del Dr. Francisco Corzana en la Universidad de la Rioja. Concretamente, el encargado de realizar la síntesis de todos los glicopéptidos y péptidos fue el Dr. Ismael Compañón.

4.6.3. Preparación de las disoluciones de péptidos y glicopéptidos.

Los péptidos y glicopéptidos que fueron sintetizados en el laboratorio del Dr. Corzana fueron recibidos liofilizados y puros. A causa del carácter ácido de la mayoría de sustratos, fue necesario ajustar el pH, tras disolverlos en 25 mM TRIS pH 7,5, con NaOH 2 M y tiras de indicación de pH (Fisher) hasta alcanzar un pH entre 7 y 8. En la mayoría de los casos se intentó dejar los péptidos a 100 mM, pero en algunos casos debido a problemas de solubilidad fue necesario usar concentraciones más diluidas.

4.7. Técnicas de análisis estructural

4.7.1. Cristalización de proteínas.

La cristalización es el proceso mediante el que moléculas homogéneas se ordenan de forma periódica, repetitiva y ordenada en las tres dimensiones espaciales

generando cristales. La cristalización de proteínas es este proceso en el que las unidades proteicas – ya sean oligómeros o monómeros – forman cristales de manera ordenada, en unas condiciones determinadas. Es de vital importancia para la formación de cristales que la muestra sea lo más homogénea posible. Mediante la difracción por rayos X de dichos cristales se puede obtener, tras el procesamiento

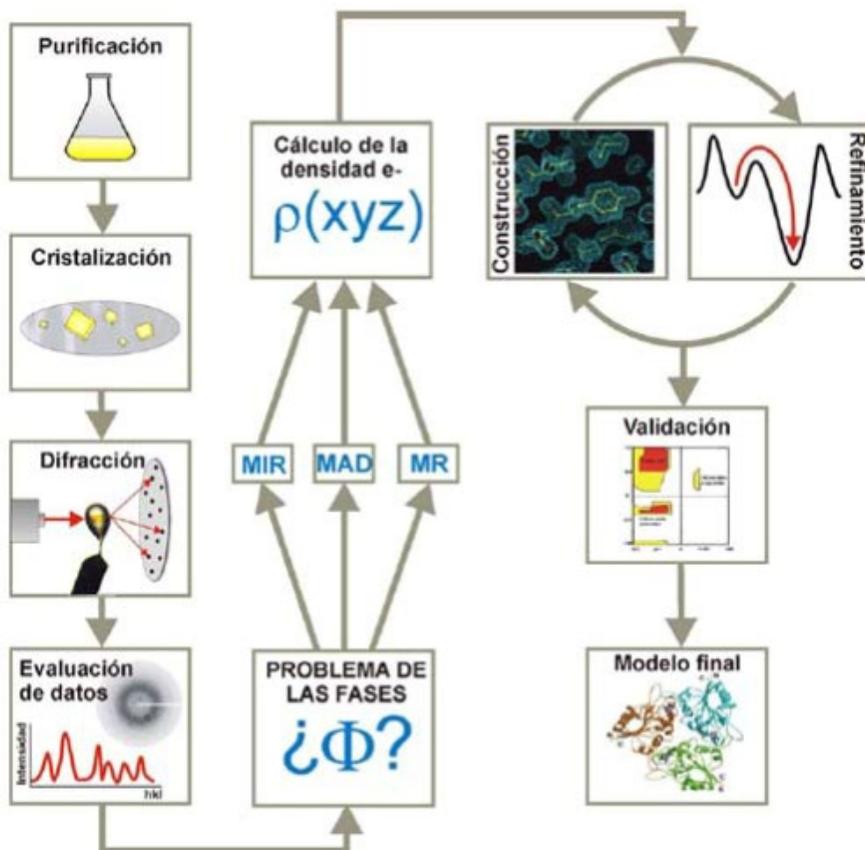


Figura 4.12. Esquema del proceso de resolución estructural de proteínas por cristalización.

Tras la obtención de cristales a partir de proteína purificada en unas condiciones determinadas, los cristales se difractan gracias a una fuente de rayos X. Los datos recogidos de la difracción, una vez procesados otorgan las coordenadas espaciales de los diferentes átomos, aunque la posición absoluta de los mismos no puede dilucidarse hasta que se resuelve el problema de las fases, que se equipara a la obtención de un eje de coordenadas para los datos obtenidos con anterioridad. La resolución de este problema se solucionó en nuestro caso gracias al modelo de *DmC1GalT1* predicho por AlphaFold2 (Jumper et al., 2021). Con esto, se obtiene el mapa de densidad electrónica de la molécula, que se procede a construir y refinar con los datos experimentales hasta obtener la estructura final de la proteína o complejo. Fuente: (Párraga Martínez et al., 2010).

correspondiente, la resolución de la estructura tridimensional de la proteína en el cristal (Fig. 4.12). Aunque la obtención de cristales de proteínas es un proceso extremadamente complicado a nivel experimental (I. Müller, 2017), a día de hoy es responsable de la mayoría de estructuras depositadas en el Protein Data Bank (PDB).

4.7.2. Cocrystalización de *DmC1GalT1*.

Para obtener la estructura tridimensional de una proteína en conjunto con sus ligandos es necesario obtener cristales del complejo que forman la proteína y sus ligandos. El procedimiento que se ha realizado en esta tesis doctoral para la búsqueda de condiciones óptimas de cristalización ha sido el de difusión de vapor en gota sentada (D. R. Davies & Segal, 1971). Este experimento se realiza en placas de 96 pocillos especiales que se componen de una cavidad en la que se añaden 60 μl de una de las condiciones del cribado y dos semicírculos en las que se pondrán las gotas para formar cristales. Del volumen de la solución comercial que se encuentra en el reservorio se mezclan 0,5 μl con 0,5 μl de la proteína, con ligandos si corresponde. La concentración de la proteína se ajusta de manera experimental en cada caso, verificando que la precipitación de las gotas del primer cribado no supere el 40% del total de gotas. Dicho esto, las concentraciones suelen oscilar entre los 5 y 20 mg/ml de manera general. Los cribados comerciales que se usaron fueron: JCSG HTS +, Classic I, Basic, PEG-SALT (Jena Bioscience), Proplex, Midas, Morpheus I, Morpheus II (Molecular Dimensions) y AmSO₄ (QUIAGEN).

La mezcla de proteína usada para los cribados contenía 15 mg/ml *DmC1GalT1*^{S73-Q388}, 5 mM UDP, 2 mM MnCl₂ y 5 mM APDT(GalNAc)RP en buffer 25 mM TRIS, 150 mM NaCl. Una vez se realiza el montaje de la placa, se almacena a 18°C aislada de vibraciones y ruidos, revisándola asiduamente hasta que se hayan formado cristales.

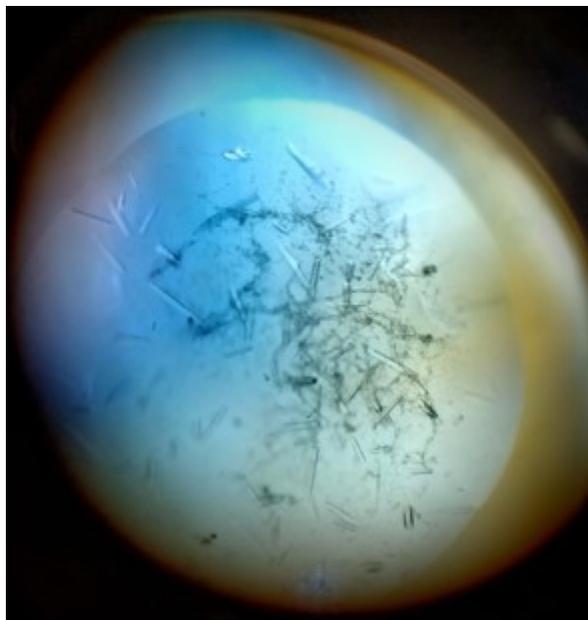


Figura 4.13. Ejemplo de cristales de proteína. Cristales de la enzima *DmC1GalT1* cocrystalizados con 5 mM UDP, 2 mM $MnCl_2$ y 5 mM APDT(GalNAc)RP.

En nuestro caso, los cristales se formaron en diferentes condiciones, pero la disolución del reservorio con la que mejor resolución se obtuvo al difractar los cristales se componía de 100 mM Tiocinato de potasio y 30 % (v/v) Polietilenglicol (PEG) monometil éter 2000 (Figura 4.13).

4.7.3. Toma de datos: difracción de rayos X.

Para la difracción con rayos X es necesario congelar los cristales, una vez se hayan formado, en nitrógeno líquido (Parkin & Hope, 1998). Debido a que la congelación a temperaturas tan bajas (72 K) puede producir la formación de hielo intermolecular en los cristales y desestructurar el empaquetamiento de los mismos se usan soluciones crioprotectoras. Estas soluciones crioprotectoras están formadas por la solución madre en la que se ha formado el cristal con la adición de alcoholes de bajo peso molecular como el etilenglicol o el glicerol (10-30 %). En algunas ocasiones, los precipitantes pueden ser crioprotectores por sí mismos, como es el caso de los compuestos tipo PEG de bajo peso molecular (por ejemplo, PEG 400)

(Kwong & Liu, 1999). En los casos en los que el precipitante de la solución madre sea una sal, se usa como crioprotector una disolución saturada de sulfato de litio (Bujacz et al., 2010). Los cristales difractados de *DmC1GalT1*^{S73-Q388} se crioprotegieron mediante la solución madre con 10% de glicerol antes de ser congelados en nitrógeno líquido.

La recogida de datos tras la difracción de los cristales se llevó a cabo en el sincrotrón Diamond Light Source (Hardwell Science and Innovation Campus, Oxfordshire, UK), concretamente en la línea I24, a una longitud de onda de 0,97 Å y a una temperatura de 100 K.

4.7.4. Resolución de la estructura.

Los datos obtenidos de la difracción de *DmC1GalT1*^{S73-Q388} fueron procesados y escalados con los programas XDS (Kabsch, 2010) y CCP4 (Collaborative Computational Project, 1984; Winn et al., 2011). La estructura cristalina se resolvió mediante reemplazamiento molecular con Phaser (Collaborative Computational Project, 1984; Winn et al., 2011), usando el modelo de *DmC1GalT1* obtenido del servidor de AlphaFold 2 (Jumper et al., 2021).

Tras las fases iniciales, se llevaron a cabo mejoras manuales de la asignación de los aminoácidos correspondientes al mapa de densidad electrónica mediante Coot (Emsley & Cowtan, 2004) ciclos de *building*, seguidos del refinamiento de la estructura mediante REFMAC5 (Murshudov et al., 2011). La estructura final de *DmC1GalT1*^{S73-Q388} en complejo con UDP, Mn²⁺ y APDT*RP se validó con PROCHECK, los datos estadísticos del modelo se mostrarán en el apartado de resultados (Laskowski et al., 1993). Cada unidad asimétrica del grupo especial P2₁ del cristal contenía dos moléculas de *DmC1GalT1*^{S73-Q388}. El diagrama de Ramachandran para el complejo muestra que el 87,8%, 11,2%, 1% y 0% de aminoácidos se encuentran en regiones favorables, permitidas, generosamente

permitidas y no permitidas, respectivamente. Las coordenadas y factores estructurales fueron depositados en el World Wide Protein Data Bank (wwPDB).

4.7.5. Resonancia magnética nuclear (RMN).

La cuantificación de las transiciones que se producen entre dos niveles de excitación próximos de energía, presentes en los núcleos atómicos con spin $\frac{1}{2}$ cuando absorben energía a una frecuencia determinada, en presencia de un campo magnético externo es el fenómeno al que se le denomina como resonancia magnética nuclear (Ferencz & Wagner, 2000). En el caso de las proteínas, los únicos átomos con este tipo de spin son los átomos de hidrógeno, lo que permite obtener espectros monodimensionales de ^1H de los que pueden obtenerse información sobre el plegamiento de las proteínas, presentando señales que se caracterizan por su desplazamiento químico (δ) y su intensidad (I). Cada núcleo atómico, dependiendo del entorno del mismo hace que varíe el valor de δ , permitiendo detectar variaciones que hayan sido causadas por la presencia o ausencia de algún ligando.

Todos los experimentos de RMN se llevaron a cabo en un espectrómetro Bruker Avance III 600 MHz equipado con una sonda criogénica de triple resonancia inversa de 5mm capaz de realizar gradientes de campo de pulso en el eje z. Las resonancias de los espectros ^1H -RMN de los glicopéptidos fueron asignados a través de experimentos 2D-TOCSY (espectroscopía de correlación total) (30 y 80 ms de tiempo de mezcla), 2D-NOESY (espectroscopía de efecto nuclear Overhauser) (400 ms de tiempo de mezcla) y ^1H , ^{13}C -HSQC (coherencia heteronuclear cuántica simple) a 283 y 298 K. El α -O-metil-GalNAc (Carbosynth, MM06786) se asignó mediante experimentos 2D-TOCSY/NOESY y ^1H , ^{13}C -HSQC a 298 K. Las concentraciones típicas que se usaron fueron alrededor de 1 mM para los experimentos heteronucleares en tampón 25 mM TRIS(D₁₁)-DCl, pH 7,5 y 150 mM de NaCl en una mezcla 90:10 de H₂O/D₂O. En los experimentos de ^1H -RMN se tomó como referencia de desplazamiento químico la resonancia del

TSP (ácido 2,2,3,3-tetradeutero-3-trimetilsilil propiónico) (δ TSP = 0 ppm).

4.7.6. Resonancia magnética de diferencia de transferencia de saturación (STD-RMN).

Dependiendo de las interacciones proteicas que se deseen medir existen diferentes formas de usar RMN para ello. En este estudio se ha usado STD-RMN como principal técnica complementaria, la cual permite caracterizar, tras su detección, interacciones receptor-ligando en disolución. Las señales que se detectan mediante esta técnica son las de los ligandos debido a que presentan las propiedades espectroscópicas necesarias para estos estudios a alta resolución (Angulo & Nieto, 2011). En la figura 4.14 se representa un esquema del funcionamiento de STD-RMN

El ensayo de STD-RMN está constituido por las tres etapas o procesos:

- El primer paso se trata de realizar un espectro monodimensional ^1H estándar de la proteína, con los ligandos y cofactores necesarios para su correcto funcionamiento, y el ligando, en cuestión a estudiar, en exceso a una frecuencia de irradiación lo suficientemente lejana de la frecuencia de resonancia tanto del ligando como de la proteína. A este primer espectro se le denomina *off-resonance* y es el que se tomará como referencia para obtener las intensidades del ligando.
- El segundo paso consiste en la saturación selectiva de la proteína mediante radiación en una frecuencia específica (en la frecuencia de los carbonos alifáticos, por ejemplo). Esta saturación selectiva se transmite favorablemente por toda la macromolécula gracias al denominado efecto nuclear Overhauser (NOE), incluso llegando a aquellas zonas que están en contacto directo con el ligando. Las señales saturadas no aparecerán en este espectro.
- Por último, se realiza la diferencia entre los dos espectros, obteniendo el

espectro STD, en el que se pueden observar las señales asignables a los protones del ligando que están en contacto directo con la superficie proteica. La intensidad de esas señales se corresponderá con la proximidad de los protones a la proteína.

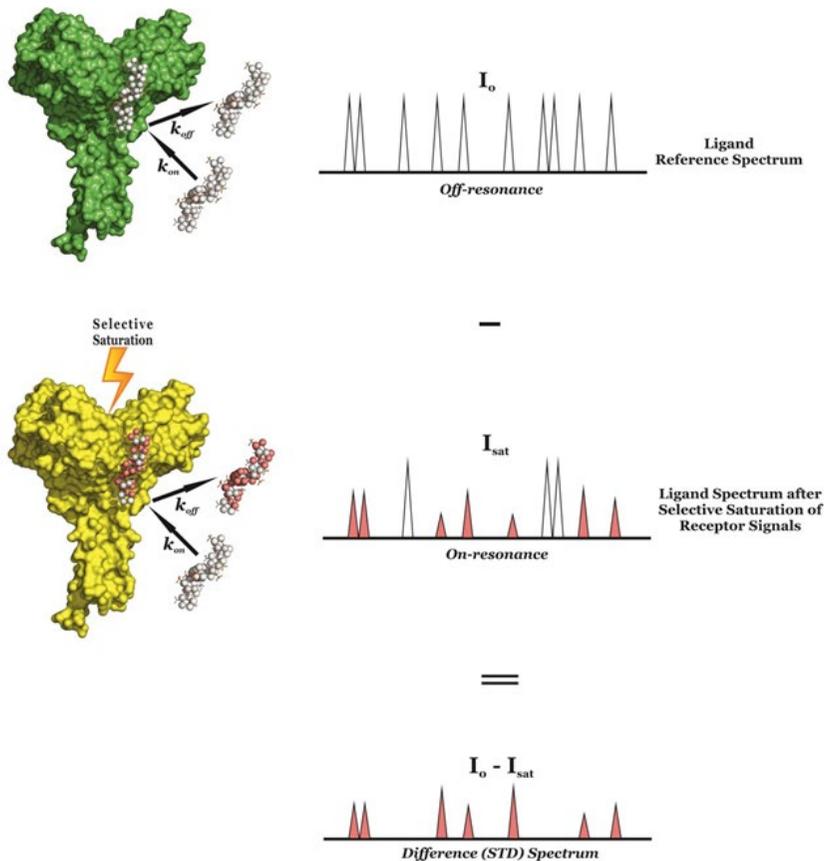


Figura 4.14. Esquema de un experimento de STD-RMN. En la imagen aparece la representación de la proteína junto a los protones no intercambiables del ligando (esferas). En primer lugar, en el espectro monodimensional ^1H superior se representan las intensidades de referencia del sistema en equilibrio. En el espectro intermedio se realiza la saturación selectiva de la proteína, lo que transmite vía difusión de spin (NOEs) la saturación a los protones del ligando que está interactuando con la superficie proteica directamente, los cuales se encuentran representados en naranja. El espectro inferior es la diferencia de los dos primeros espectros, obteniéndose un espectro diferencial en el que se muestran sólo las señales de los protones que han participado en la interacción mayor o menor intensidad dependiendo de la proximidad a la superficie. Fuente: <http://www.glycopedia.eu/chapters/NMR-for-Structural-Glycoscience-35/Saturation-Transfer-Difference>.

Gracias a estos experimentos se ha conseguido caracterizar de manera más

específica, las interacciones proteína-ligando. Todos los experimentos de STD realizados en este proyecto se han llevado a cabo por el grupo de investigación de la Doctora Filipa Marcelo de la Universidad Nova de Lisboa. A continuación, se detallarán las condiciones de los ensayos realizados.

Los experimentos STD se realizaron usando un *ratio* molar de 1:35, definido por 20 μM *DmC1GalT1*^{T43-Q388} y 710 μM ligandos (α -O-metil-GalNAc y glicopeptidos), y tampón 25 mM TRIS(D11)-DCl, pD 7,5, 150 mM NaCl y 150 μM MnCl₂ en D₂O. Algunos de los STD se realizaron en ausencia y presencia de UDP (135 μM). En la presencia de 150 μM MnCl₂ se observan fuertes aumentos en la relajación paramagnética del UDP, lo que impide la detección de las señales de los protones del UDP en el espectro de RMN. De cualquier forma, la presencia de MnCl₂ no impide observar las señales de protones de los ligandos, permitiendo extraer información de estos ensayos de STD. Los espectros STD-RMN (secuencia de pulso *stddiffesgp* de la biblioteca del software Bruker) fueron adquiridos con 1728 escaneos y 64 K puntos de datos, en una ventana espectral de 12335,53 Hz centrada a 2818 Hz. La saturación selectiva (*on resonance*) se realiza irradiando a 7 y/o 0,5 ppm – dependiendo de si el ligando contenía o no residuos aromáticos – usando un tren de pulsos a 90° (50 ms) con forma de 40 Eburp1.10000 (de la biblioteca de pulsos con forma de Bruker) para conseguir un tiempo de saturación total de 2 s y un tiempo de retraso de relajación de 4 s. Para el espectro *off-resonance*, las muestras se irradiaron a 110 ppm. Para cada uno de los ligandos se llevaron a cabo controles apropiados en la ausencia de proteínas y se observaron señales residuales de STD debidas a los grupos metilos de las alaninas y treoninas. Estos resultados se tuvieron en cuenta a la hora de analizar los experimentos de STD en presencia de *DmC1GalT1*^{T43-Q388}. Se realizaron además controles de STD con la proteína en ausencia de ligandos y se sustrajeron a la hora de realizar los análisis.

El espectro STD (I_{STD}) se obtuvo restando el espectro *on resonance* (I_{on}) al espectro *off resonance*. El porcentaje de STD ($I_{STD}/I_{off} \times 100$) se estimó comparando la intensidad de las señales de espectro STD (I_{STD}) con las señales del espectro de referencia (I_{off}). El factor de amplificación de STD (STD_{AF}) también fue estimado, multiplicando los porcentajes de los valores de STD producidos por el exceso de ligando (Mayer & Meyer, 2001), que en todos los experimentos realizados fue 35 para cada ligando. Para determinar el mapa de epítomos derivado de los STD se calculó el porcentaje relativo al ajuste de STD al 100 %, determinando el 100 % como el protón con mayor intensidad y calculando los demás protones acorde a estos parámetros, que se mostrarán en el apartado de resultados. Algunos protones no pudieron evaluarse con precisión debida a la supresión de la señal del agua y a la baja señal/ruido de fondo, por lo que se muestran con un círculo azul en los mapas de epítomos derivados de los STD. Además, las resonancias que se superponían en los espectros 1H -RMN fueron consideradas en la estimación de los STD y marcadas con un *.

4.8. Técnicas biofísicas para el análisis y caracterización de proteínas.

4.8.1. Absorción UV-visible y cuantificación de proteínas.

El espectro de absorción ultravioleta (UV) de las proteínas se conforma por dos picos principales característicos, el primero de ellos corresponde a la absorción del enlace peptídico y se encuentra alrededor de los 225 nm; el segundo está en torno a los 280 nm y es causado por la absorción de los enlaces tipo sp^2 de los residuos con grupos funcionales con anillos aromáticos, como la tirosina o el triptófano principalmente (Creighton, 1997).

Durante esta tesis doctoral, esta técnica ha sido la forma principal de cuantificación de proteínas puras usando su coeficiente de extinción teórico y la absorción UV-visible a 280 nm, obtenido mediante la introducción de la secuencia aminoacídica

en servidores como Protparam. La relación entre la absorción a 280 nm y el coeficiente de extinción está definido por la ecuación de Lambert-Beer (Luong et al., 2011; Swinehart, 1962):

$$A = \epsilon_{280 \text{ nm}} \times C \times l$$

A representa el valor de la absorbancia medida a una longitud de onda específica, siendo ϵ el coeficiente de extinción molar de la proteína a esa longitud de onda ($M^{-1}cm^{-1}$), l es la longitud del paso óptico de la muestra (cm) y C es la concentración (M). Las mediciones de absorbancia a 280 nm se realizaron en un nanodrop espectrofotómetro/fluorímetro DS-11 Series (DeNovix). Los coeficientes de extinción de las proteínas silvestres y de los mutantes se han mostrado en el apartado 4.4.5.

4.8.2. Determinación de la constante de unión por ITC.

Para obtener las constantes de equilibrio y los parámetros termodinámicos entre nuestra enzima y sus ligandos usamos la calorimetría de titulación isotérmica o ITC. Esta técnica permite medir el calor generado o absorbido cuando se mezclan dos soluciones de especies diferentes. Una de las soluciones se encuentra en lo que se denomina celda, mientras que la otra solución – que suele ser la que se encuentra a mayor concentración – está en otro compartimento denominado jeringa. La medición ocurre cuando se producen inyecciones de la jeringa en la celda de la muestra. Esta técnica (Figura 4.15) se usó para caracterizar la interacción de *DmC1GalT1*^{T43-Q388} con UDP, APDT*RP en presencia y ausencia de UDP, APDTRP con UDP y P4/P7 con y sin UDP. Los experimentos se llevaron a cabo en un aparato Auto-iTC200 (Microcal, Cytiva) con un programa de 19 inyecciones de 2 μ l en la celda, con 150 segundos de reposo entre las mismas, a una temperatura de 20°C.

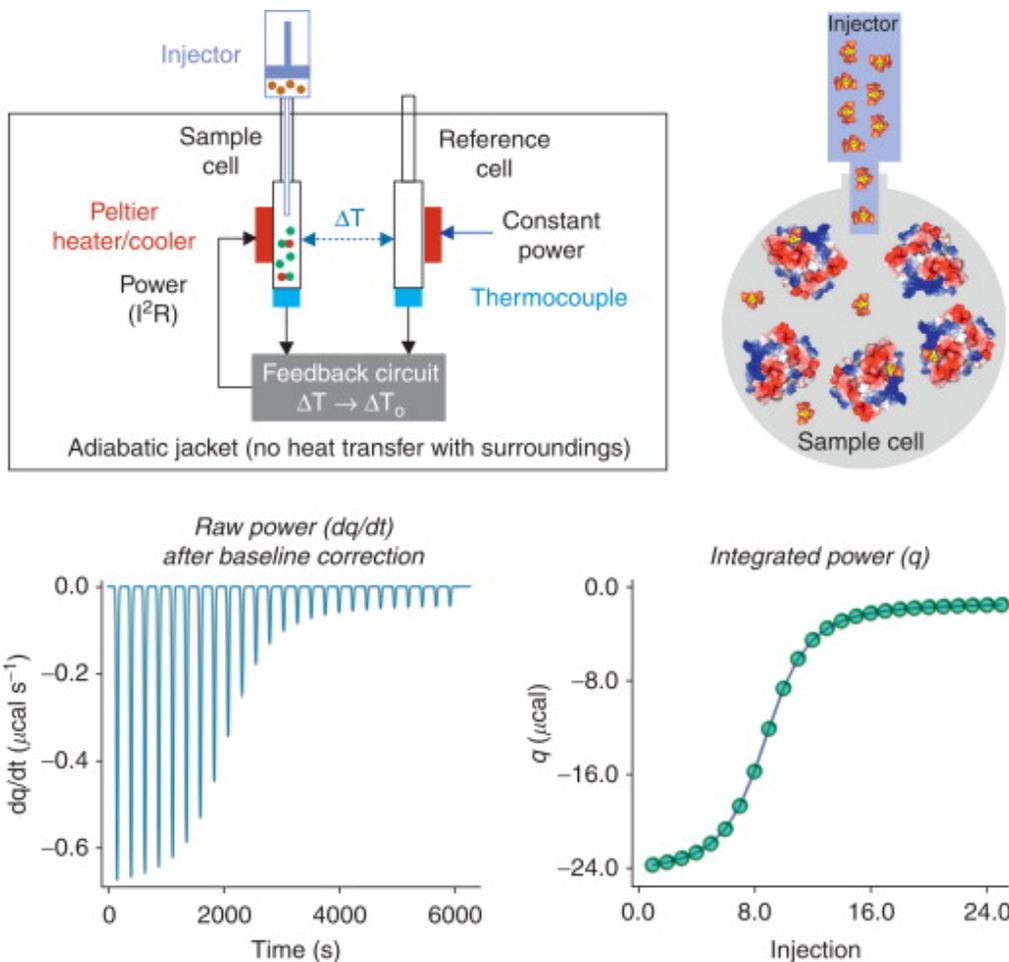


Figura 4.15. Esquema general del funcionamiento de la calorimetría de titulación isoterma. En la parte superior se representa un esquema del equipo en el que se puede observar una cámara adiabática con dos celdas, una de referencia (agua) y una con la muestra de proteína en la que se va inyectando progresivamente el ligando. En la parte inferior izquierda de la imagen se observan los datos brutos obtenidos a partir de la diferencia de calor entre la celda de referencia y la celda de la muestra. Por último, en la parte inferior derecha de la imagen se muestran los datos integrados de cuyo ajuste se obtiene la constante unión (K), la entalpía (ΔH) y la estequiometría de unión (n). Fuente: (Dam & Fred Brewer, 2007).

Las concentraciones usadas fueron las siguientes:

- $DmC1GalT1^{T43-Q388}$ y UDP: 100 μM de enzima con 800 μM de UDP en tampón 25 mM TRIS pH 7,5, 150 mM NaCl y 1 mM MnCl_2 .
- $DmC1GalT1^{T43-Q388}$ y APDT*RP, P4 y P7 en ausencia de UDP: 60 μM de enzima con 2 mM de glicopéptidos en tampón 25 mM TRIS pH 7,5, 150

mM NaCl y 1 mM MnCl₂.

- *DmC1GalT1*^{T43-Q388} y APDT*RP, P4 y P7 en exceso de UDP: los experimentos se realizaron en tampón 25 mM TRIS pH 7,5, 150 mM NaCl, 1 mM MnCl₂ y 1 mM de UDP. La concentración de enzima fue 60 μM para las titulaciones de APDT*RP y APDTRP y 50 μM para las titulaciones de P4 y P7. La concentración de los glicopéptidos y péptido fue 2 mM en todos los casos.

Los experimentos se realizaron en duplicado y la integración, corrección y análisis de los datos se llevó a cabo mediante el software Origin 7 (Microcal). Los datos se ajustaron al modelo de unión de equilibrio de un solo sitio de unión. La estequiometría (n) de todos los experimentos fue en todos los casos ~1:1, excepto para el UDP, cuya n resultó ser 0,4.

4.8.3. Análisis de actividad enzimática.

Con el objetivo de caracterizar y comprender el funcionamiento de C1GalT1 y su relación sus diferentes ligandos y sustratos, se realizaron ensayos de actividad enzimática. Debido a que no existen métodos para poder medir la reacción en tiempo real, se tuvieron que usar kits de detección de ATP en los que se acopla la reacción de transferencia de galactosa, liberando UDP y usando ese sustrato para transformarlo en ATP, uniéndose al reactivo de detección para producir luminiscencia. En concreto, el kit que se usó fue el UDP-GloTM (Promega) (Figura 4.16).

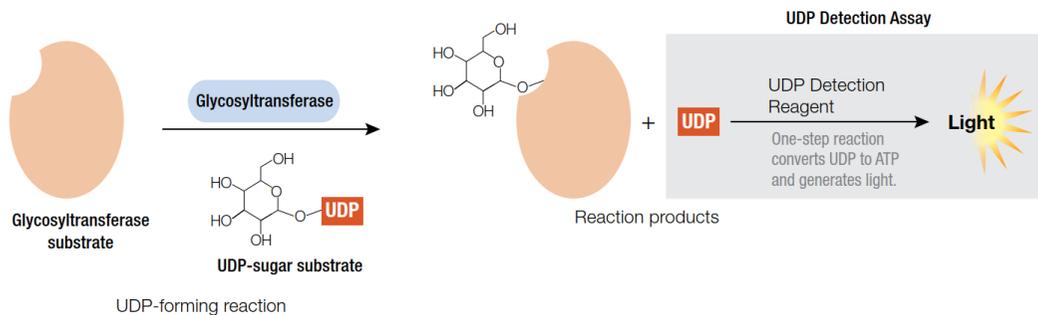


Figura 4.16. Principios del funcionamiento del kit UDP-Glo para las cinéticas de C1GalT1.
Fuente: (Promega, 2020).

Las reacciones se incubaron durante 30 minutos a 37°C y se pararon usando el reactivo de detección de UDP, reactivo de detección que se mezcla en proporción 1:1 con la reacción y que debe incubarse durante 1 hora a temperatura ambiente en oscuridad. Tras la incubación, se debe medir la luminiscencia, lo cual se realizó en un lector de placas Synergy HT (Biotek). Es necesario que las placas sean blancas y opacas, es irrelevante si son de 96 o de 384 pocillos, aunque estos experimentos se realizaron en las segundas ya que se podía usar menor volumen de reacción y de reactivo de detección (5 µl). Con el objetivo de estimar la cantidad de UDP producido en la reacción de transferencia, se realizó en cada uno de los ensayos una recta de calibrado de UDP, sometiendo dicho sustrato a la misma temperatura y tiempo de incubación que las reacciones.

Los valores obtenidos se corrigieron para eliminar la hidrólisis de UDP-Galactosa y se ajustaron al modelo no lineal de Michaelis-Menten usando el programa Graphpad Prism 8, del cual se obtuvieron los valores aparentes de K_m , k_{cat} y V_{max} junto con el error estándar de cada uno de los parámetros. Todos los experimentos se realizaron por duplicado excepto la actividad de los mutantes de *DmC1GalT1* en los que se realizaron por triplicado.

Antes de poder realizar las cinéticas de los diferentes sustratos, fue necesario establecer la cantidad de UDP-Gal que se iba a mantener constante en las reacciones, calculando sus parámetros cinéticos para usar condiciones saturantes.

Para ello se realizaron reacciones con 500 nM de *DmC1GalT1*^{T43-Q388} y concentraciones variables de UDP-Gal (12,5 μ M – 1 mM) en presencia de P4 y APDT*RP bajo condiciones saturantes (250 μ M y 1 mM respectivamente) en tampón 25 mM TRIS pH 7,5, 150 mM NaCl, 50 μ M MnCl₂ y 1 mg/ml BSA.

Las reacciones enzimáticas de *DmC1GalT1*^{T43-Q388}, *DmC1GalT1*^{S73-Q388} y los mutantes contenían 500 nM de las enzimas en 25 mM TRIS pH 7,5, 150 mM NaCl 50 μ M MnCl₂, 1 mg/ml BSA y 500 μ M UDP-Gal en presencia de concentraciones variables de los péptidos y del α -O-metil-GalNAc, oscilando entre 12,5 y 500 μ M y entre 125 μ M y 2 mM, respectivamente. En el caso de los glicopéptidos APDT*RP y APDS*RP se usaron concentraciones hasta 1 mM para obtener un mejor ajuste cinético no lineal de Michaelis-Menten. Para los mutantes, el ensayo de actividad se realizó en el mismo tampón mencionado anteriormente con 500 nM de las enzimas, 500 μ M UDP y 500 μ M APDT*RP.

4.9. Técnicas computacionales.

4.9.1. Estudio de homología: Alineamiento de secuencias de C1GalT1.

El alineamiento de secuencias homólogas de C1GalT1 se realizó mediante el servidor online de alineamiento múltiple de secuencias Clustal Omega (Sievers et al., 2011) introduciendo las secuencias aminoacídicas de diferentes especies en formato FASTA tras realizar una búsqueda mediante BLAST (*Basic Local Alignment Search Tool*) (States & Gish, 1994).

4.9.2. Predicción de la estructura secundaria de *DmC1GalT1*.

La predicción de la estructura secundaria de *DmC1GalT1* para obtener construcciones en las que se eliminasen regiones desordenadas que dificultasen la cristalización de la proteína se realizó mediante el servidor de predicción de la estructura secundaria de proteínas JPRED4 (<https://www.compbio.dundee.ac.uk/jpred/>) introduciendo la secuencia

aminoacídica de *DmC1GalT1* (Drozdetskiy et al., 2015).

4.9.3. Dinámicas moleculares (MD).

Una de las técnicas computacionales que permite extrapolar, previsualizar e interpretar los datos obtenidos experimentalmente es la simulación de dinámicas moleculares (MD, de sus siglas en inglés), obteniendo información enriquecedora y complementaria a los datos experimentales.

A partir de estructuras de proteínas de interés y definiendo parámetros estéricos y de interacción con el entorno de manera matemática se predicen mediante programas informáticos el movimiento y posición de los átomos, además de sus posiciones respectivas, a lo largo de un determinado tiempo.

Gracias a esta técnica, que se ha llevado a cabo en el grupo del Doctor Francisco Corzana, de la Universidad de la Rioja, se ha conseguido obtener información adicional y a complementar los estudios sobre *C1GalT1*. A continuación se describirá detalladamente los parámetros usados en las MDs realizadas.

La estructura cristalina de *DmC1GalT1*-UDP-Mn²⁺-APDT*RP fue superpuesta con el complejo de la enzima humana B3GNT2-UDP-GlcNAc (PDB 7JHL), proporcionando las coordenadas de la molécula de UDP-GlcNAc, encontrándose en la misma posición que la molécula de UDP unida a *DmC1GalT1*. Una vez se generó el complejo *DmC1GalT1*-UDP-GlcNAc-Mn²⁺-APDT*RP, se reemplazó el UDP-GlcNAc por una molécula de UDP-Gal, formando el complejo UDP-Gal-Mn²⁺-APDT*RP. Los otros complejos fueron generados añadiendo o eliminando los residuos correspondientes mediante PyMOL 2.5. Los cálculos se realizaron mediante el paquete AMBER 20, implementados con los campos de fuerza ff14SB y GLYCAM06 (Maier et al., 2015). Cada complejo fue sumergido en un cubículo de agua con 10 Å de tampón de moléculas de agua TIP3P. El sistema se neutralizó con la adición explícita de contraiones (Na⁺ o Cl⁻). A continuación, se realizó una

optimización geométrica en dos etapas. La primera tiene como único objetivo minimizar las posiciones de las moléculas del solvente, mientras que la segunda minimiza sin restricciones todos los átomos en la celda de simulación. Tras ello, los sistemas fueron suavemente calentados incrementando la temperatura desde 0 a 300 K bajo una presión constante de 1 atm y condiciones de contorno periódicas. Se aplicaron restricciones harmónicas de 30 kcal mol⁻¹ al soluto, y se usó el esquema de acoplamiento de temperatura Andersen para controlar y ecualizar la temperatura. Las etapas de tiempo se mantuvieron a 1 fs durante las etapas de calentamiento para permitir que las inhomogeneidades se autoajustasen.

Los efectos electrostáticos de largo alcance se calcularon usando las condiciones de contorno periódicas con malla de Ewald (método de mallas de partículas de Ewald). Se aplicó un límite de 8 Å a las interacciones de tipo Lennard-Jones. Cada sistema se equilibró durante 2 ns con etapas de reposo de 2 fs a volumen constante y temperatura de 300 K. Tras ello, se ejecutaron las trayectorias de producción durante 0,5 μs adicionales bajo las mismas condiciones de simulación.

Para el cálculo de los mapas de energía libre de los glicopéptidos APDT*RP y APDS*RP, en agua a 300K, se implementó un método en AMBER 20 de dinámica molecular con sesgo adaptativo (ABMD o “Adaptively Biased Molecular Dynamics”) (Babin et al., 2008).

5. RESULTADOS

5.1. Alineamiento múltiple de C1GalT1. Elección de ortólogos para la expresión, purificación y caracterización biofísica.

Debido a la necesidad de co-expresar Cosmc con C1GalT1 en organismos superiores, se realizó un alineamiento múltiple de diferentes especies para revisar los niveles de homología entre diferentes ortólogos (Figura 5.1) con el objetivo de seleccionar una C1GalT1 que mantuviese una elevada homología con la C1GalT1 humana pero que no necesitase la chaperona específica. En este alineamiento se observan los motivos mencionados en el apartado 2.4.1., DDD y CCSD que se mantienen en todos los ortólogos (Figura 5.1)(Ju et al., 2006). Debido a que los ortólogos de C1GalT1 en invertebrados no necesitan de la chaperona específica, son de especial interés para la expresión recombinante de los mismos. Cabe destacar la existencia de varios sitios diferentes de *N*-glicosilación en los ortólogos de C1GalT1 de invertebrados. Es por ello que se decidió realizar la prueba de expresión recombinante tanto del ortólogo de *C. elegans* como de *D. melanogaster* tanto en células de mamífero como en el sistema de expresión de baculovirus en células de insecto. Para esto último, se tuvo que realizar la optimización de este sistema de expresión, puesto que no se había realizado antes en el grupo de investigación, tomando como referencia el manual de Invitrogen Bac to Bac (Invitrogen, 2015) junto a la transfección de bácmidos en células de insecto en suspensión desarrollada por “Scholz & Suppmann, 2017” y a través de la ayuda de diferentes grupos de investigación con mayor experiencia en la expresión en este sistema.

Figura 5.1. Alineamiento múltiple de las secuencias de C1GalT1. Las secuencias que se muestran pertenecen a las siguientes especies: humano (*Hs*), chimpancé (*Pt*), perro (*Cf*), vaca (*Bt*), rata (*Rn*), ratón (*Mm*), gallo (*Gg*), rana (*Xl*), pez cebra (*Dr*), mosca (*Dm*) y *Caenorhabditis elegans* (*Ce*). Los residuos se encuentran coloreados según el nivel de conservación de la secuencia, donde el negro denota secuencia aminoacídica idéntica o muy alta, las tonalidades de gris identidad o similitud media y blanco ninguna similitud. En rojo se encuentran resaltados los motivos DDD y CCSD y en verde los posibles sitios de *N*-glicosilación solo presentes en organismos inferiores.

5.2. Análisis computacional de la estructura secundaria de *DmC1GalT1* y diseño de construcciones para su cristalización: pruebas de expresión con baculovirus recombinantes.

Para comenzar, se realizó el diseño de un constructo en el que se eliminó la secuencia señal y el dominio transmembrana predicho en el ortólogo de *D. melanogaster* (T43-Q388). Se decidió realizar los ensayos con el ortólogo de mosca pues en una primera prueba se comprobó que se expresaba de manera recombinante en HEK293F mucho mejor que el equivalente en *C. elegans*. Sin embargo, también procedimos a expresar la C1GalT1 de gusano usando el mismo razonamiento mencionado arriba, sin obtener éxito en la expresión (datos no mostrados). Por tanto, se centró el esfuerzo en expresar, purificar e intentar cristalizar la construcción de *DmC1GalT1*. Lamentablemente, al no tener éxito en la producción de cristales para obtener la estructura tridimensional, se cambió el sistema de expresión, realizando la expresión del mismo constructo, pero en el sistema de expresión de baculovirus. El cambio de sistema de expresión se debe a que en insectos la *N*-glicosilación es menos compleja que en mamíferos, por lo que es más sencillo y susceptible su tratamiento con endoglicosidasas (en este caso Endo-H) para obtener una población proteica homogénea para su cristalización. En este caso, se observó que la proteína sufría una clara degradación apareciendo dos bandas diferentes en los geles SDS-PAGE. Ambas bandas se identificaron como C1GalT1 de *D. melanogaster* mediante MALDI-TOF-TOF por lo que se procedió a realizar un análisis de la estructura secundaria de la secuencia aminoacídica para identificar regiones de la cadena polipeptídica que se encontrasen desordenadas y que pudiesen

ser la causa de dicha degradación, mejorando así la facilidad de cristalización de la proteína (Figura 5.2).

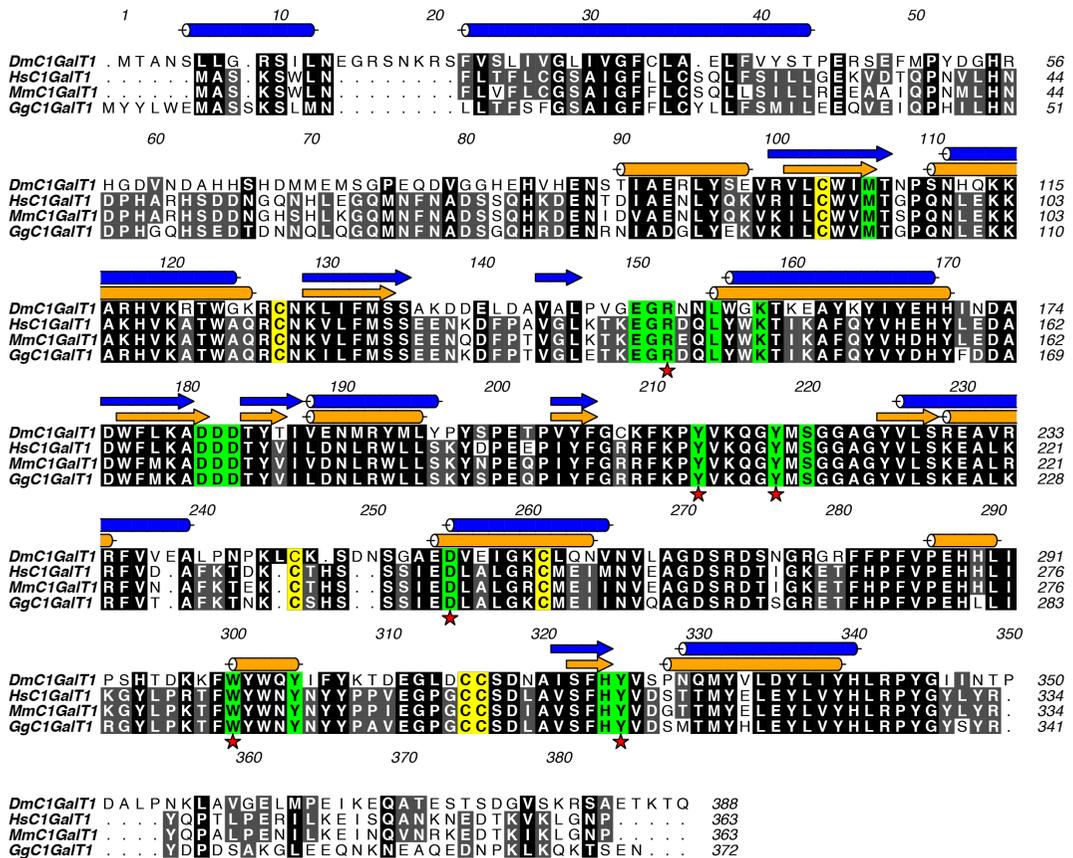


Figura 5.2. Alineamiento múltiple de las secuencias de C1GalT1 y predicción de la estructura secundaria. Las secuencias que se muestran pertenecen a las siguientes especies: mosca (*Dm*), humano (*Hs*), gallo (*Gg*) y ratón (*Mm*). Los residuos se encuentran coloreados según el nivel de conservación de la secuencia, donde negro, gris y blanco representan, respectivamente, alta, media o ninguna similitud/identidad. Justo arriba de la secuencia se muestra los elementos de estructura secundaria (hélices α y láminas β) predichos mediante el servidor JPRED4 en azul y naranja las estructuras obtenidas basándose en la estructura tridimensional de *Dm*C1GalT1. Las cisteínas que participan en la formación de puentes disulfuro se encuentran resaltados en amarillo. Los residuos involucrados en el reconocimiento de ligandos se muestran subrayados de color verde y los residuos que se han mutado en este trabajo se indican con una estrella roja.

El análisis realizado mediante JPRED4 (<https://www.compbio.dundee.ac.uk/jpred/>) permitió el diseño de cinco nuevas construcciones. El diseño de las mismas se basó en la eliminación de regiones de la

cadena polipeptídica que no tuviesen estructura secundaria (Figura 5.2). De esta forma, se realizaron dos constructos en los que fueron eliminadas las zonas del extremo N-terminal: el primero hasta el aminoácido His55 y el segundo hasta el residuo Ser73. De igual forma, se realizaron otras tres construcciones eliminando zonas del extremo C-terminal, eliminando la secuencia desde Q388 hasta Pro343, Pro354 y Ser373.

5.2.1. Pruebas de expresión de construcciones truncadas de *DmC1GalT1*.

Se realizó la prueba de expresión de las cinco construcciones realizadas tanto en células Hi5 como en células Sf9 recogiendo muestras a las 48 y 72 horas. Las construcciones que mejor expresión presentaban eran las que tenían el extremo C-terminal intacto, siendo esas H55-Q388 y S73-Q388. Además, con respecto a las otras tres construcciones, en la que menos secuencia del C-terminal se había eliminado seguía presentado una ligera buena expresión (T43-S373) mientras que

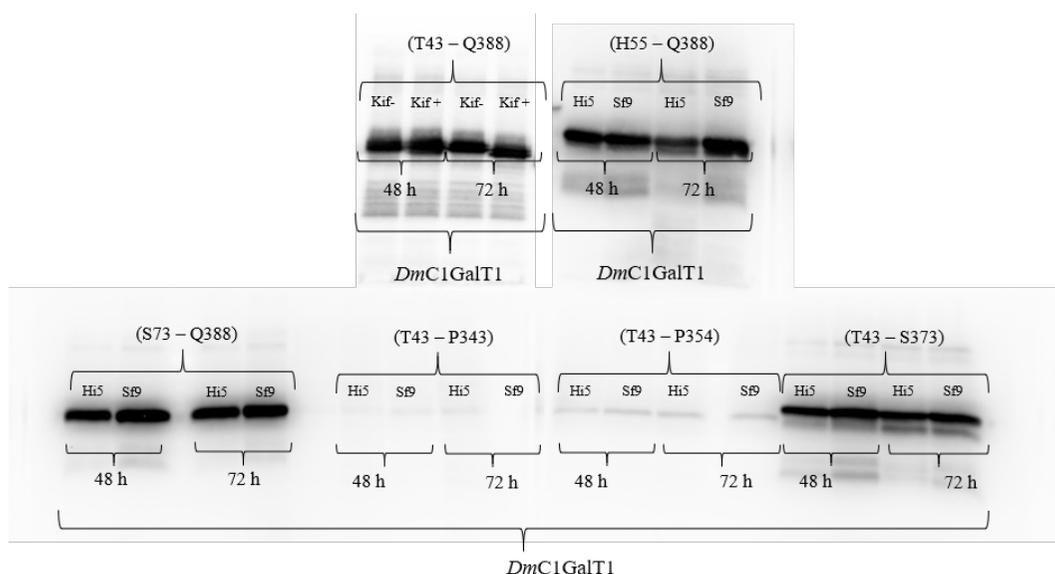


Figura 5.3. Pruebas de expresión de construcciones de *DmC1GalT1*. Se muestran las pruebas de expresión de la construcción inicial con y sin kifunensina y de las construcciones diseñadas a partir de la predicción de estructura secundaria de *DmC1GalT1* mediante JPRED4.

los otros dos apenas se expresaban (T43-P343 y T43-P354). Tanto el constructo

H55-Q388 como el T43-S373 mantenían la degradación de la construcción inicial, mientras que en el S73-Q388 no se observaba esa degradación (Figura 5.3). Esta última construcción fue la que se eligió para la formación de cristales y su expresión se llevó a cabo en células Hi5, aunque la expresión fuese parecida en las células Sf9. De este análisis de expresión se pudo concluir también que la parte que se degrada de la proteína se encuentra en el N-terminal y que la buena expresión de esta proteína disminuye mucho conforme se elimina más secuencia del C-terminal.

Además, puesto que estas construcciones se realizaron con el fin de producir cristales, era necesario usar kifunensina en la producción para poder realizar el corte de los N-glicanos con Endo-H y obtener una muestra homogénea y con N-glicanos simples conteniendo solo GlcNAc para facilitar la obtención de cristales. En la Figura 5.3 se puede observar el cambio de tamaño en la construcción inicial (T43-Q388), a causa de una menor N-glicosilación, entre la muestra sin y con kifunensina debido a la inhibición de la manosidasa de clase I (Elbein et al., 1990).

5.3. Diseño racional de glicopéptidos específicos para C1GalT1.

Para evaluar la actividad catalítica de *DmC1GalT1*, se diseñaron una serie de glicopéptidos (denominados P1 a P7) basándose en el estudio realizado por (Perrine et al., 2009)(Tabla 5.1). Estos glicopéptidos, de 7 aminoácidos de tamaño y un residuo de Thr glicosilado con GalNAc en la posición 0, contenían un residuo de Gly en la posición +1 y bien un residuo de Phe o de Tyr en la posición +3, residuos que claramente mejoraban la actividad de la enzima humana. Se podía observar según los experimentos llevado a cabo por Perrine que los residuos de Tyr, Phe y Pro en la posición -3 aumentaban ligeramente la actividad de la T-sintasa humana. En P6 y P7 se añadieron un residuo de Glu y de Asp en la posición -1, respectivamente, para comparar la adición de un residuo ácido comparándolo con P1 (Ala -1). Las demás posiciones se encontraban ocupadas por residuos de Pro en +2 y Ala en -2 debido a que ambos residuos parecían afectar positivamente a la

actividad enzimática (Perrine et al., 2009).

Tabla 5.1. Secuencia aminoacídica de la serie de glicopéptidos diseñados racionalmente. El diseño se basa en el estudio realizado por (Perrine et al., 2009).

	-3	-2	-1	0	+1	+2	+3
P1	Y	A	A	T*	G	P	F
P2	Y	A	A	T*	G	P	Y
P3	P	A	A	T*	G	P	F
P4	P	A	A	T*	G	P	Y
P5	F	A	A	T*	G	P	F
P6	Y	A	E	T*	G	P	F
P7	Y	A	D	T*	G	P	F

Además, se incluyeron el péptido APDTRP y los glicopéptidos APDT*RP y APDS*RP, donde * representa un residuo de GalNAc unido al aminoácido subyacente, ya sea Thr o Ser. APDTRP se considera una secuencia aminoacídica que hace las veces de epítipo inmunogénico de las repeticiones en tándem de MUC1. En particular, el grupo del Dr. Hurtado-Guerrero en colaboración con el grupo del Dr. Corzana resolvieron hace unos años la estructura del anticuerpo SM3 en complejo con APDT*RP (Martínez-Sáez et al., 2015), siendo esta la base en la que se sustenta el desarrollo de varias vacunas contra el cáncer (Martínez-Sáez et al., 2017). Además, es un sustrato natural de C1GalT1 en el entorno de MUC1 (Martínez-Sáez et al., 2015; Nason et al., 2021). También se usó el glicopéptido APDS*RP para confirmar si la actividad de C1GalT1 era mejor con un glicopéptido que contuviese un residuo de α -GalNAc-Thr frente a uno con uno con un residuo de α -GalNAc-Ser, tal y como se había señalado con anterioridad (Granovsky et al., 1994).

5.4. Cinéticas de *DmC1GalT1* frente a diferentes sustratos.

Para la realización de los experimentos biofísicos usando *DmC1GalT1*, se usó la construcción sin secuencia señal ni dominio transmembrana secretada en células de mamíferos HEK293F (residuos T43-Q388) (Apartado 4.3.1.). Para iniciar los estudios cinéticos, se establecieron las condiciones experimentales usando *DmC1GalT1* con UDP-Gal y APDT*RP (Figura 5.4, 5.5 y Tabla 5.2).

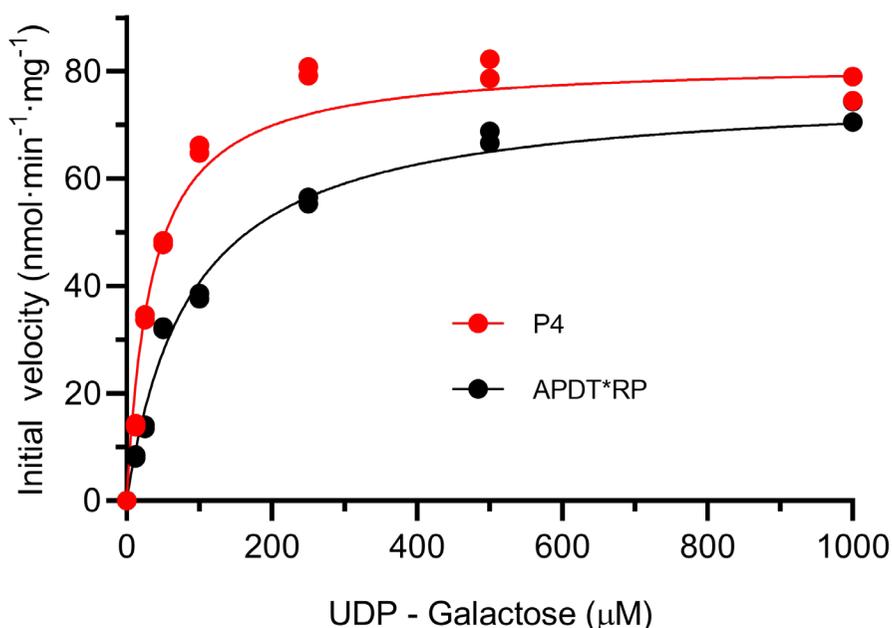


Figura 5.4. Cinéticas de *DmC1GalT1*^{T43-Q388} con concentraciones variables de UDP-Gal. Se usaron concentraciones fijas de APDT*RP (1000 μM) y de P4 (250 μM). Las velocidades iniciales se obtuvieron por duplicado (n=2 de ensayos independientes) para cada concentración de sustrato.

En primer lugar, se realizaron pruebas para determinar la concentración de enzima óptima a usar en los ensayos, variando la concentración de enzima y manteniendo los sustratos constantes. Al realizar dicho ensayo, se obtuvo que la concentración óptima de enzima era 500 nM debido a que la diferencia de señal entre los blancos y las muestras con más enzima se mantenía constante.

Además, se realizaron reacciones en presencia y ausencia de BSA, usándolo como *crowder* para observar si aumentaba la actividad enzimática. Como resultado se obtuvo que la actividad de *DmC1GalT1*^{T43-Q388} junto con BSA mejoró entre 2 y 3 veces en comparación a la reacción en ausencia de BSA.

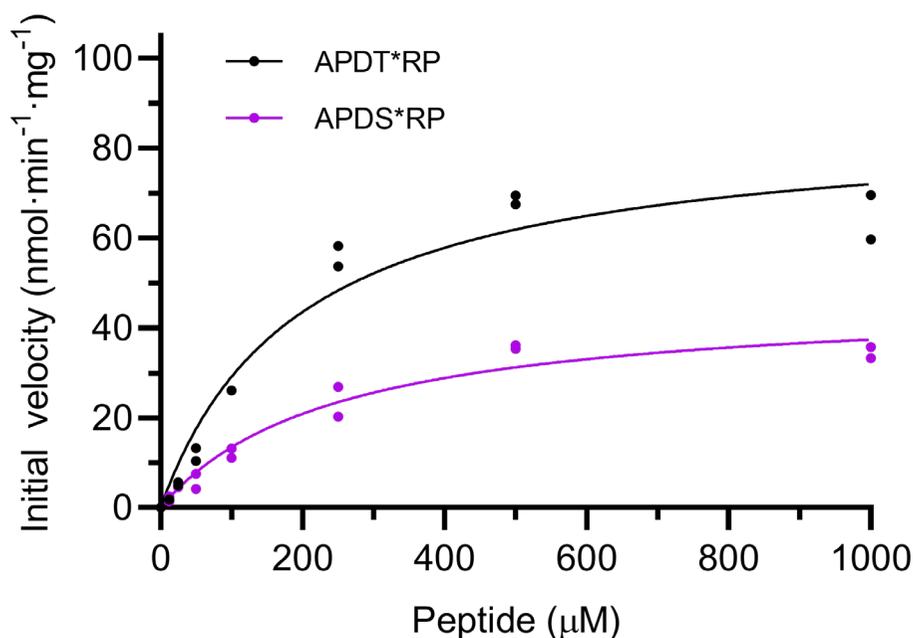


Figura 5.5. Cinéticas de *DmC1GalT1*^{T43-Q388} con concentraciones variables de APDT*RP y APDS*RP. Las concentraciones de los glicopéptidos variaron desde 12,5 μM hasta 1 mM; usando 500 μM de UDP-Gal. Las velocidades iniciales se obtuvieron por duplicado (n=2 de ensayos independientes) para cada concentración de sustrato.

DmC1GalT1 mostraba un perfil hiperbólico bajo concentraciones variables tanto de UDP-Gal como de APDT*RP, lo cual puede observarse también en presencia del resto de glicopéptidos (Figura 5.4 y 5.6a). Las constantes de Michaelis-Menten aparentes (K_m^{app}) para UDP-Gal y APDT*RP eran 88 ± 8 y 195 ± 43 μM, respectivamente, y la constante catalítica aparente (k_{cat}^{app}) $\sim 3,5$ min⁻¹ en ambos casos (Figura 5.6b y Tabla 5.2), un valor consistente con otros bajos valores de k_{cat}^{app} previamente registrados en otras “follow-up” glicosiltransferasas como

POMGnT1/POMGnT2, donde la k_{cat}^{app} variaba de 7 a 1020 min^{-1} dependiendo de la secuencia del glicopéptido (Akasaka-Manya et al., 2011; Clement et al., 2008).

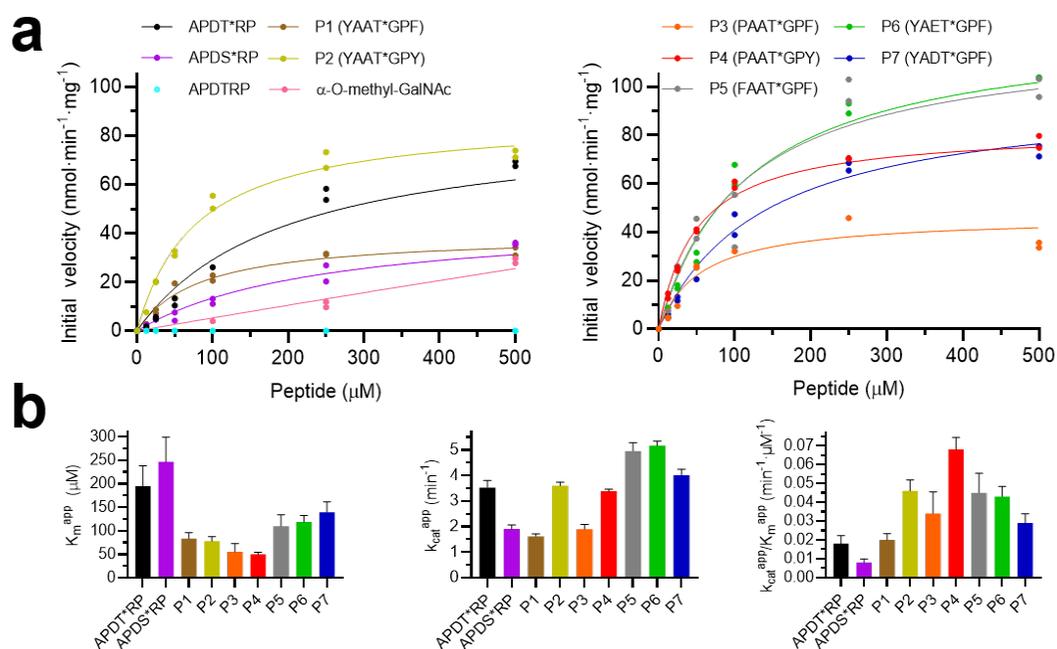


Figura 5.6. Ensayos cinéticos de *DmC1GalT1*^{T43-Q388} con varios (glico)péptidos y α -O-metil-GalNAc. a. Cinéticas de glicosilación de *DmC1GalT1*^{T43-Q388} frente a los diferentes glicopéptidos específicos diseñados, APDT*RP, APDS*RP, el péptido APDTRP y α -O-metil-GalNAc. **b.** Gráficos comparando la K_m^{app} , k_{cat}^{app} y eficiencia catalítica (k_{cat}^{app}/K_m^{app}) entre los diferentes sustratos. Las velocidades iniciales se obtuvieron por duplicado ($n=2$ de ensayos independientes) para cada concentración de sustrato. Las barras de error representan el error estándar calculado mediante los ajustes obtenidos con GraphPad Prism. Los valores exactos de los parámetros obtenidos se encuentran en la Tabla 5.2.

Tabla 5.2. Parámetros cinéticos de los (glico) péptidos usados en esta tesis doctoral tanto con *DmC1GalT1*^{T43-Q388} como con *DmC1GalT1*^{S73-Q388}. Los valores de error de la tabla representan el error estándar calculado mediante los ajustes obtenidos con GraphPad Prism.

	K_m^{app} (μM)	R^2	V_{max}^{app} ($nmol \cdot min^{-1} \cdot mg^{-1}$)	k_{cat}^{app} (min^{-1})	k_{cat}^{app}/K_m^{app} ($min^{-1} \cdot \mu M^{-1}$)
UDP-Gal ^{APDT*RP}	88,27 ± 7,86	0,99	76,45 ± 1,7	3,14 ± 0,07	0,036 ± 0,003
^{oo} UDP-Gal ^{P4}	35,33 ± 5,03	0,96	81,86 ± 2,52	3,36 ± 0,1	0,095 ± 0,014
APDTRP	^	^	^	^	^
APDT*RP	195 ± 43,34	0,96	86,02 ± 6,71	3,53 ± 0,27	0,018 ± 0,004
APDS*RP	246,2 ± 52,59	0,96	46,61 ± 3,72	1,91 ± 0,15	0,008 ± 0,002
P1 (YAAT*GPF)	83,02 ± 13,42	0,97	39,61 ± 2,19	1,62 ± 0,09	0,02 ± 0,003
P2 (YAAT*GPY)	78,17 ± 9,39	0,99	87,69 ± 3,49	3,6 ± 0,14	0,046 ± 0,006
P3 (PAAT*GPF)	55,14 ± 17,51	0,9	46,38 ± 4,45	1,9 ± 0,18	0,034 ± 0,011
P4 (PAAT*GPY)	50,06 ± 4,56	0,99	82,54 ± 1,93	3,39 ± 0,08	0,068 ± 0,006
P5 (FAAT*GPF)	110 ± 24,16	0,95	120,9 ± 7,86	4,96 ± 0,32	0,045 ± 0,01
P6 (YAET*GPF)	118,8 ± 13,97	0,98	125,9 ± 4,57	5,16 ± 0,19	0,017 ± 0,004
P7 (YADT*GPF)	139,6 ± 21,8	0,98	97,79 ± 5,8	4,01 ± 0,24	0,043 ± 0,005
APDT*RP ^{DmC1GalT1 (S73-Q388)}	322,6 ± 88,09	0,95	126,7 ± 14,16	4,71 ± 0,58	0,029 ± 0,005

^inactivo

Tal y como se esperaba, *DmC1GalT1* era incapaz de realizar cualquier transferencia en el péptido no glicosilado APDTRP y su actividad se veía ligeramente reducida frente al glicopéptido con Ser (APDS*RP). En concreto, la K_m^{app} , k_{cat}^{app} y eficiencia catalítica con el glicopéptido APDS*RP eran 1,26, 1,85 y 2,25 veces peor que las obtenidas en presencia del APDT*RP (Figura 5.4), resultado que concuerda con los descubiertos en los ensayos realizados con la enzima de rata (Granovsky et al., 1994) y, sugiere que el grupo metilo de Thr es probablemente relevante para la obtención de parámetros cinéticos ligeramente mejores. También fue determinado si el residuo de GalNAc se comportaba como un sustrato aceptor. Se usó para ello como sustrato α -O-metil-GalNAc, que terminó siendo mucho peor sustrato que APDT*RP no pudiendo los parámetros cinéticos ser calculados al no conseguir la

saturación de la enzima, un hallazgo en sintonía con conclusiones previas en las que se usaron análogos similares frente a glicopéptidos (R. Müller et al., 2005). Bajo condiciones saturantes de glicopéptido ($\sim 500 \mu\text{M}$), la actividad con α -O-metil-GalNAc era aproximadamente la mitad obtenida con APDT*RP y no se conseguía llegar a la saturación enzimática ni con 2 mM de sustrato (Figura 5.6a y 5.7). Con estos datos se muestra que el ambiente peptídico alrededor del residuo de azúcar es clave para la obtención de parámetros cinéticos óptimos y que, además, el residuo de GalNAc solo no es suficiente para conseguirlos.

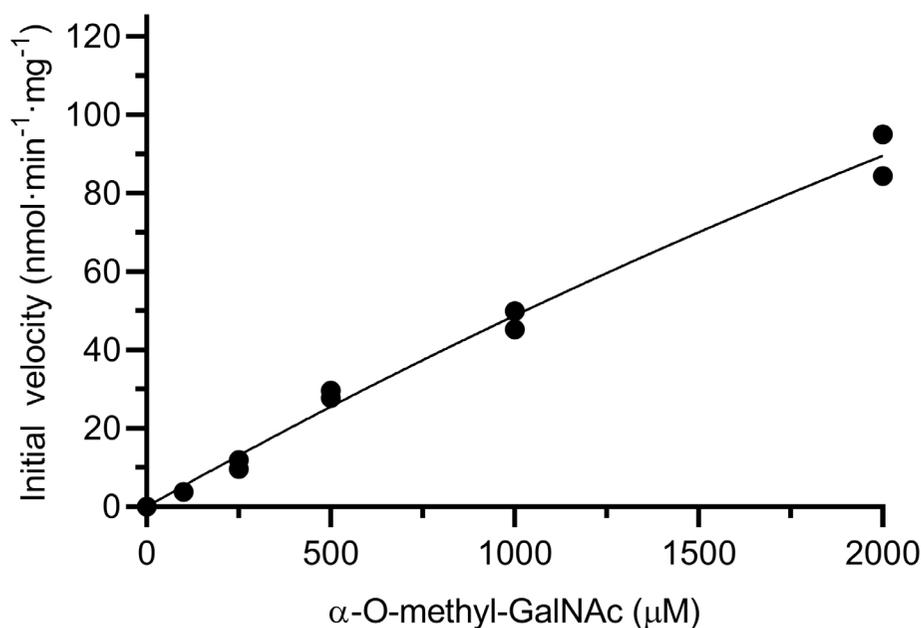


Figura 5.7. Cinéticas de *DmC1GalT1*^{T43-Q388} con concentraciones variables de α -O-metil-GalNAc. La concentración se varió desde 12,5 μM hasta 2 mM usando 500 μM de UDP-Gal. Las velocidades iniciales se obtuvieron por duplicado ($n=2$ de ensayos independientes) para cada concentración de sustrato.

Con respecto a las cinéticas con los glicopéptidos P1 – P7 (Figura 5.6a), en primer lugar, se determinaron los parámetros cinéticos de UDP-Gal bajo concentraciones saturantes de P4, obteniendo una ligera mejoría en la K_m^{app} y casi la misma $k_{\text{cat}}^{\text{app}}$ comparándolo con los parámetros obtenidos de UDP-Gal en presencia de

APDT*RP (Figura 5.4 y Tabla 5.2).

Las K_{ms}^{app} de los glicopéptidos P1 – P7 eran todas levemente mejores que la obtenida con el glicopéptido nativo APDT*RP, con mejoras de 1,3 a 4 veces, siendo el P4 el que mejor K_m^{app} presentaba (Figura 5.6b). Estos datos sugieren también que la Pro en la posición –3 es mejor para la catálisis que el residuo de Tyr en la misma posición (P2 frente a P4) y que el Glu o el Asp en la posición –2 son levemente peores para la transferencia del GalNAc que una Ala en la misma posición (P6/P7 frente a P1). El rango de valores de los parámetros de k_{cat}^{app} es algo más restrictivo, siendo el P1 el sustrato más lento y el P4 el más rápido (Figura 5.6b). Con respecto a la eficiencia catalítica, los valores calculados eran menos restrictivos que los obtenidos en la k_{cat}^{app} , sugiriendo que para la serie de glicopéptidos que contenían como aceptor glicosilado un residuo de Thr, el mejor sustrato era el P4 y los peores el P1 y APDT*RP (Figura 5.6b).

En general, los datos muestran que las diferencias en los parámetros cinéticos entre glicopéptidos son pequeñas y que el residuo de GalNAc por sí solo no es suficiente para obtener parámetros cinéticos óptimos, siendo la secuencia peptídica crucial para conseguirlos, Todo esto sugiere que C1GalT1 podría interactuar directamente con el péptido de los sustratos aceptores. Es importante destacar que no se pudo alcanzar la saturación enzimática con el sustrato α -O-metil-GalNAc y que ésta solo se consiguió con los diferentes glicopéptidos como sustratos.

5.5. Experimentos de STD-RMN.

Tras la realización de los ensayos cinéticos y para comprender la manera en la que interaccionan los glicopéptidos con *DmC1GalT1*, se realizaron experimentos de STD-RMN. Estos ensayos se basan en observar cómo el ligando interacciona con la proteína, analizando las señales de asignación de los ^1H -RMN sustentándose en la transferencia de saturación mediante el efecto nuclear Overhauser, desde la

resonancia del protón del receptor (proteína o enzima en este caso) a los protones del ligando (azúcar o glicopéptido en el caso actual) cambiando entre el estado de proteína libre al estado en el que el ligando y la proteína están unidos (Mayer & Meyer, 1999). El análisis de las respuestas de STD permite deducir qué átomos del ligando se encuentran en contacto más cercano con el receptor y, de esa forma, mapear los epítodos de unión de ligandos en complejo con grandes moléculas (Mayer & Meyer, 2001). Para estos experimentos se seleccionaron los siguientes sustratos: α -O-metil-GalNAc, el péptido APDTRP sin glicosilar, uno de los peores sustratos (APDT*RP) y uno de los mejores sustratos (P4).

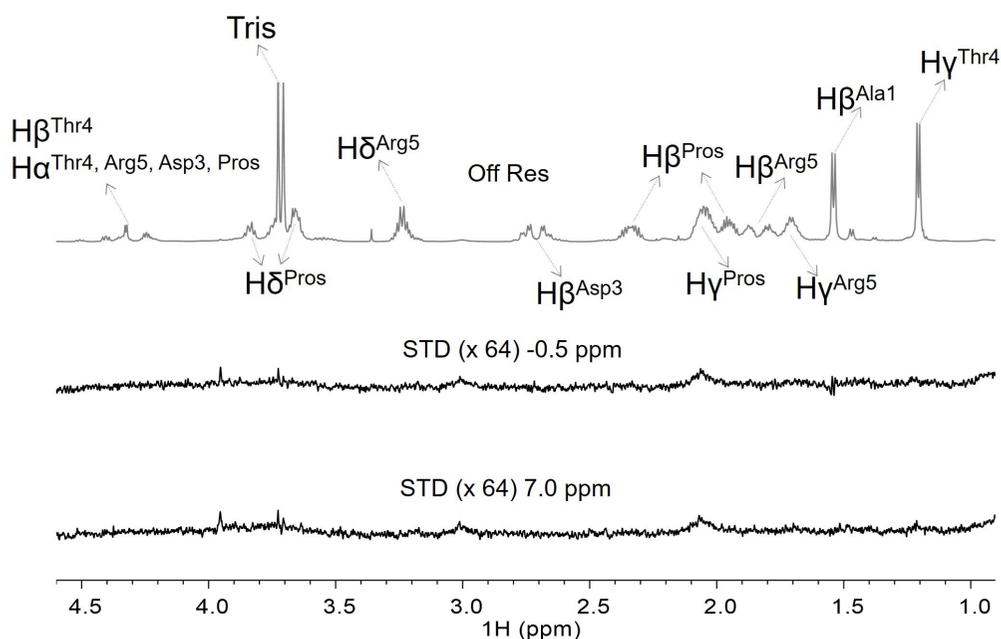


Figura 5.8. Espectro STD-RMN de APDTRP. Fue obtenido a 600 MHz y 298 K con APDTRP a 730 μ M en presencia de 20 μ M *DmC1GalT1*^{T43-Q388}, 135 μ M UDP y 150 μ M MnCl₂. El espectro *off-resonance* (etiquetado como Off Res) se muestra en gris y el espectro STD (etiquetado STD) obtenido mediante la saturación de la proteína a -0,5 ppm y 7,0 ppm se muestra en negro. No se detectó señal de STD-RMN.

En el caso del α -O-metil-GalNAc, el péptido no glicosilado y el APDT*RP, se usaron dos frecuencias de irradiación diferentes para alcanzar la saturación de la proteína, en la región alifática (-0,5 ppm) y la aromática (7 ppm). Sin embargo,

para el P4 (debido a la presencia del residuo aromático de Tyr) solo se usó la frecuencia de irradiación alifática.

Usando un exceso de aproximadamente 6 y 7 veces de UDP y MnCl₂ con respecto a la enzima, se observó que, aunque el péptido no glicosilado no devolvía ningún tipo de respuesta de STD (Figura 5.8 y Tabla 5.3), el α -O-metil-GalNAc sí que presentaba diferencias con respecto al espectro sin saturar, indicando la importancia del residuo de GalNAc en el reconocimiento por parte de C1GalT1 (Figura 5.9a y Tabla 5.3) A continuación, se realizaron los experimentos STD-RMN de los glicopéptidos APDT*RP y P4 en ausencia y presencia de UDP y MnCl₂. Estos experimentos en ausencia del nucleótido fueron realizados porque previamente se habían reportado casos en los que otras glicosiltransferasas, bastante lejanas evolutivamente a C1GalT1 pero con el mismo tipo de plegamiento GT-A, como la GalNAc-T2 y las proteínas efectoras NleB/SseK, necesitan la presencia de UDP para que pueda llevarse a cabo la unión al péptido o proteína aceptora, implicando la existencia de un mecanismo de ajuste inducido (de las Rivas, Paul Daniel, et al., 2018; García-García, Hicks, et al., 2021; Lira-Navarrete et al., 2014). Por ejemplo, el caso de GalNAc-T2 es muy interesante porque su conformación activa está caracterizada por un cambio en el loop flexible de una conformación abierta a una cerrada, la cual se consigue de manera completa en presencia de UDP-GalNAc, pero de manera intermedia en presencia de UDP (de las Rivas, Coelho, et al., 2018). Recientemente, se descubrió que FUT8, una glicosiltransferasa con el tipo de plegamiento GT-B, muestra propiedades similares a las otras dos proteínas mencionadas, con la particularidad que se une mejor a su sustrato aceptador (*N*-glicanos) en presencia de GDP, pero sin ser esencial la unión de GDP para que tenga lugar la unión del *N*-glicano a FUT8 (García-García et al., 2020; García-García, Serna, et al., 2021).

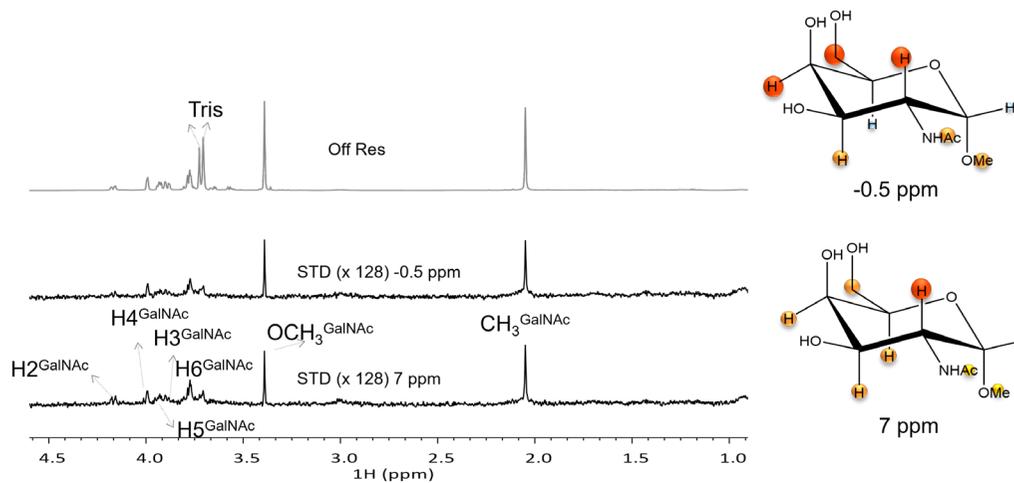
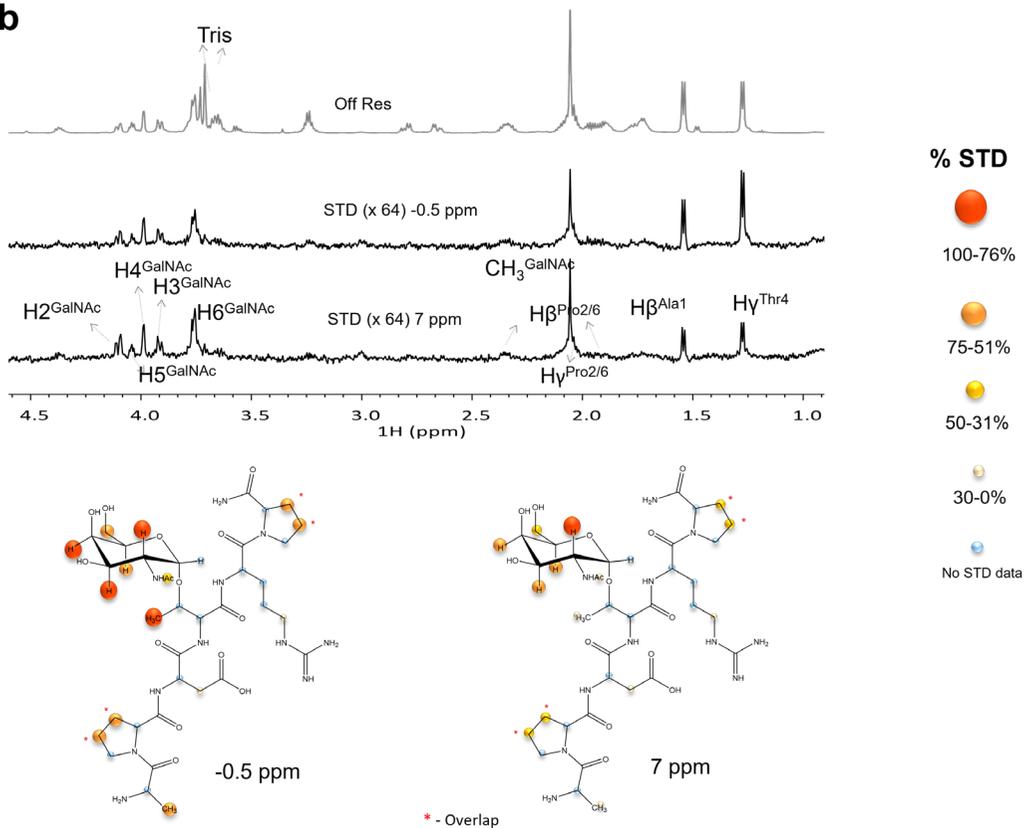
a**b**

Figura 5.9. Espectro STD-RMN de α -O-methyl-GalNAc (a) y de APDT*RP (b). Fueron obtenidos a 600 MHz y 298 K. **a.** Espectro STD de α -O-methyl-GalNAc a 745 μ M en presencia de 20 μ M *DmC1GalT1*^{T43-Q388}, 135 μ M UDP y 150 μ M MnCl₂. **b.** Espectro STD de APDT*RP a 710 μ M en presencia de 20 μ M *DmC1GalT1*^{T43-Q388}, 135 μ M UDP y 150 μ M MnCl₂. Los protones que no pudieron analizarse de manera precisa en los espectros STD se mapearon en azul. Específicamente, la señal del protón anomérico H₁ del residuo de GalNAc no se podía analizar a causa de su cercanía con la resonancia de HDO ni tampoco las de los H α de los aminoácidos por el solapamiento con los protones del azúcar. El espectro no saturado (*off-resonance*, etiquetado como Off Res) se muestra en gris y el espectro STD (etiquetado STD) obtenido mediante la saturación de la proteína a -0,5 ppm y 7,0 ppm se muestra en negro. Las resonancias de los protones claves se encuentran señaladas en los espectros de STD (paneles inferiores). En la imagen también se muestran los mapas de epítipo para los dos ligandos tras ser estimados. Las resonancias de protones que se solapan en los espectros han sido identificadas y marcadas en la figura con un * de color rojo. El modo de unión es similar en ambos ligandos según estos resultados.

Tanto NleB/SseK como FUT8 contienen también loops flexibles que se ordenan en presencia de los nucleótidos tal y como se describió para GalNAc-T2, implicando que el sitio activo adopta una conformación activa una vez estos loops flexibles se unen al nucleótido azucarado (de las Rivas, Coelho, et al., 2018; García-García, Hicks, et al., 2021; García-García, Serna, et al., 2021; Lira-Navarrete et al., 2014). En resumen, se ha propuesto que estas enzimas necesitan que se una el nucleótido azucarado en primer lugar, ejerciendo algún tipo de cambio conformacional, para que luego pueda unirse el sustrato aceptor (aunque en el caso de FUT8 la previa unión a GDP no es esencial para que se una a un *N*-glicano) y en consecuencia para que tenga lugar la glicosilación.

Tabla 5.3. Tabla de asignación de los picos del espectro RMN-H¹. Se muestran las asignaciones de los picos de RMN-H¹ de α -O-methyl-GalNAc, APDT*RP y P4 los factores de amplificación de STD y el % relativo de STD para cada protón con respecto al protón con mayor señal (ya sea a -0,5 ppm o 7 ppm). El H₁ anomérico y el H_a de los aminoácidos no pudieron ser analizados en el espectro de STD debido a su cercana distancia a la resonancia de HDO y a la supresión del agua. La relación entre señal y ruido no permitía medir con precisión la respuesta STD de algunos protones de aminoácidos. En el caso de APDT*RP, algunos protones de Pro2 y Arg1, y en el caso de P4 los protones de Pro6.

	Asignación de pico	Ppm	STD _{AF} (-0,5 ppm)	% STD (-0,5 ppm)	STD _{AF} (7 ppm)	%STD (7 ppm)
α-O-methyl-GalNAc	H ₂ GalNAc	4,169	30,7	100%	45,0	100%
	H ₃ GalNAc	3,891	20,0	65%	24,9	55%
	H ₄ GalNAc	3,993	28,7	93%	27,9	62%
	H ₅ -GalNAc	3,932	-	-	25,8	57%
	H ₆ -GalNAc	3,776	23,6	77%	29,5	66%
	CH ₃ (NHAc) GalNAc	2,049	18,9	61%	20,0	44%
	OMe	3,391	17,6	57%	16,6	37%
	Asignación de pico	Ppm	STD _{AF} (-0,5 ppm)	% STD (-0,5 ppm)	STD _{AF} (7 ppm)	%STD (7 ppm)
APDT*RP	H ₂ GalNAc	4,104	44,8	100%	67,1	100%
	H ₃ GalNAc	3,915	35,5	79%	41,1	61%
	H ₄ GalNAc	3,990	35,5	79%	42,7	64%
	H ₅ -GalNAc	4,043	30,5	68%	35,2	53%
	H ₆ -GalNAc	3,764	27,1	60%	35,8	53%
	CH ₃ (NHAc) GalNAc	2,056	15,6	35%	20,5	27%
	H _{β} -A1	1,544	25,2	56%	16,7	25%
	H _{β3} -P's	2,342	22,7	51%	24,4	36%
	H _{β2} -P's	1,949	21,6	48%	-	-
	H _{γ} -P's	2,070-2,007	22,7	51%	23,7	35%
	H _{γ} -T4	1,276	39,8	89%	18,6	28%
	H _{δ} -R5	3,244	9,1	20%	10,3	15%
	Asignación de pico	Ppm	STD _{AF} (-0,5 ppm)	% STD (Con UDP)	STD _{AF} (-0,5 ppm)	%STD (Sin UDP)
P	H ₂ GalNAc	4,106	63,5	99%	45,1	88%

H₃ GalNAc	3,925	59,3	93%	40,9	80%
H₄ GalNAc	3,993	64,1	100%	45,5	89%
H₅-GalNAc	4,060	52,7	82%	43,4	84%
H₆-GalNAc	3,770	40,9	64%	30,3	59%
CH₃ (NHAc) GalNAc	2,029	29,1	46%	30,5	59%
H₇-P1/H_{β2}P1	2,096-2,067	61,8	96%	39,9	78%
H_{β3}-P1	2,478	56,6	88%	37,7	73%
H_δ-P1	3,419	35,0	55%	21,7	42%
H_β-A2	1,427	29,8	47%	18,1	35%
H_β-A3	1,458	37,7	59%	22,4	44%
H_γ-T4	1,293	50,9	80%	44,5	86%
H_{β3}-Y7	3,157	41,6	65%	34,8	68%
H_ε-Y7	6,873	59,9	93%	51,5	100%
H_δ-Y7	7,176	40,5	77%	40,5	79%

En el caso específico de *DmC1GalT1* y en ausencia de UDP, solo el P4 mostraba claras señales de STD-RMN. Aun así, ambos glicopéptidos mostraban respuesta clara de STD en presencia de UDP. Estos resultados sugieren la existencia de diferencias en el reconocimiento de estos dos glicopéptidos y que, en ambos casos, el residuo de GalNAc presentaba una elevada transferencia de saturación indicando que debe estar en mayor contacto con la enzima. Los mapas de epítipo de GalNAc de los STDs para los glicopéptidos y el α -O-metil-GalNAc eran bastante parecidos, significando una orientación de unión similar de la unidad de GalNAc en todas las estructuras (Figura 5.9, Figura 5.10 y Tabla 5.3). Sin embargo, el factor de amplificación fue menor en el caso del α -O-metil-GalNAc que en el de los glicopéptidos (Tabla 5.3), lo que reflejando la esperada menor afinidad de unión del α -O-metil-GalNAc frente a APDT*RP y P4, deducida de los experimentos cinéticos. De hecho, al nivel de la secuencia peptídica, el grupo metilo de la Thr mostraba una clara respuesta STD en todos los casos, mientras que los resultados eran bastante diversos para el resto de aminoácidos de ambos glicopéptidos.

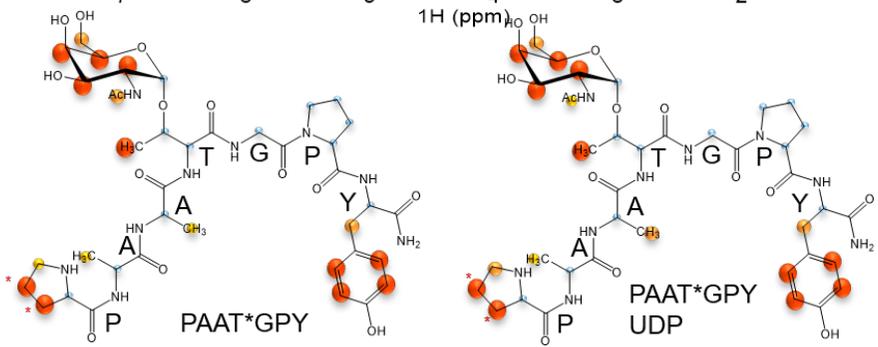
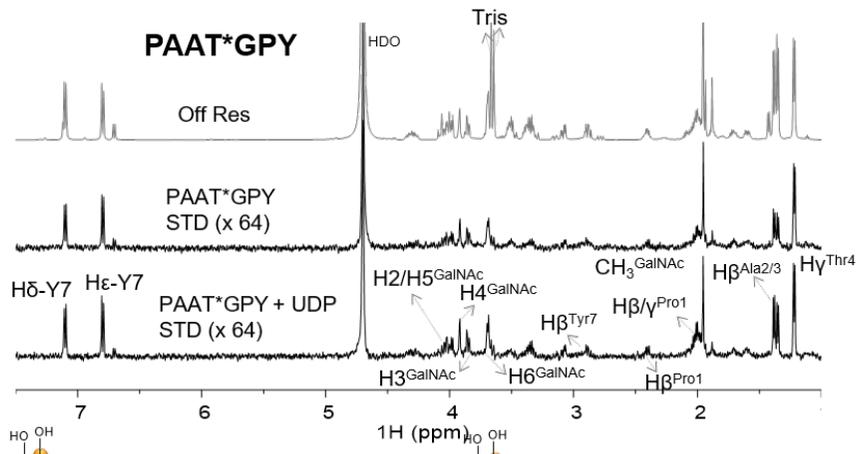
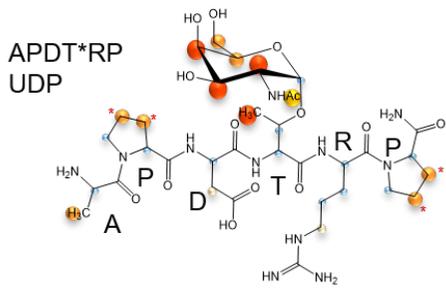
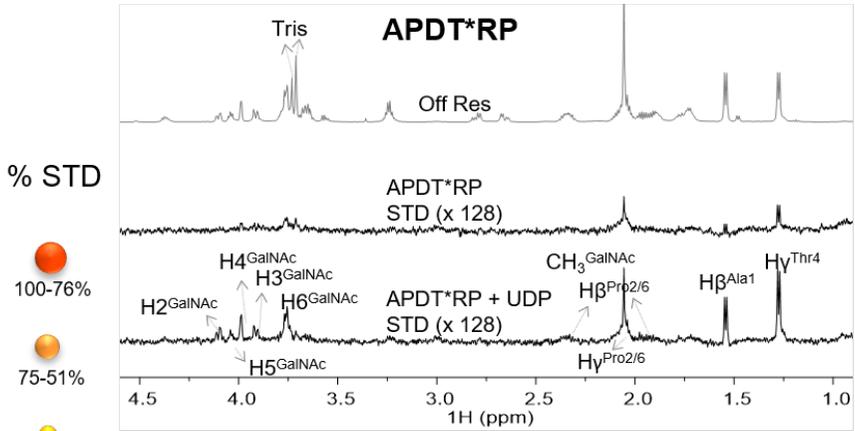


Figura 5.10. Espectros STD–RMN y mapas de epítipo de los glicopéptidos APDT*RP y P4 en presencia y ausencia de UDP. Los espectros fueron obtenidos a 600 MHz y 298 K. En ambos caso se muestra el espectro *off-resonance* (Off Res) en gris en la parte superior de los espectros, los espectros STD se obtuvieron saturando la proteína a $-0,5$ ppm y se muestran en negro. Arriba en el panel se encuentra el espectro STD de APDT*RP a $710 \mu\text{M}$ en presencia de $20 \mu\text{M}$ *DmC1GalT1*^{T43-Q388} y $150 \mu\text{M}$ MnCl_2 en ambos espectros saturados y con $135 \mu\text{M}$ UDP en el espectro inferior. En la parte inferior de la imagen se muestra el espectro STD de P4 a $710 \mu\text{M}$ en presencia de $20 \mu\text{M}$ *DmC1GalT1*^{T43-Q388} y $150 \mu\text{M}$ MnCl_2 en ambos espectros saturados y con $135 \mu\text{M}$ UDP en el espectro inferior. Bajo los espectros se encuentran los mapas de epítipo para cada ligandos, en el caso de P4 tanto en presencia como en ausencia de UDP. Los protones que no pudieron analizarse de manera precisa se muestran en azul. Las señales tanto del protón anomérico H_1 del residuo de GalNAc como de los H_α de los aminoácidos no pudieron analizarse debido a la cercanía con las resonancias de HDO y de los protones del azúcar respectivamente. Las diferentes esferas coloreadas indican la señal STD normalizada (%) para cada protón. Las resonancias de protones que solapan en los espectros fueron identificadas y marcadas con * de color rojo.

En el caso del APDT*RP, se detectaron ligeras mejoras de la respuesta de STD para el metilo de Ala1 y algunos protones de Pro2 y Pro6 superpuestos en el espectro. No se detectó respuesta para los protones pertenecientes a Asp3 (Figura 5.10). Curiosamente en el caso del P4, se observó una respuesta significativa procedente de las cadenas laterales de Pro1 y Tyr7, tanto en ausencia como en presencia de UDP (Figura 5.10), sugiriendo que estos aminoácidos deberían encontrarse en contacto cercano con *DmC1GalT1*. Estas interacciones adicionales podrían explicar las diferencias en las K_m^{app} entre ambos glicopéptidos, siendo cuatro veces mejor la del P4 que la del APDT*RP. En general, estos datos sugieren que la unión del residuo de GalNAc es lo más importante de los glicopéptidos a la hora del reconocimiento de los mismos y que la unión óptima se alcanza en presencia tanto de la unidad de GalNAc como del péptido.

5.6. *DmC1GalT1* no presenta un mecanismo con características alostéricas para la unión a los glicopéptidos.

Con el objetivo de contrastar el comportamiento diferente que mostraban los glicopéptidos con la enzima en ausencia y presencia de UDP, se decidió realizar experimentos de calorimetría de titulación isotérmica (ITC). En primer lugar, se determinó la constante de disociación (K_d) de la unión de UDP a *DmC1GalT1* en presencia de $MnCl_2$ (K_d $18,39 \pm 4,67 \mu M$) (Figura 5.11 y Tabla 5.4). Como se esperaba, el péptido APDTRP no se unía a la proteína bajo exceso de UDP (Figura 5.11).

Tabla 5.4. Parámetros termodinámicos de la unión de los ligandos a *DmC1GalT1*^{T43-Q388}. La K_d representa la constante de disociación ($=1/K_a$), y ΔG , ΔH y $-T\Delta S$ son los parámetros termodinámicos. En la mayoría de casos la estequiometría de unión es cercana a $\sim 1:1$. Los coeficientes de correlación de los ajustes (R^2) de ITCs eran $< 0,99$. Los valores de error representan el error calculado a través del ajuste por iteración de los grupos de datos mediante Origin 7 (Microcal).

	K_d (μM)	ΔH (kcal/mol)	ΔG (kcal/mol)	$-T\Delta S$ (kcal/mol)	n
UDP	$18,39 \pm 4,67$	$-4,26 \pm 0,47$	-6,46	-2,19	0,40
APDT*RP (exceso UDP)	$173,5 \pm 50,79$	$-1,77 \pm 1,02$	-5,13	-3,36	0,88
APDTRP	^	^	^	^	^
APDT*RP	^	^	^	^	^
PAAT*GPY (exceso UDP)	$60,35 \pm 8,76$	$-4,29 \pm 0,87$	-5,75	-1,46	0,88
PAAT*GPY	$40,99 \pm 5,94$	$-1,88 \pm 0,24$	-5,98	-4,1	0,98
YADT*GPF (exceso UDP)	$59,99 \pm 7,01$	$-2,78 \pm 0,25$	-5,76	-2,98	1,50
YADT*GPF	$37,09 \pm 6,24$	$-6,2 \pm 1,69$	-6,04	0,16	0,650

^No cuantificable bajo nuestras condiciones.

Tras los experimentos iniciales, se evaluó si esta glicosiltransferasa requería de la unión de una molécula de UDP antes de poder unirse a los glicopéptidos. Aunque *DmC1GalT1* era capaz de unirse tanto a P4 como a APDT*RP en presencia de UDP, no ocurría lo mismo en ausencia de UDP, donde sólo se observaba unión a P4, tal y como mostraban los resultados de los experimentos de STD-RMN (Figura 5.12 y 5.13).

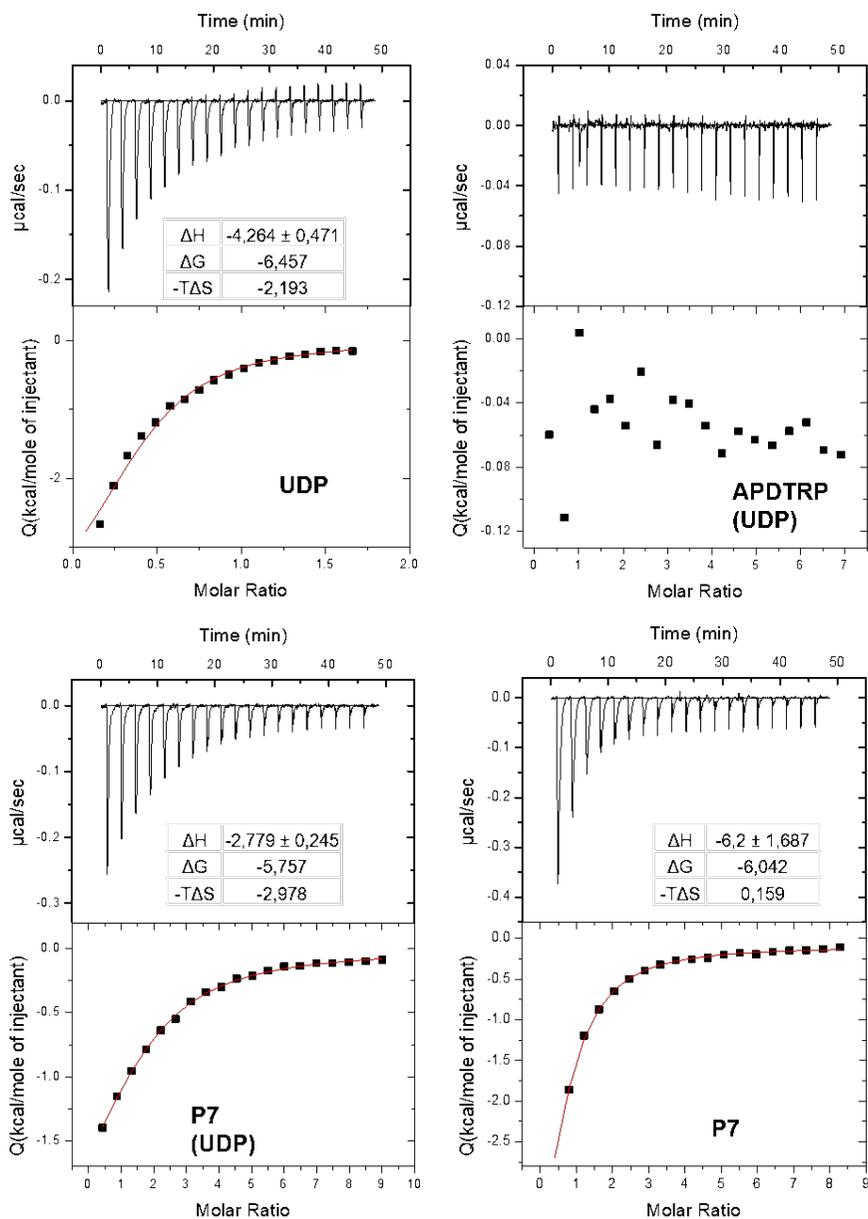


Figura 5.11. Datos de ITC de la unión entre UDP, APDTRP y P7 en presencia y ausencia de exceso de UDP a *DmC1GalT1*^{T43-Q388}. En la parte superior de los gráficos se encuentran los termogramas (potencia térmica frente a tiempo) y en la inferior las isothermas de unión (calor normalizado frente a “ratio molar”). Los ITCs de P7 se realizaron en presencia o ausencia de un exceso de UDP y el de APDTRP en presencia de exceso de UDP. Los experimentos se llevaron a cabo en duplicado (n = 2 experimento independientes).

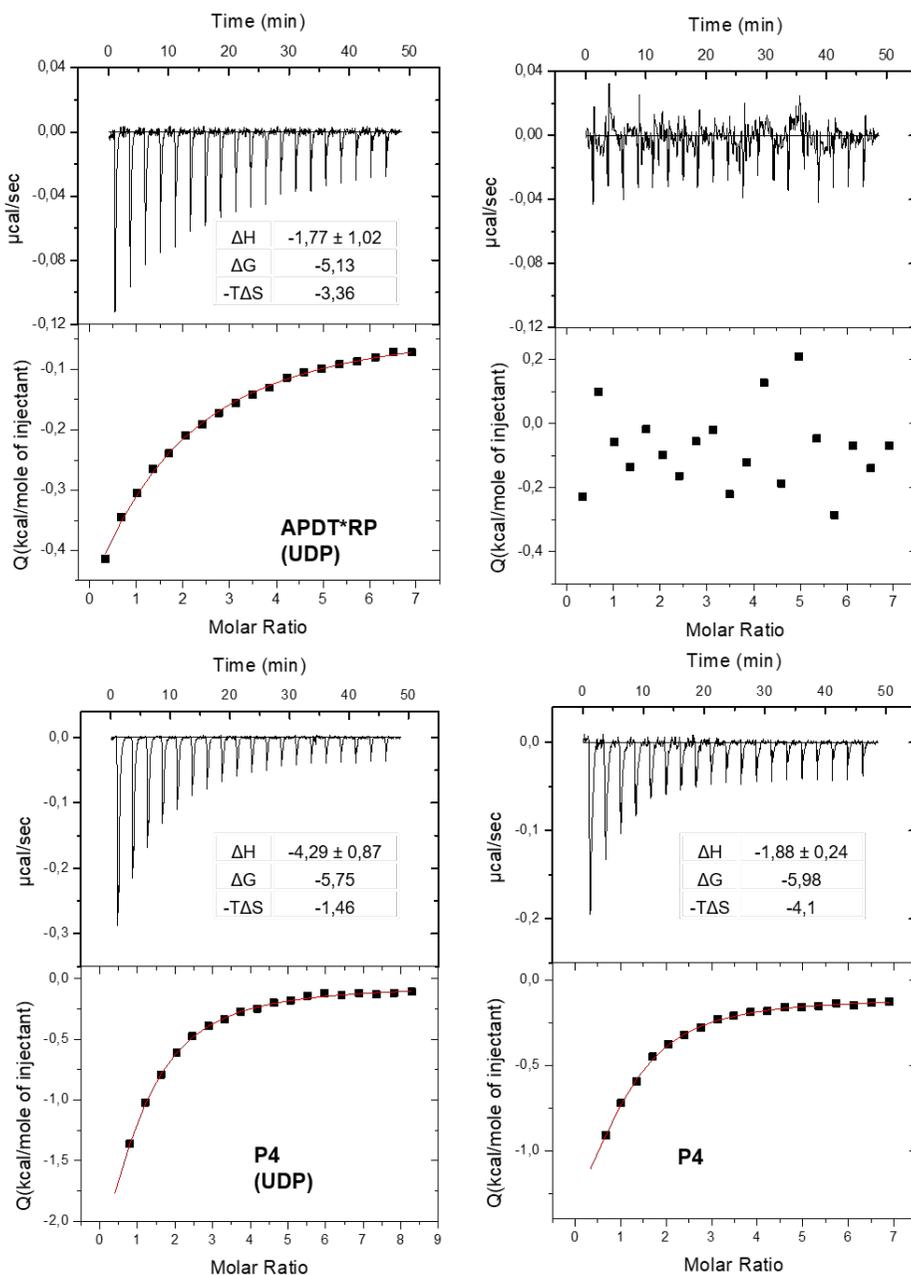


Figura 5.12. Datos de ITC de la unión entre APDT*RP y P4 a $DmC1GalT1^{T43-Q388}$. En la parte superior de los gráficos se encuentran los termogramas (potencia térmica frente a tiempo) y en la inferior las isothermas de unión (calor normalizado frente a ratio molar). Los ITCs se realizaron en presencia o ausencia de un exceso de UDP. Los experimentos se llevaron a cabo en duplicado (n = 2 experimento independientes).

Las K_{ds} de los glicopéptidos coincidían con las K_{ms}^{app} y con las diferencias entre ellas; la K_d de P4 era aproximadamente 3,5 veces mejor que la de APDT*RP. A causa de que el glicopéptido APDT*RP es bastante inusual por contener dos residuos cargados (Asp y Arg), surgió la duda de si esa podría ser la razón de su comportamiento en la ausencia de UDP. Para esclarecer esto, se realizó otro ensayo de ITC con el P7, el cual contiene un residuo cargado negativamente. P7 se comportaba de manera similar al P4 y se unía a la enzima indistintamente de la presencia o ausencia de UDP (Figura 5.11, 5.12 y 5.13 y Tabla 5.3), lo que sugiere que el residuo de Arg de APDT*RP o su propia conformación podrían ser los causantes de que no sea capaz de unirse en la ausencia de UDP.

En cuanto a los análisis de los parámetros termodinámicos de la interacción, eran bastante complejos y difíciles de interpretar en el caso de los glicopéptidos, impidiendo la obtención de una conclusión significativa (Tabla 5.4).

Estos resultados revelan que *DmC1GalT1* no sigue un mecanismo inducido como otras glicosiltransferasas como *NleB1* (García-García, Hicks, et al., 2021), *GalNAc-T2* (de las Rivas, Coelho, et al., 2018; Lira-Navarrete et al., 2014) y *FUT8* (García-García et al., 2020), y que por lo tanto no necesita unirse en primer lugar al nucleótido azucarado para unirse a los sustratos aceptores.

Cabe destacar que todos los termogramas realizados presentan un bajo parámetro “c” de Wiseman ($c < 1$), lo que explica la ausencia del plateau inicial de la curva. A pesar de esto, los datos son fiables pues el posible error en la estimación de la estequiometría (n) no afecta prácticamente al ajuste de K_a como se ha descrito previamente. El parámetro “c” de Wiseman es el producto de la concentración del receptor y de la constante de asociación, K_a (Tellinghuisen, 2008; Turnbull & Daranas, 2003).

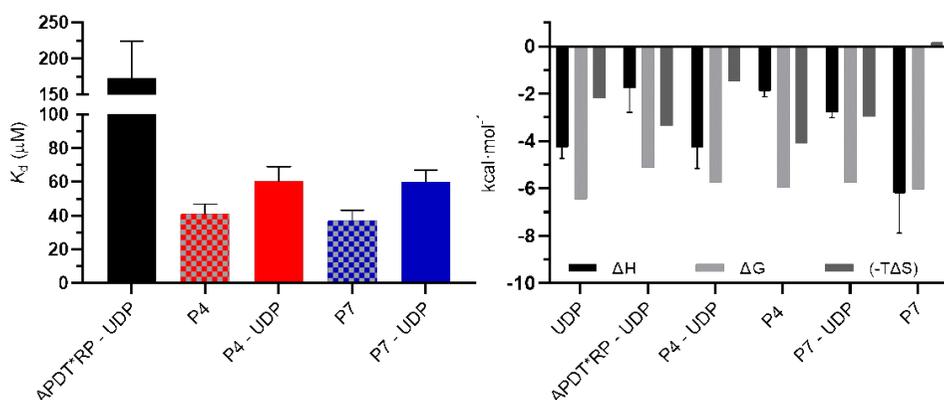


Figura 5.13. Parámetros termodinámicos de la unión a *DmC1GalT1*^{T43-Q388} de los glicopéptidos en presencia o ausencia de UDP. (Izquierda) Gráfica que muestra los valores de K_d s de los glicopéptidos. **(Derecha)** Gráfica que muestra los parámetros termodinámicos de los ajustes calculados en los que muestran el componente entálpico (ΔH), la energía libre de Gibbs (ΔG) y el componente entrópico en función de la temperatura ($-T\Delta S$) en $\text{kcal}\cdot\text{mol}^{-1}$. Los resultados con valor negativo muestran una contribución favorable a la unión, y los positivos una contribución desfavorable. Las barras de error representan el error calculado a través del ajuste por iteración de los grupos de datos mediante Origin 7 (Microcal).

5.7. Detalles del complejo formado por *DmC1GalT1*-UDP-APDT*RP.

Con el objetivo de proporcionar detalles a nivel atómico de la estructura de *DmC1GalT1* y su interacción con UDP-Gal/UDP y los glicopéptidos, se trabajó con una versión truncada de *DmC1GalT1* (residuos S73-Q388) (Resultados 5) secretada de células de insecto High FiveTM (Hi5) (Métodos 4.3.3).

Para asegurar la similitud y actividad de este constructo en comparación con el expresado en mamíferos, y usado en los experimentos mencionados anteriormente, se realizaron ensayos cinéticos que demostraron que los parámetros cinéticos de esta enzima eran muy parecidos a aquellos obtenidos con el constructo más largo *DmC1GalT1* T43-Q388 (Figura 5.14 y Tabla 5.2), verificando que la eliminación adicional de residuos en el extremo N-terminal no afectaba a las propiedades cinéticas de la glicosiltransferasa.

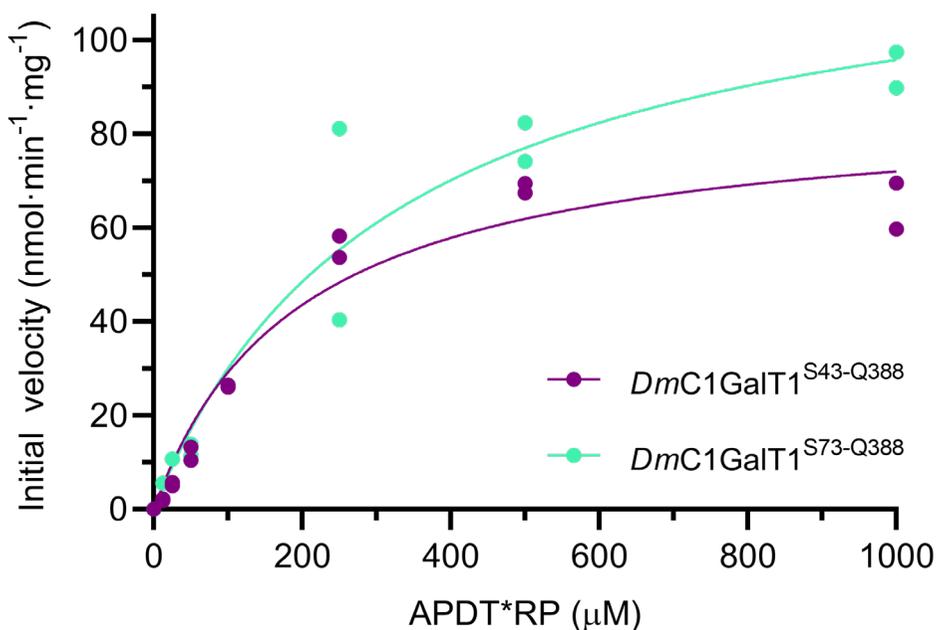


Figura 5.14. Cinéticas de *DmC1GalT1*^{T43-Q388} y *DmC1GalT1*^{S73-Q388} con concentraciones variables de APDT*RP. Se usaron concentraciones que variaban desde 12,5 μM a 1 mM. Las velocidades iniciales se obtuvieron por duplicado (n=2 de ensayos independientes) para cada concentración de sustrato.

Se obtuvieron cristales de *DmC1GalT1* en el grupo espacial P2₁ en presencia de UDP, MnCl₂ y APDT*RP. A pesar de que se intentó obtener cristales con otros glicopéptidos, no se pudieron conseguir condiciones en la que se formasen cristales que pudiesen difractar. Previo a la obtención de la estructura tridimensional, se intentaron realizar otros ensayos (con metales pesados y con ligandos unidos a metales pesados) debido a que no se disponía de un modelo molecular fiable con el que realizar el reemplazamiento molecular y resolver el problema de las fases con los datos obtenidos de las difracciones. La estructura cristalina se obtuvo finalmente mediante reemplazamiento molecular a 2,40 Å usando como base el modelo de *DmC1GalT1* procedente del servidor de Alpha Fold 2 (Jumper et al., 2021) (Materiales y métodos 4.7.4, Figura 5.15 y Tabla 5.5).

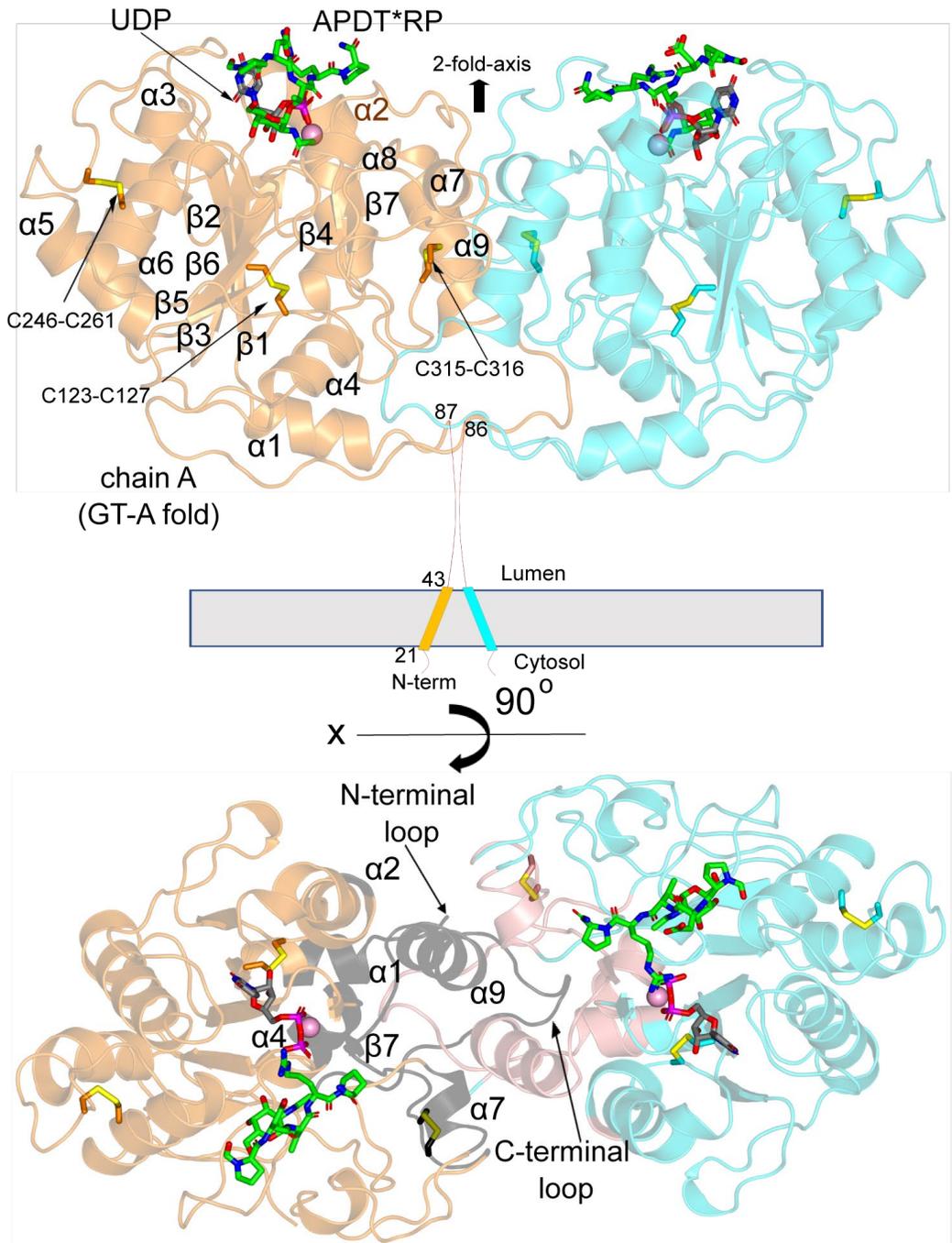


Figura 5.15. Estructura cristalina de la forma dimérica del complejo *DmC1GalT1*^{S73-Q388}-UDP-Mn²⁺-APDT*RP en formato *ribbon*. Las cadenas A y B se encuentran coloreadas en naranja y cian respectivamente. Los átomos del nucleótido de UDP están representados en gris mientras que el átomo de manganeso se muestra como una esfera rosa. Los átomos de carbono del glicopéptido son los coloreados en verdes y los puentes disulfuros en amarillo, estando el último de ellos formado por dos residuos de Cys contiguas. En la parte inferior de la imagen el dímero se encuentra rotado 90° resaltando las estructuras secundarias de los extremos N y C terminal involucradas en las interacciones en la interfase entre dímeros (negro y rosa para la cadena A y B respectivamente). En el centro de la imagen se representa la membrana del aparato de Golgi y las secuencias que no se encuentran en la estructura cristalina, las cuales anclarían el dímero al interior del aparato de Golgi y el extremo N-terminal estaría expuesto al citosol.

La unidad asimétrica (AU) de los cristales P2₁ contenía dos moléculas de *DmC1GalT1* ordenadas formando un homodímero con cada uno los monómeros adoptando el plegamiento típico de las GT-A (Figura 5.15). La forma dimérica de la proteína se confirmó mediante cromatografía de exclusión molecular (Figura 4.11) además de haber sido descrita este estado con anterioridad para el ortólogo humano (Ju, Brewer, et al., 2002). El servidor PISA confirmó esta estructura dimérica y reveló que el dímero presentaba una gran superficie enterrada (7621 Å²), lo que implicaba una interfase estrecha y estable entre ambas unidades proteicas. Los residuos en la interfase involucrados en la estabilización dimérica se encontraban en el loop del extremo N-terminal, $\alpha 1$, $\alpha 2$, $\alpha 4$, loop $\alpha 6$ - $\alpha 7$, $\alpha 7$, loop $\alpha 7$ - $\alpha 8$, loop $\alpha 8$ - $\beta 7$, $\beta 7$, loop $\beta 7$ - $\alpha 9$, $\alpha 9$, y el largo y desestructurado loop en el extremo C-terminal (Figura 5.15 panel inferior y Figura 5.2).

Tabla 5.5. Recolección de datos de la difracción cristalográfica y datos estadísticos del refinamiento de la estructura de *DmC1GalT1*^{S73-Q388}.

<i>DmC1GalT1</i> en complejo con UDP-Mn²⁺ y APDT*RP			
Recolección de datos		Refinamiento	
Grupo espacial	P2 ₁	Resolución (Å)	2,4
Dimensiones de la celda		Nº reflexiones	75579
<i>a. b. c</i> (Å)	50,24. 80,44. 71,95	<i>R</i>_{work} / <i>R</i>_{free}	0,175/0,234
$\alpha. \beta. \gamma$ (°)	90. 93,19. 90	Nº átomos	
Resolución (Å)	20-2,40	Proteína	4580
	(2,53-2,40*)	APDT*RP	112

R_{merge}	0,143 (0,825)	UDP	50
R_{pim}	0,091 (0,526)	Mn ²⁺	2
I / σI	7,1 (2,6)	Aguas	74
Integridad (%)	99,8 (100)	Etilenglicol	12
Redundancia	3,4 (3,3)	Factores - B (Å²)	
Mn(I) half-set correlation CC (1/2)	0,981 (0,412)	Proteína	34,49
<p>Se usó un solo cristal para determinar la estructura cristalográfica.</p> <p>*Los valores en paréntesis se refieren a la capa de mayor resolución.</p> <p>Dos subunidades se encuentran en la unidad asimétrica formando un dímero.</p>		APDT*RP	41,05
		UDP	29,45
		Mn ²⁺	24,55
		Aguas	31,57
		Etilenglicol	34,9
		R.m.s. desviaciones	
		Longitud de unión (Å)	0,0108
		Ángulos de unión (°)	1,7686

Curiosamente, uno de los residuos que se encontraba altamente conservado, localizado en la interfase de la cadena $\alpha 9$ se encontró mutado en ratones los cuales acababan con trombocitopenia y nefropatía. Concretamente se encontraba mutado el residuo Tyr321 de ratones (326 en *DmC1GalT1*) a Asn, (Figura 5.2). De forma adicional, ha sido demostrado que Cosmc se une a los residuos 83-97 de la *C1GalT1* humana (*HsC1GalT1*)(Aryal et al., 2014), localizados en las cadenas $\alpha 1$, $\beta 1$ y en el loop $\beta 1$ - $\alpha 2$ de la estructura de *DmC1GalT1*. Uno de estos aminoácidos, Leu95 en *DmC1GalT1* (Leu82*HsC1GalT1*) se encuentra en $\alpha 1$ en la interfase del dímero (Figura 5.2), sugiriendo que Cosmc es importante en la formación obligada del dímero de *C1GalT1*. Por otro lado, esta región peptídica se encuentra parcialmente conservada dentro de los ortólogos de *C1GalT1*, tanto vertebrados como invertebrados, implicando que este péptido particular en *DmC1GalT1* es poco probable que sea reconocido por Cosmc. De cualquier manera, ambos ejemplos, aunque opuestos, ilustran la gran importancia de estos residuos de la interfase dimérica en la estabilidad y función de *C1GalT1* (Taujale et al., 2021). La

desviación media cuadrática (RMSD en inglés; es una medida estándar de la distancia estructural entre dos coordenadas que mide la distancia media entre un grupo de átomos) entre ambas moléculas pertenecientes a la cadena A y B de la unidad asimétrica es de 0,24 Å en un total de 278 átomos de $C\alpha$ equivalentes, lo que de otro modo enfatiza que no existen apenas cambios conformacionales entre las dos unidades monoméricas, siendo ambas prácticamente iguales. A partir de este punto, se discutirá solo la molécula A debido a que la densidad electrónica de los ligandos se encuentra mucho mejor definida en la misma.

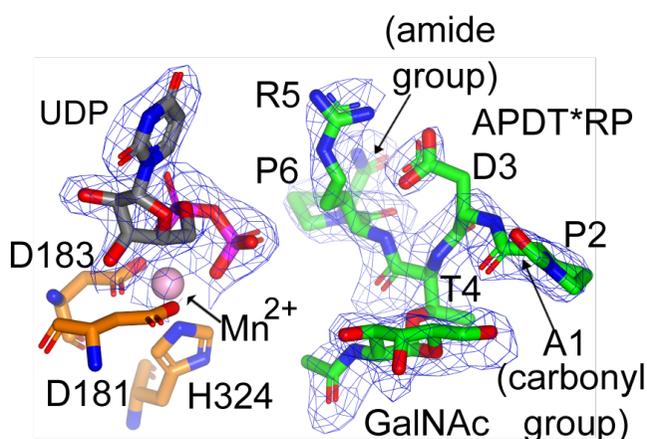


Figura 5.16. Vista en primer plano del sitio activo de *DmC1GalT1*^{S73-Q388} mostrando Mn^{2+} , UDP y APDT*RP unidos a la proteína. El mapa de densidad electrónica (azul) son los equivalentes a $F_o - F_c$ contorneados a $2,2 \sigma$ (densidad de carga superficial, que es la cantidad de carga por unidad de área (σ)). A excepción del primer residuo del extremo N-terminal (Ala1) de APDT*RP, la densidad electrónica del glicopéptido y del UDP estaban bien definidas. Cabe destacar que Pro6 acaba como un grupo amida.

Se observó que esta proteína tiene también las cuatro características más conservadas entre las glicosiltransferasas de tipo GT-A (Taujale et al., 2021): el motivo DxD para las interacciones con metales catiónicos (Asp181-X-Asp183); un loop rico en Gly frente a los sitios de unión del sustrato aceptor y el azúcar donador, localizado en la estructura de *Drosophila melanogaster* en el loop $\beta 5 - \beta 6$; un motivo

“xED” al principio de $\alpha 6$ albergando la base catalítica de la enzima (Asp255, la cual se definirá en el siguiente apartado); y un residuo “C-His” que se coordina con el ion de metal (His324) (Figura 5.16 y 5.2).

Una inspección minuciosa del sitio activo de *DmC1GalT1* y su comparación con otros ortólogos de C1GalT1 como el de humano, ratón y pollo desveló que tanto el sitio de unión a UDP-Gal como el del glicopéptido eran idénticos (Figura 5.17 panel superior y 5.2), demostrando que *DmC1GalT1* es un modelo excelente para entender los aspectos bioquímicos de la enzima humana.

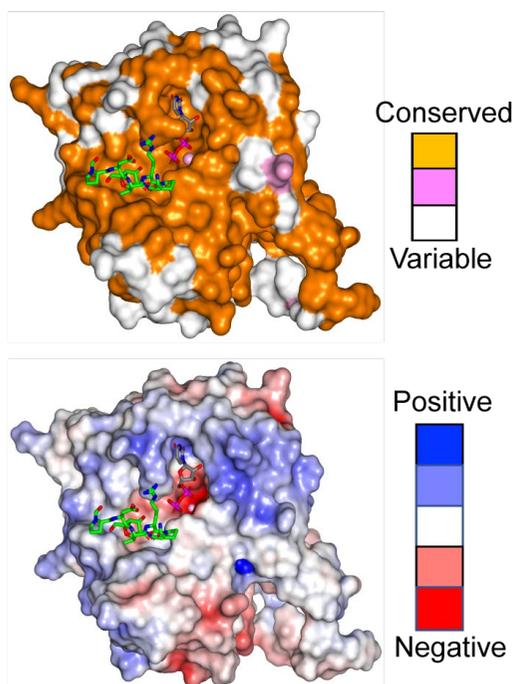


Figura 5.17. Representación superficial de homología de C1GalT1 y de la superficie electrostática de *DmC1GalT1*^{S73-Q388}. En la parte superior se encuentra la homología superficial, codificada por colores según el grado de conservación entre los ortólogos de mosca, humano, ratón y pollo. En la parte inferior se muestra la representación de la superficie electrostática de *DmC1GalT1* en la que la escala varía desde -5 kTe^{-1} a $+5 \text{ kTe}^{-1}$ (unidades que representan la constante de Boltzmann con respecto a la temperatura, refiriéndose a energía, por unidad de carga).

El análisis del potencial electrostático de la superficie proteica mostraba una zona cargada negativamente correspondiente al sitio de unión de UDP-Gal, la cual es requerida para coordinar el Mn^{2+} , y zonas cargadas de manera positiva y neutra en los sitios de unión del péptido. Además, el sitio de unión a GalNAc se encontraba moderadamente cargado negativa y positivamente frente al núcleo central y al grupo acetamido/OH6, respectivamente (Figura 5.17 panel inferior). Con respecto a la homología estructural de *DmC1GalT1* a otras estructuras ya descritas, el servidor DALI (Holm & Laakso, 2016) reveló cierta homología estructural con otras dos galactosiltransferasas, la B3GNT2 humana en forma de homodímero (entradas del PDB 7JHN (Hao et al., 2021) y 6WMO (Kadirvelraj et al., 2021) y la “Manic fringe” monomérica de ratón (entrada del PDB 2J0A y 2J0B (Jinek et al., 2006), ambas pertenecientes a la familia CAZy31 (Figura 5.18).

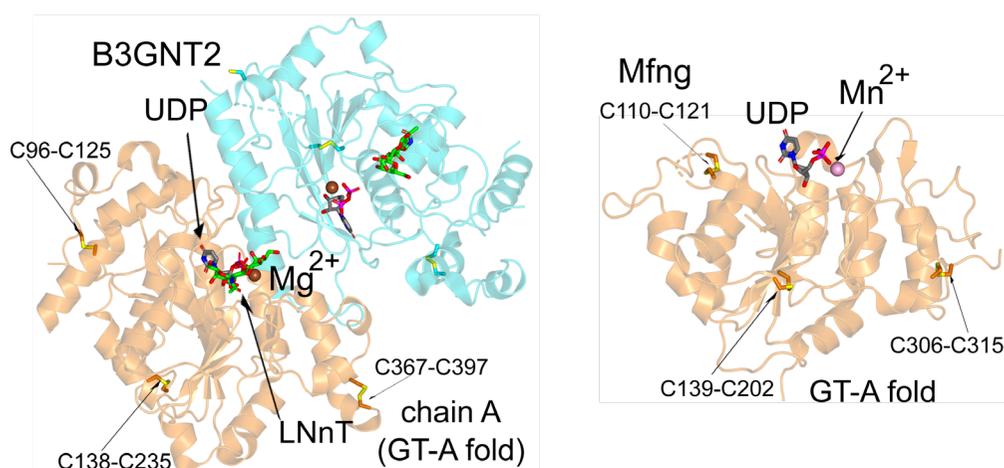


Figura 5.18. Homólogos estructurales de *DmC1GalT1*, los ligandos y los cofactores. Representados siguen el mismo código de color definido en imágenes anteriores.

Aunque los sustratos aceptores son muy diferentes entre las tres enzimas, estas se superponían bastante bien, siendo el servidor capaz de procesar buenos resultados. Las RMSDs obtenidas fueron de aproximadamente 1,7 y 3,17 Å entre las

estructuras cristalinas de *DmC1GalT1* y B3GNT2, y *DmC1GalT1* y Mfng, respectivamente (189 y 151 residuos superpuestos). Las fuertes similitudes entre *DmC1GalT1* y B3GNT2 en el plegamiento general resultan muy interesantes ya que además se observaba una excelente superposición del UDP y de los sustratos aceptores (Figura 5.19). Merece la pena mencionar que los grupos OH3 tanto del GalNAc de APDT*RP como del residuo de galactosa de LNnT (el sustrato aceptor de B3GNT2) se encontraban en posiciones casi idénticas (con un desplazamiento atómico entre ambas moléculas de aproximadamente 0,92 Å) y cerca del fosfato β, en concordancia con su papel como sitios aceptores. Hay que tener en cuenta que el UDP en Mfng está también muy bien superpuesto con el UDP correspondiente a la estructura cristalina de *DmC1GalT1*, aunque la estructura de Mfng solo se obtuviese con UDP y Mn²⁺ y no con sustrato aceptor.

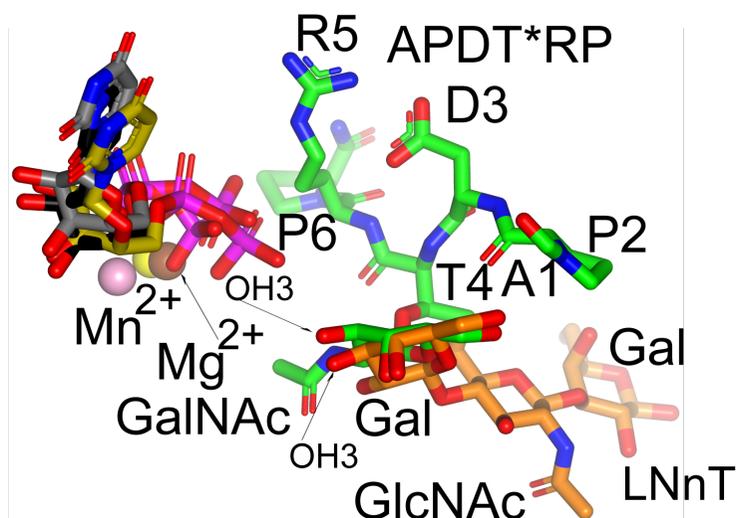


Figura 5.19. Superposición de los ligandos de *DmC1GalT1*, B3GNT2 y Mfng. Los carbonos del sustrato LNnT se muestran en naranja y los del glicopéptido en verde. Los carbonos del UDP se muestran en gris en *DmC1GalT1*, verde oliva en B3GNT2 y negro en Mfng. Los átomos de Mn²⁺ están coloreados en rosa para *DmC1GalT1* y en amarillo para Mfng. El átomo de Mg²⁺ de B3GNT2 está representado en marrón.

5.8. El sitio activo de *DmC1GalT1*.

El sitio activo de *DmC1GalT1* está formado por los sitios de unión de UDP-Gal y del glicopéptido (Figura 5.20). El residuo de uridina que forma parte de la molécula de UDP establece una interacción CH- π con Leu155 mientras que el residuo de uracilo se une a través de puentes de hidrógeno con el *backbone* de Gly151 y con las cadenas laterales de Glu150 y Lys158. La ribosa de la molécula de uridina interacciona con la cadena lateral de Asp192 y el *backbone* de Met160 y el pirofosfato de la misma molécula, a su vez, interactúa con las cadenas laterales de Arg152, His324 y Tyr325. Los átomos de oxígeno del grupo pirofosfato, los aminoácidos Asp181 y Asp183 del motivo DxD y His324 se coordinan hexagonalmente con el ion metálico Mn^{2+} , formando una hexacoordinación octaédrica.

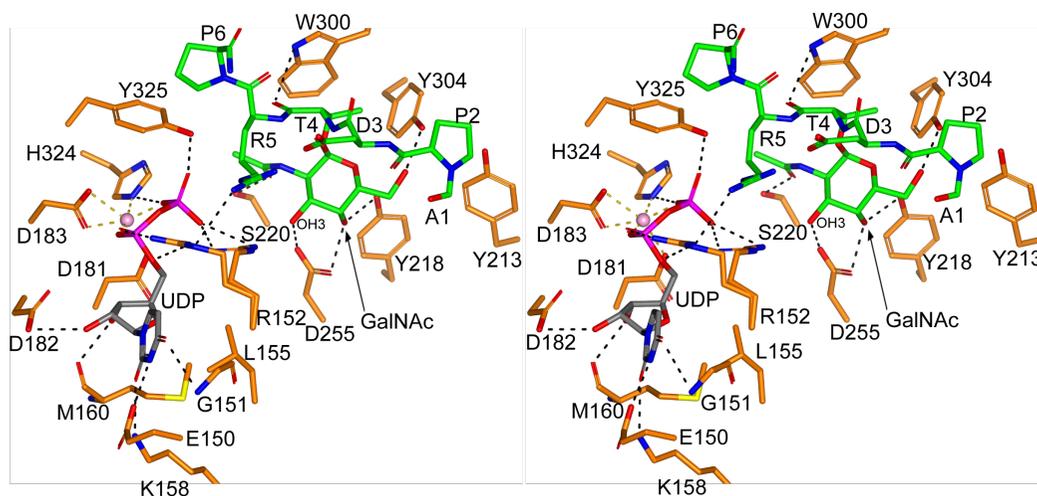


Figura 5.20. Vista estereoscópica del sitio activo de *DmC1GalT1*^{S73-Q388} en complejo ternario con UDP- Mn^{2+} y APDT*RP. Los carbonos de los residuos que forman parte del sitio activo se encuentran representados en naranja, los del UDP en gris y los del glicopéptido en verde. El átomo de manganeso se muestra como una esfera rosa. Las interacciones de puentes de hidrógenos se representan mediante líneas discontinuas negras.

A diferencia del reconocimiento íntimo que sucede entre *DmGalT1* y la molécula

de UDP, con APDT*RP se observa mucho menos contacto con la enzima (Figura 5.20), concordando con los datos obtenidos mediante los experimentos de ITC, en los cuales la unión de UDP era unas 9,5 veces más fuerte que la unión del APDT*RP a la enzima (Tabla 5.4). El grupo GalNAc del glicopéptido es reconocido mediante puentes de hidrógeno que se forman entre: el carbonilo de la acetamida y la cadena lateral de Ser220; el OH3 con la cadena lateral de Asp255; el OH4 con las cadenas laterales de Asp255/Tyr218; y el OH6 con la cadena lateral de Tyr304. A nivel peptídico, la cadena lateral de Pro2 y el grupo metilo de Thr4 establecen interacciones tipo CH- π con Tyr213 y Trp300/Tyr304, respectivamente, y el *backbone* de Thr4 forma un puente de hidrógeno con la cadena lateral de Trp300. La cadena lateral de Arg5 establece un puente de hidrógeno con Arg152, y Pro6 establece una interacción de tipo CH- π con Tyr325 (Figura 5.20).

Estas interacciones desvelan que el residuo de GalNAc está reconocido de manera más estrecha que el péptido a causa de que el primero se encuentra enlazado mediante numerosos puentes de hidrógeno mientras que el segundo es reconocido mediante menos interacciones, de las cuales algunas son interacciones hidrofóbicas además de los puentes de hidrógeno. En general, los datos obtenidos se alinean con los mapas de epítomos derivado de los espectros de STD-RMN, sugiriendo que el residuo de GalNAc es la principal fuerza responsable para que ocurra el reconocimiento del sustrato aceptor y, además, que el péptido mejora la unión del mismo estableciendo interacciones directas con la enzima.

5.9. El reconocimiento molecular de α -GalNAc-Thr es posible gracias a la conformación de alta energía que adopta su enlace glicosídico.

Una de las características estructurales más peculiares que se ha podido deducir gracias a la obtención de la estructura cristalina en complejo con el sustrato aceptor ha sido la conformación energética menos favorable que adopta el enlace glicosídico de la Thr unida a GalNAc (α -GalNAc-Thr) (Figura 5.21 izquierda).

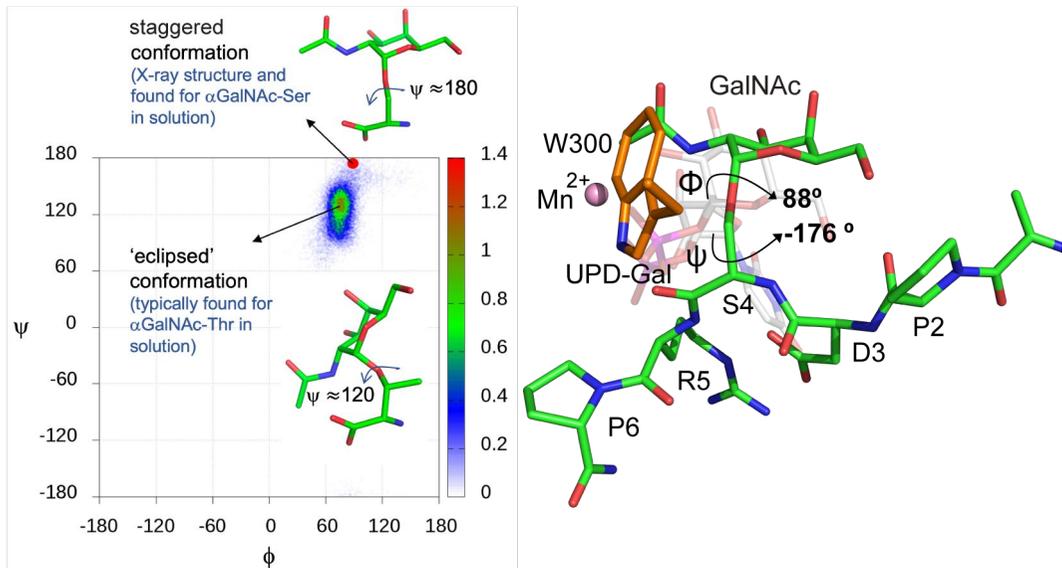


Figura 5.21. Geometría del enlace glicosídico del glicopéptido APDT*RP en solución obtenida mediante simulaciones de dinámica molecular de 0,5 μ s (conformación eclipsada). Los ángulos diédricos están definidos de la diferente manera: $\phi = \text{O5-C1-O1-C}\beta$, y $\psi = \text{C1-O1-C}\beta\text{-C}\alpha$. El círculo rojo indica la conformación adoptada por este enlace en APDT*RP en la estructura cristalina del complejo ternario de *DmC1GalT1* en presencia de UDP y Mn^{2+} . Esta conformación es en la que se encuentra de manera habitual el glicopéptido APDS*RP en solución (conformación alternada) (izquierda de la imagen). A la derecha se muestra una vista tridimensional del glicopéptido APDS*RP con *DmC1GalT1* y en presencia de UDP-Gal, obtenida mediante simulaciones de dinámica molecular de 0,5 μ s.

La conformación alternada suele encontrarse en residuos de serina unidos a una molécula de α -GalNAc en solución. La característica que define esta conformación es que el ángulo ψ es de $\approx 180^\circ$ (ángulo formado entre $\text{C1-O1-C}\beta\text{-C}\alpha$). Este ángulo en residuos de α -GalNAc-Thr, ya sea en su forma libre (Corzana et al., 2007) o bien unida a proteínas, adopta una conformación que se denomina eclipsada en la que el ángulo del rotámero es la forma habitual con una torsión de $\psi \approx 120^\circ$ (Bermejo et al., 2018; Lira-Navarrete et al., 2015; Macías-León et al., 2020; Madariaga et al., 2014). Para comprobar estos datos, se realizaron simulaciones de dinámica molecular (MD) con *DmC1GalT1* en complejo con UDP-Gal y APDS*RP,

obteniéndose en dichos cálculos que la conformación alternada también era predicha en el caso de α -GalNAc-Ser (Figura 5.21 panel derecho y Materiales y métodos 4.9.3), lo que implica la necesidad de adoptar esa conformación alternada con ambos sustratos para una glicosilación efectiva de los aminoácidos Ser y Thr unidos a α -GalNAc. A la hora de realizar simulaciones con el glicopéptido APDT*RP pero fijando la conformación de α -GalNAc-Thr a su forma eclipsada (Figura 5.22 y Materiales y métodos 4.9.3), se observó una enorme pérdida de interacciones entre el péptido y la proteína, en comparación con aquellas encontradas en la estructura cristalográfica, que adoptaba la conformación “alternada”.

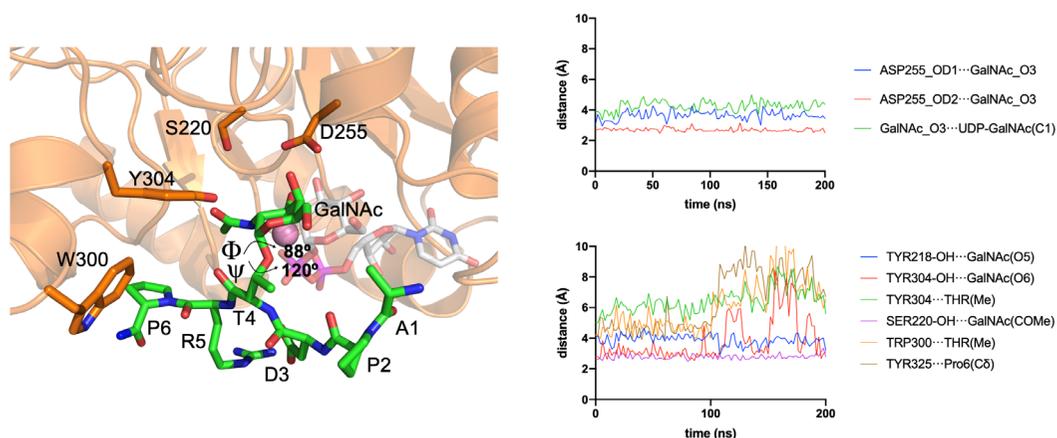


Figura 5.22. Vista en primer plano del complejo *DmC1GalT1-APDT*RP* en presencia de UDP-Gal-Mn²⁺. Estos datos se obtuvieron de MDs con restricciones durante 0,2 μ s en las que el enlace glicosídico fue forzado a adoptar una conformación “eclipsada” ($\psi \approx 120^\circ$). A la derecha se muestran las distancias de las interacciones entre el glicopéptido y la enzima con el tiempo.

Específicamente, las interacciones de tipo CH- π entre el grupo metilo de Thr4 y Trp300/Tyr304 y entre Pro6 (C δ) del glicopéptido y Tyr325 se debilitaron significativamente a causa del aumento de la distancia entre los anillos aromáticos y el péptido. El puente de hidrógeno entre el grupo carbonilo de Thr4 y Trp300 se volvía prácticamente inexistente durante esta simulación de dinámica molecular a causa de las restricciones impuestas. Con respecto al residuo de GalNAc, el puente

de hidrógeno entre la cadena lateral de Tyr304 y el OH6 del mismo desaparecía. Por otro lado, el glicopéptido APDS*RP tiene una K_m^{app} ligeramente peor que su derivado con treonina a pesar de no tener la posible penalización conformacional que sí que ocurre en el péptido que tiene la Thr*. Estos resultados podrían sugerir que efectos de energía libre muy sutiles estén probablemente guiando la unión. En el caso de la treonina, la penalización correspondiente de energía libre asociada al cambiar la conformación del enlace glicosídico de una conformación eclipsada a una alternada es de 2,5 kcal/mol (Figura 5.23). Por otro lado, este cambio conformacional es favorecido por 1,9 kcal/mol en el derivado de serina, indicando que es muy probable que en solución se encuentre en esa conformación alternada (Figura 5.23).

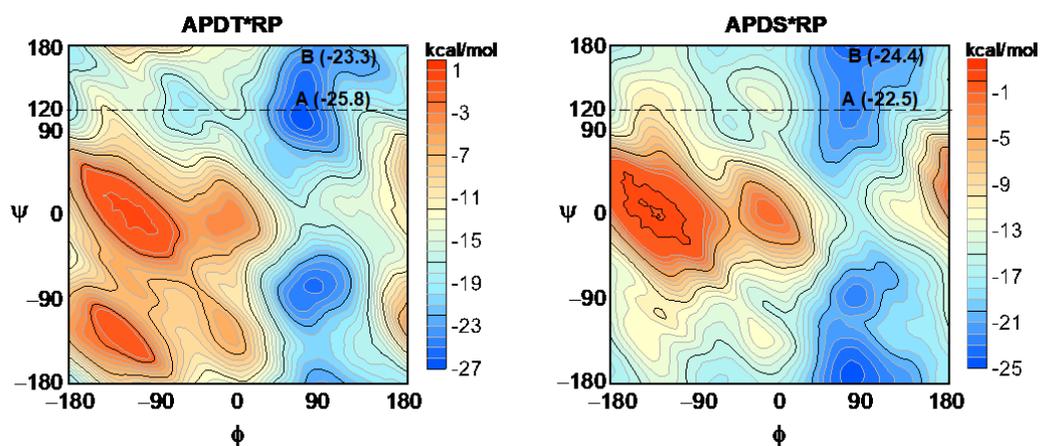


Figura 5.23. Mapas de energía libre (ϕ , ψ) de los ángulos diédricos glicosídicos. Los ángulos fueron calculados con los péptidos libres en agua usando el método ABMD (dinámica molecular adaptativamente sesgada) implementado en AMBER 20 a 300 K. Los mapas de contorno están dibujados con un espaciado de 1 kcal/mol. Las regiones en las que no se observa presencia de los péptidos se muestran como naranja oscuro. “A” refiere a la conformación “eclipsada” encontrada típicamente en derivados de α -GalNAc-Thr en solución (Bermejo et al., 2018; Corzana et al., 2007). “B” se refiere a la conformación “alternada” que se encuentra en glicopéptidos con α -GalNAc-Ser en solución (Corzana et al., 2006).

Este descubrimiento explica cómo gracias al aumento en las interacciones que se

producen entre el glicopéptido de treonina y la enzima con respecto al glicopéptido de serina se llega a equilibrar el penalti de energía que requiere el cambio conformacional. También explica esto probablemente por qué C1GalT1 tiene parámetros cinéticos similares para glicopéptidos con Ser* y Thr*, y por qué puede glicosilar indistintamente tanto los residuos de α -GalNAc-Ser como de α -GalNAc-Thr.

5.10. La transferencia de la galactosa se realiza mediante un mecanismo de inversión a través de *DmC1GalT1*.

Para poder conseguir mayor información sobre el mecanismo de inversión de C1GalT1, se superpuso la estructura tridimensional resuelta en complejo con UDP-Mn²⁺ y APDT*RP con la estructura de la enzima humana B3GNT2 en complejo con UDP-GlcNAc (entrada del PDB 7JHL) reemplazando las coordenadas de la molécula de UDP-GlcNAc por las equivalentes a una molécula de UDP-Gal. Las posiciones de los átomos del complejo resultante, *DmC1GalT1*-UDP-Gal-Mn²⁺-APDT*RP, fueron minimizadas a través del cálculo de mecánica molecular descrito en Materiales y métodos 4.9.3 (Figura 5.24). En la estructura obtenida mediante la simulación se observa que el OH3 de la molécula de GalNAc se encuentra correctamente alineado y lo suficientemente cerca como para atacar al carbono anomérico, encontrándose la molécula de galactosa en una posición que permite la inversión de la configuración de α a β durante la reacción.

Para confirmar la importancia de Asp255 en la catálisis, se mutó ese mismo aminoácido a Ala. La actividad del mutante D255A era completamente inexistente, confirmando dicho aminoácido como base catalítica (Figura 5.25a). Por otra parte, los demás mutantes que podían tener un rol importante en la glicosilación presentaban una actividad del mismo modo nula o reducida, la cual se discutirá en el siguiente apartado.

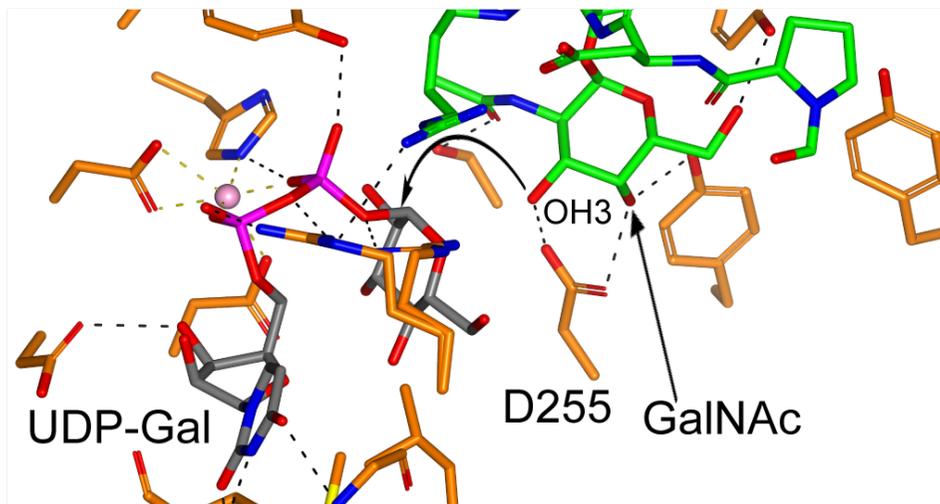


Figura 5.24. Simulación y vista detallada del sitio activo de *DmC1GalT1* con los sustratos de la reacción. Gracias a la simulación e incorporación en la estructura de la molécula de UDP-Gal, se puede observar cómo el Asp255 podría ser la base catalítica en un plausible mecanismo de reacción S_N2 de desplazamiento simple. La orientación y proximidad del OH3 del GalNAc hacia el carbono anomérico (3,81 Å) es compatible con la inversión de la configuración del azúcar durante la reacción, apoyando el mecanismo de inversión enzimático de C1GalT1.

Tras estos resultados se puede concluir que C1GalT1 sigue el típico mecanismo de inversión en el cual la base catalítica desprotona el OH3 del GalNAc de manera que el oxoanión resultante puede proceder a atacar al carbono anomérico del residuo de Gal, el cual sufre un estado de transición semejante a un ion oxocarbenio (Figura 5.25b). Estos datos que se han descrito son completamente compatibles con el mecanismo de reacción de desplazamiento único S_N2 , el cual es el más común dentro de las GTs de inversión (Moremen & Haltiwanger, 2019).

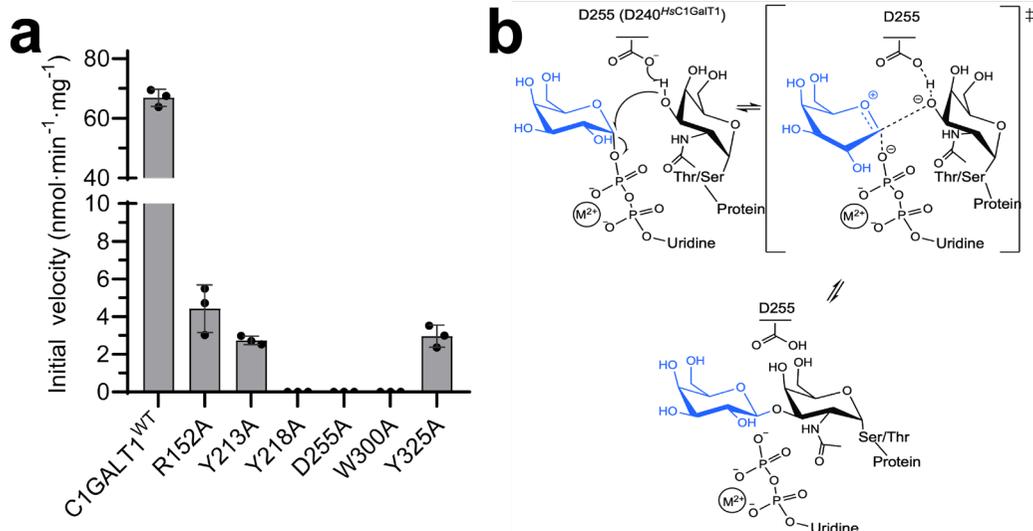


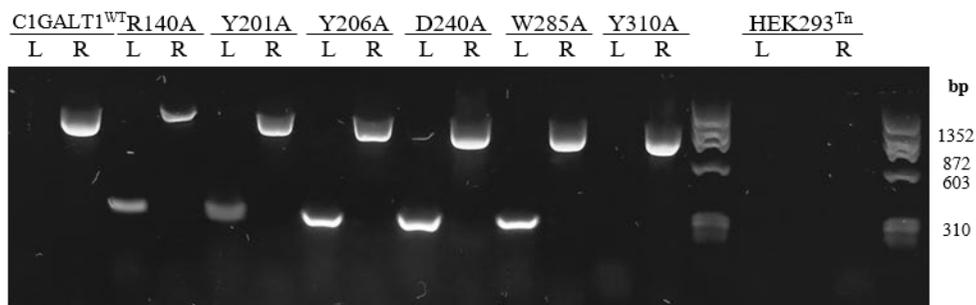
Figura 5.25. Actividades relativas de los mutantes de *DmC1GalT1* (a) y mecanismo de reacción propuesto (b). **a.** Histograma con las actividades relativas de los mutantes de *DmC1GalT1*^{T43-Q388} en comparación con la construcción *wild-type*. Todos los experimentos se realizaron en triplicado (n = 3 experimentos independientes). Las barras de error representan la desviación estándar calculada mediante el ajuste de los datos por GraphPad Prism. **b.** Mecanismo de reacción S_N2 de desplazamiento simple propuesto para C1GalT1.

5.11. Comportamiento y actividad de los mutantes de *DmC1GalT1 in vitro* y de los mutantes de *HsC1GalT1 in vivo*.

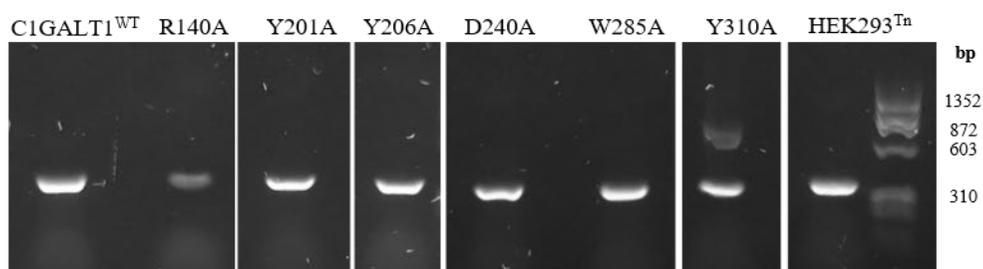
Con el objetivo de aclarar el rol que juegan los residuos de *DmC1GalT1* que están más estrechamente involucrados en las interacciones con el glicopéptido, se realizaron mutaciones en los residuos Arg152, Tyr213, Tyr218, Trp300 y Tyr325 a Ala, basándose en las interacciones observadas en el sitio activo del complejo tridimensional. Los mutantes resultantes fueron caracterizados *in vitro* bajo las mismas condiciones que el mutante D255A. Los resultados mostraban que tanto Y218A como W300A eran inactivos mientras que R152A mostraba una disminución de la actividad de 15 veces con respecto a la enzima silvestre. Por otro lado, los mutantes Y213A y Y218A sufrían una disminución de la actividad de 25 veces en comparación la enzima salvaje (Figura 5.25a). Esto sugería que tanto el

knockin dirigido (KI) (Métodos 4.2.3) (Figura 5.27 y 5.28c).

a Junction PCR



b WT PCR



c mAb 3C9

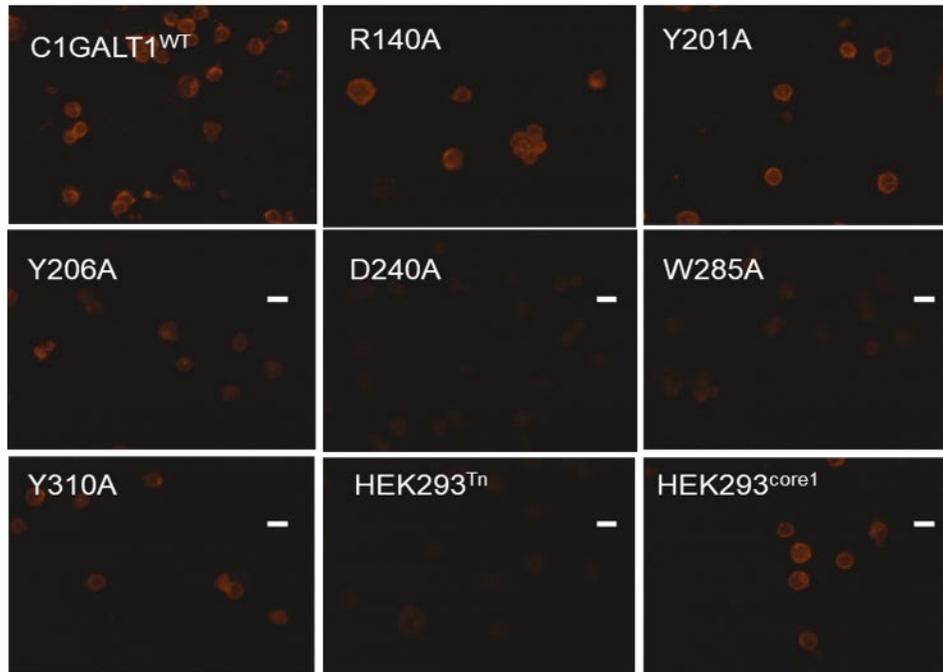


Figura 5.27. PCR de la inserción de los constructos y análisis inmunocitológicos con el mAb 3C9 específico para el antígeno T en los clones *knockins*. **a.** La PCR de la inserción confirmó (Izquierda (L) y derecha (R)) los constructos de la enzima silvestre y los mutantes de *HsC1GalT1* en el locus AAVS1 de las células HEK293^{Tn}. **b.** Para determinar el número de copias alélicas integradas en el ADN genómico se usó PCR alélica silvestre. La presencia de la banda deseada en las PCRs alélicas silvestres indican la amplificación del locus AAVS1 sin la integración del *knockin* objetivo de *HsC1GalT1* en al menos uno de los alelos. Todos los clones KI mostraban una integración monoalélica de los genes *HsC1GalT1* en el locus AAVS1. **c.** Análisis inmunocitológico de los clones KI con el anticuerpo específico mAb 3C9 para el antígeno T. Las imágenes son representativas de tres experimentos individuales. La barra de escala se corresponde con 20 μm .

La inducción de la expresión del *core 1* en la superficie celular se evaluó mediante citometría de flujo gracias al uso del anticuerpo monoclonal (mAb) anti-T 3C9 (Figura 5.28a y 5.28b). Este anticuerpo no tenía la capacidad de unirse a las células HEK293^{Tn} a causa de la ausencia de antígeno T, pero sí que se unían consistentemente a las células tras el KI de la *HsC1GalT1* silvestre.

Los KI de los mutantes R140A, Y201A, Y206A y Y310A recuperaban parcialmente algo de la unión al anticuerpo, indicando la expresión del antígeno T, siendo el mutante Y206A el menos efectivo de los cuatro. Los *knockins* de D240 y W285A no produjeron ningún tipo de unión al anticuerpo, sugiriendo que eran completamente inactivos (Figura 5.28b). Gracias a la obtención de estos resultados se apoya la hipótesis de que el residuo D240 de *HsC1GalT1* actúa ($\text{D255}^{\text{DmC1GalT1}}$) como la base catalítica de la enzima y que los residuos Y206 ($\text{Y218A}^{\text{DmC1GalT1}}$) y W285 ($\text{W300}^{\text{DmC1GalT1}}$) cumplen funciones esenciales en el reconocimiento y la catálisis enzimática. Estos resultados con los mutantes de la C1GalT1 humana están en concordancia con los obtenidos con los mutantes de *DmC1GalT1*, lo que implica una vez más la validación de este ortólogo de mosca como un buen modelo para poder entender la enzima humana.

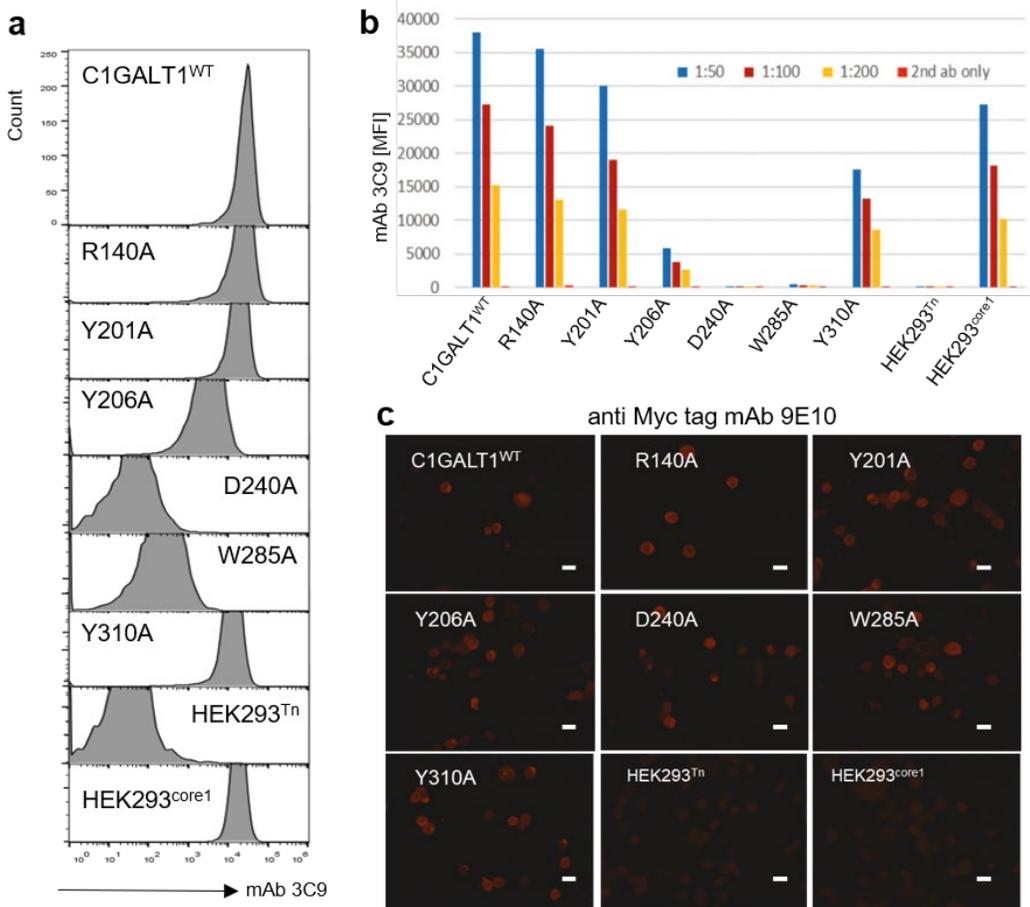


Figura 5.28. Análisis por citometría de flujo de la reinsertión del antígeno T con los mutantes de *HsC1GalT1*. **a.** Para evaluar el nivel de antígeno T o core 1 en la superficie celular, se realizaron análisis de citometría de flujo con un anticuerpo monoclonal específico denominado 3C9 (sobrenadante de hibridoma diluido 1:100). **b.** El diagrama de barras muestra la intensidad media de unión del mAb 3C9. **c.** Análisis inmunocitológicos de clones singulares con *knockin* de *HsC1GalT1* o los mutantes con el mAb 9E10 anti-Myc para detectar la unión a la etiqueta-Myc de C1GalT1. Es necesario tener en cuenta que los clones de HEK293^T tienen expresión endógena de *HsC1GalT1* sin la etiqueta-Myc. Las imágenes son representativas de tres experimentos independientes (n = 3) y la barra de la escala representa 20 μ m.

5.12. Simulaciones de dinámicas moleculares (MD) para la obtención de complejos estructurales putativos tridimensionales de *DmC1GalT1*.

A través de las simulaciones de dinámica molecular pudieron generarse estructuras

tridimensionales tanto de la forma apo de la enzima, como de la enzima en complejo con UDP-Gal, además de los complejos entre *DmC1GalT1* y los glicopéptidos APDT*RP, APDS*RP, P2, P4 y P7 (Figuras 5.29 – 5.35).

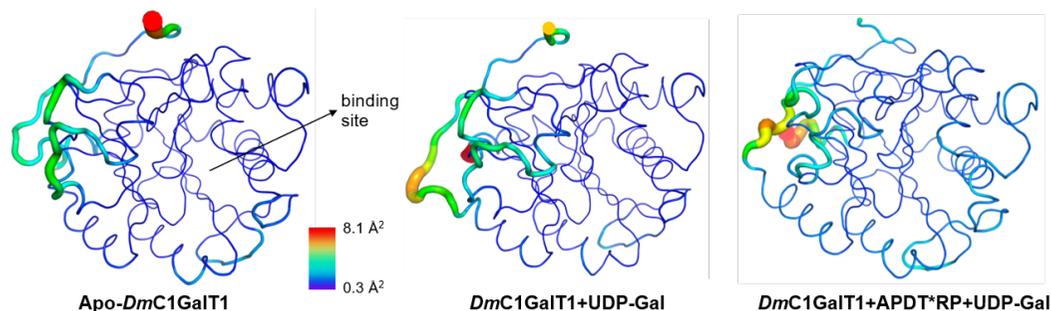


Figura 5.29. Fluctuaciones atómicas (Ca) obtenidas de las simulaciones de MD durante 0,5 μ s de *DmC1GalT1* y dos complejos binarios. Los datos corresponden a la media estructural de la proteína durante las simulaciones.

Al obtener estos datos mediante los cálculos realizados, se determinó que la enzima apenas sufría cambios en su estructura tridimensional en presencia de UDP-Gal, ni siquiera cuando se formaba el complejo ternario con APDT*RP (Figura 5.29), lo que corrobora la inexistencia de un mecanismo de inducción, como se ha mencionado anteriormente. Es extremadamente llamativo observar cómo el enlace glicosídico de todos los glicopéptidos siempre mantiene la conformación alternada del enlace, siendo esta la estrategia o mecanismo más probable que usa esta enzima para adaptarse y glicosilar indistintamente los residuos tanto de α -GalNAc-Thr como de α -GalNAc-Ser (Figura 5.30).

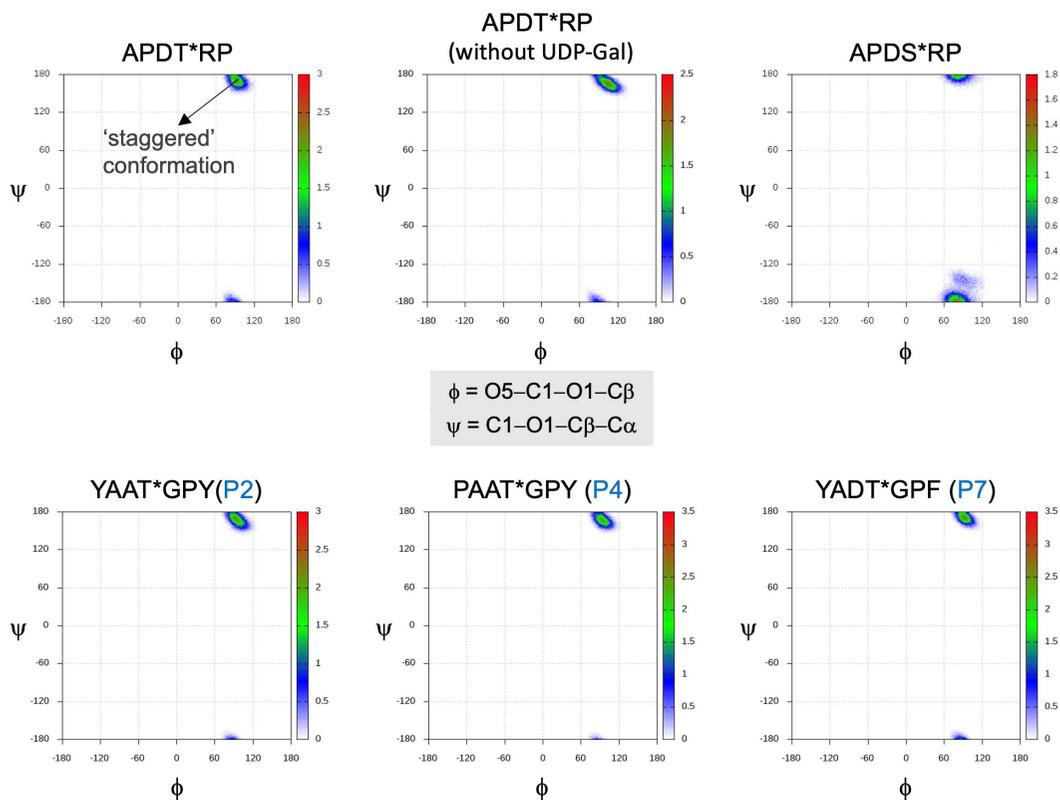


Figura 5.30. Relación entre los ángulos ϕ/ψ del enlace glicosídico obtenida mediante simulaciones de MD de diferentes complejos de *DmC1GalT1*. En todos los casos, se observaba cómo todos los péptidos unidos a la enzima adoptaban la conformación alternada.

En todos los complejos simulados se podían observar los mismos puentes de hidrógeno entre el residuo de GalNac y la enzima que los observados en la estructura obtenida por rayos X, no modificando la secuencia peptídica las interacciones entre la proteína y el azúcar (Figuras 5.30, 5.31 y 5.32). A continuación, serán detalladas las interacciones más interesantes encontradas en estas simulaciones.

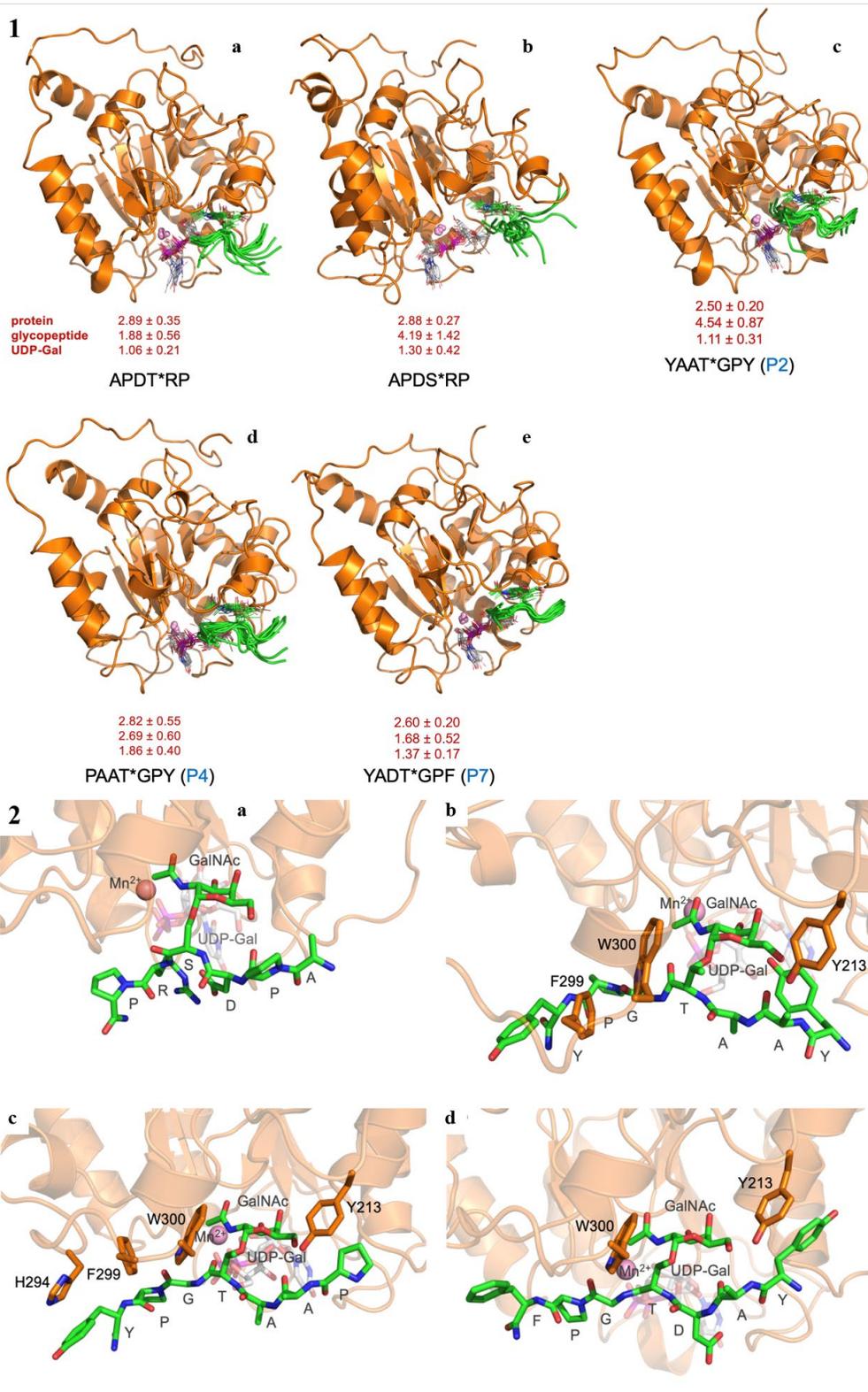


Figura 5.31. Simulaciones de MD de complejos ternarios de *DmC1GalT1*. ¹Superposición de 10 imágenes de diferentes complejos de *DmC1GalT1* obtenidas durante simulaciones de MD de 0,5 μ s. Se muestran en la imagen los valores de RMSD de la proteína, glicopéptido y UDP-Gal de arriba a abajo (la desviación estandar se muestra junto a cada valor). ²Imágenes obtenidas mediante MDs del sitio activo de *DmC1GalT1* en complejo con APDS*RP (a), P2 (b), P4 (c) y P7 (d) y UDP-Gal y Mn^{2+} .

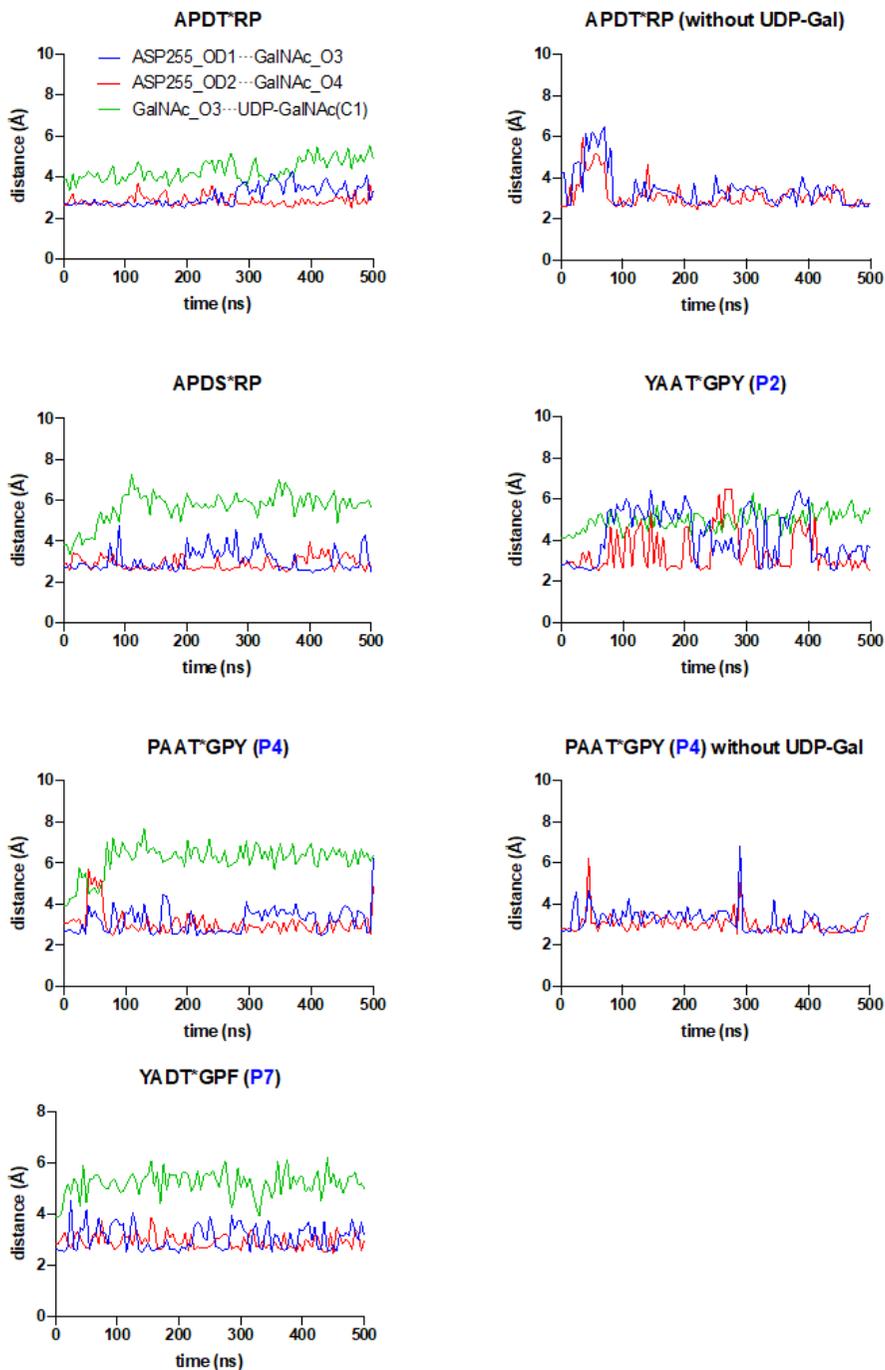


Figura 5.32. Representación de la distancia entre la enzima y el glicopéptido (Å) en diferentes complejos obtenidas mediante MDs durante 0,5 μ s. En la primera gráfica se indica qué representa cada curva.

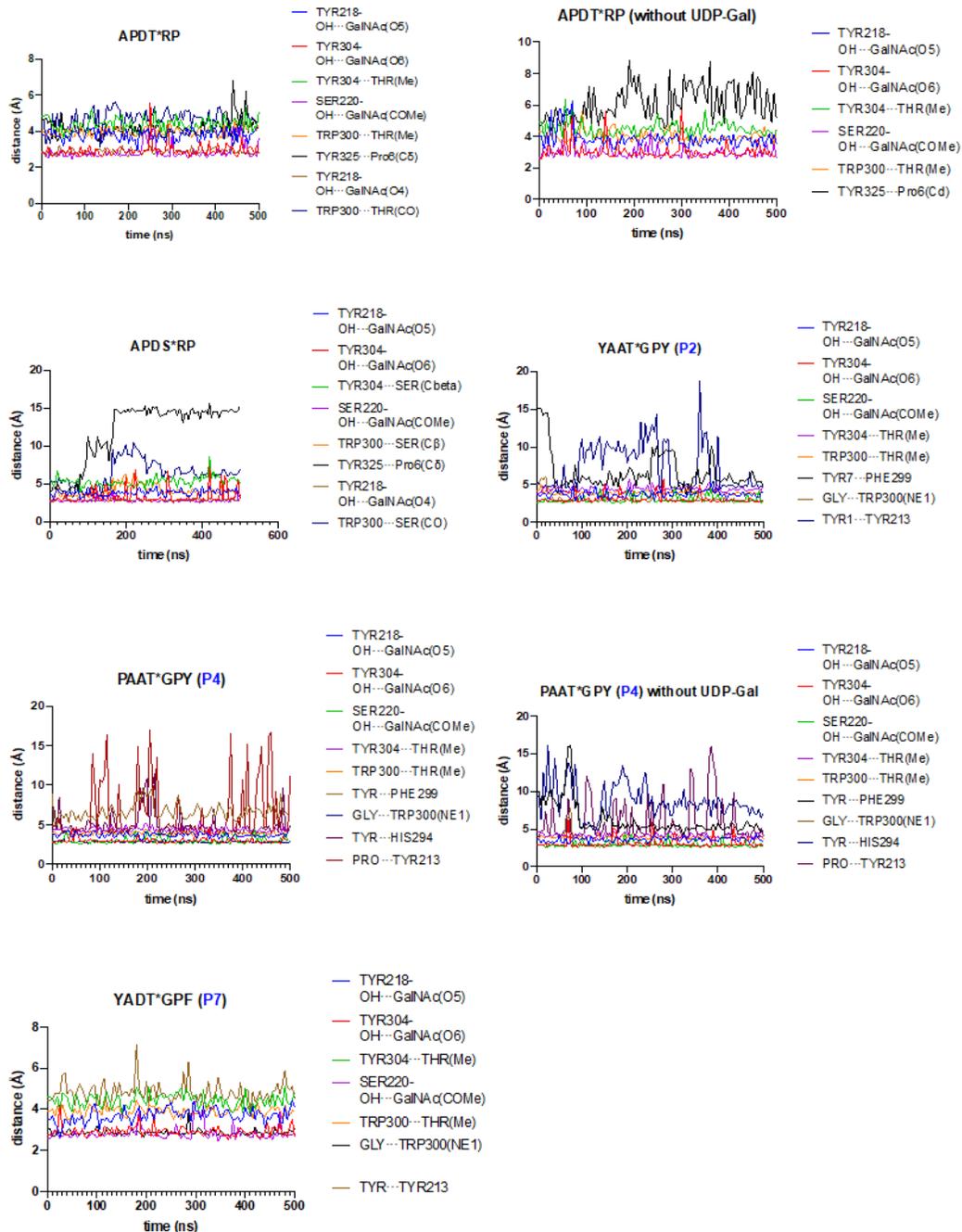


Figura 5.33. Representación de la distancia entre la enzima y el glicopéptido (Å) en diferentes complejos obtenidas mediante MDs durante 0,5 μ s.

La simulación en conjunto con el péptido APDS*RP (Figura 5.33) permitió obtener datos que demostraban la ausencia de interacciones de tipo CH- π entre Trp300 y Ser4 y la existencia de una interacción similar que parece producirse entre los hidrógenos del C β de Ser4 y la cadena lateral de Tyr304.

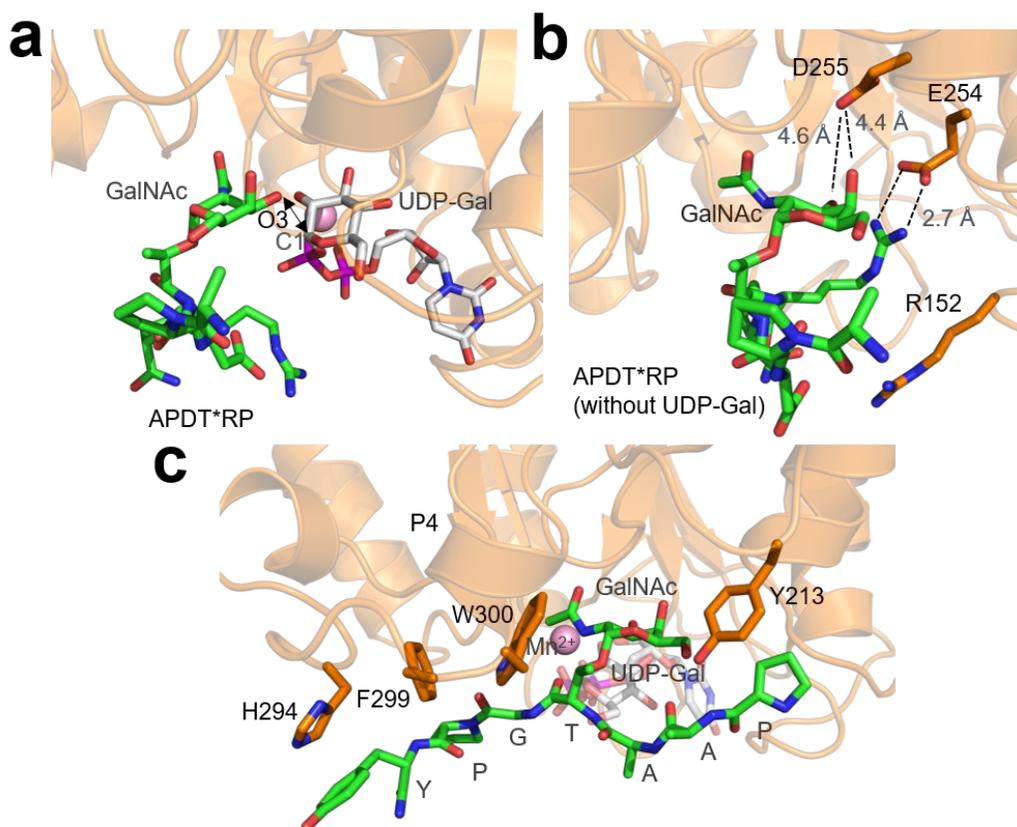


Figura 5.34. Gráfica de la RMSD derivadas de las simulaciones de MD (0,5 μ s) de la forma apo y los diferentes complejos. Para el cálculo de estos datos solo se usaron los átomos pertenecientes al *backbone* de la proteína.

En el complejo de APDT*RP con *DmC1GalT1*, el GalNAc y UDP-Gal tenían una orientación y posición adecuada, con una distancia entre el O3-GalNAc y el C1-Gal menor a 5,5 Å a lo largo de toda la simulación, reforzando la hipótesis del mecanismo de catálisis planteado y siendo consistente con el mecanismo de inversión deducido a partir de los resultados obtenidos (Figuras 5.33, 5.34 y 5.35a). Gracias a estas simulaciones de MD también se pudo comprobar que la unión entre

el glicopéptido y la enzima coincidía con los experimentos de STD-RMN realizados y descritos anteriormente (Tabla 5.6).

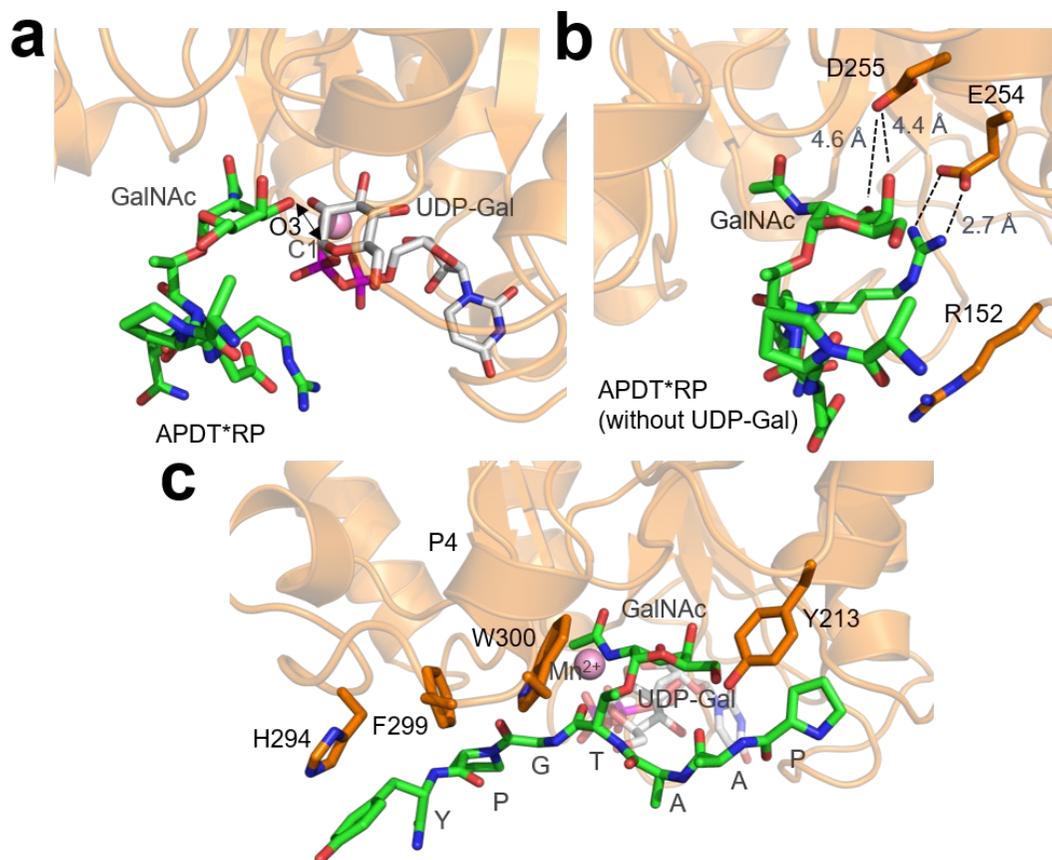


Figura 5.35. Simulaciones de MD de los complejos de *DmC1GalT1*. **a.** Complejo de *DmC1GalT1*-UDP-Gal-APDT*RP. **b.** Complejo *DmC1GalT1*-APDT*RP sin UDP-Gal. **c.** Simulación de DM del complejo de *DmC1GalT1*-UDP-Gal-P4 mostrando las interacciones entre la proteína y los extremos N y C terminal del péptido.

La ausencia de UDP-Gal no modificaba de manera significativa las interacciones entre los glicopéptidos y la enzima comparándolo con los complejos ternarios, a excepción del caso del complejo con el glicopéptido APDT*RP, coincidiendo con los resultados experimentales.

Tabla 5.6. Interacciones observadas durante la simulación de MD del complejo formado por APDT*RP y *DmC1GalT1* en presencia de UDP-Gal. Las interacciones medidas fueron las correspondientes a los protones no intercambiables del ligando y la proteína. En la tabla se muestran los contactos con una frecuencia de MD superior al 30% y una distancia inferior a 5 Å.

Complejo APDT*RP-<i>DmC1GalT1</i>-UDP-Gal		
Protones del ligando	Protones de la proteína	Respuesta de STD-RMN
H6s (GalNAc)	Tyr218	75-51%
	Val170	
	Tyr132	
	Tyr127	
	Ala167	
H2 (GalNAc)	Phe124	100-76%
	Tyr132	
	Ser134	
	Trp214	
	Gly135	
	Tyr218	
H4 (GalNAc)	Tyr132	100-76%
	Val170	
	Ala167	
	Asp169	
H5 (GalNAc)	Tyr218	75-51%
CH ₃ -NHAc (GalNAc)	Tyr239	51-30%
	His238	
	Trp214	
	Gly136	
	Pro197	
	Ser134	
	Ala137	
	Phe198	

H β -Ala1	Ala167	75-51%
Pro2/6	Tyr239	75-51%
	Tyr218	
H γ -Thr4	Tyr218	100-76%
	Trp214	
Interacciones con UDP		
H2 (GalNAc)	UDP-Gal	-
H3 (GalNAc)		
H4 (GalNAc)		
CH ₃ -NHAc (GalNAc)		
H δ -Arg5		

En ese caso excepcional, la ausencia de UDP-Gal provoca que la Arg5 del péptido interactúe con Glu254, desembocando en un desplazamiento de la molécula de GalNAc de su sitio de unión. De hecho, algunos fotogramas de las simulaciones MD del complejo binario APDTRP-*Dm*C1GalT1 muestran una carencia de puentes de hidrógeno entre el OH3 y OH4 del azúcar y Asp255 (Figura 5.35b). Es por ello que la participación de UDP-Gal en este complejo debe ser la responsable de la estabilización de la carga positiva, impidiendo la interacción de Arg5 con Glu254. En el supuesto contrario, la ausencia de UDP-Gal podría favorecer interacciones no específicas con la proteína, explicando la ausencia de unión de este glicopéptido a la enzima cuando no se añade UDP. Por otro lado, las simulaciones con el glicopéptido P2 mostraban tres interacciones relevantes entre el péptido y la proteína (Figura 5.33^{2b}): durante aproximadamente un 94 % de la trayectoria, se observa un puente de hidrógeno entre la cadena lateral de Trp300 y el grupo carbonilo de Gly5; además de participar en interacciones de tipo CH- π ambas cadenas laterales de Tyr231 y Phe299 con los extremos N y C terminales del péptido, respectivamente. El caso de P7 es bastante similar al de P2, en el que se forma un puente de hidrógeno entre su Gly y Trp300, así como una interacción de tipo CH- π entre su extremo N terminal y Tyr213 (Figura 5.33^{2d}).

Tabla 5.7. Interacciones observadas durante la simulación de MD del complejo formado por P4 y *DmC1GalT1* en presencia de UDP-Gal. Las interacciones medidas fueron las correspondientes a los protones no intercambiables del ligando y la proteína. En la tabla se muestran los contactos con una frecuencia de MD superior al 30% y una distancia inferior a 5 Å.

Complejo P4-<i>DmC1GalT1</i>-UDP		
Protones del ligando	Protones de la proteína	Respuesta de STD-RMN
H2 (GalNAc)	Phe124	100-76%
	Ser134	
	Tyr132	
	Trp214	
	Tyr218	
H4 (GalNAc)	Tyr132	100-76%
	Val170	
	Ala167	
H5 (GalNAc)	Tyr218	100-76%
H6s (GalNAc)	Tyr218	75-51%
	Tyr127	
	Val170	
	Tyr132	
	Ala167	
CH ₃ NHAc (GalNAc)	Tyr239	50-31%
	His238	
	Trp214	
	Pro197	
	Ser134	
	Gly136	
	Phe198	
Pro1	Tyr127	100-51%
	Ala167	
	Val170	
H _β -Ala2	Tyr218	50-31%

	Tyr127	
H _β -Ala3*	Sin contactos significantes	75-51%
H _γ -Thr4	Trp214	100-76%
	Tyr218	
Tyr7*	Sin contactos significantes	100-51%
UDP interactions		
H3 GalNAc	UDP-Gal	-
H4 GalNAc		

Para finalizar, las simulaciones de P4 en complejo con el residuo azucarado y *DmC1GalT1* indican que existe un puente de hidrógeno bastante fuerte entre Gly5 y la cadena lateral de Trp300 ya que, durante toda la trayectoria, aproximadamente un 95% de la población exponía dicha interacción junto a otras interacciones que estabilizaban la proteína con los extremos N y C terminal del péptido (Figura 5.35c). En el mismo caso de este glicopéptido P4, se puede observar una buena correspondencia entre las distancias interprotónicas entre el glicopéptido y la proteína obtenidas a lo largo de las simulaciones MD en presencia de UDP-Gal y las respuestas de STD estimadas para GalNAc, Pro1, Ala2 y Thr4 (Tabla 5.7) para este mismo glicopéptido. Se detectaron adicionalmente contactos temporales cercanos entre Ala3 o Tyr7 con residuos de la proteína a lo largo de toda la trayectoria de simulación de MD, lo cual podría también explicar la respuesta de STD para esos aminoácidos.

6. DISCUSIÓN Y CONCLUSIONES

6.1. Discusión.

El interés que suscita C1GalT1 se debe a su papel indispensable y primordial en la elongación y procesamiento de la *O*-glicosilación proteica de tipo GalNAc en la mayoría de células normales. En la mayoría de procesos tumorales la transcripción de su chaperona específica se encuentra silenciada, dando lugar a una glicosilación aberrante por su mal o inexistente funcionamiento. Gracias a los ensayos y estudios realizados durante esta tesis doctoral se ha conseguido proporcionar más información general y estructural sobre esta enzima y su mecanismo catalítico gracias a la resolución de la estructura cristalina del ortólogo de *Drosophila melanogaster*.

Para llegar a obtener esta información, sin embargo, hubo que sortear una serie de obstáculos, siendo el primero de ellos la expresión y purificación de proteínas recombinantes, el cual es un terreno que puede considerarse sencillo a día de hoy, aunque dependiendo de la proteína, su plegamiento y su localización celular, la dificultad de este proceso puede aumentar desmesuradamente. El segundo obstáculo es la obtención de la estructura de proteínas o complejos de proteínas mediante técnicas de biofísica estructural, y en nuestro caso particular la cristalografía de rayos-X, la cual es una técnica que hace impredecible cualquier proyecto que se lleve a cabo. El tercer obstáculo son las modificaciones postraduccionales y cómo estas son realizadas por cada organismo dependiendo del reino al que pertenezca. Por tanto, es importante elegir un sistema de expresión óptimo para la proteína objetivo dependiendo de las características o rasgos de la misma. La gran variedad de sistemas de expresión unido a la complejidad del proteoma en conjunto crea la necesidad de buscar el sistema de expresión óptimo para cristalizar proteínas, apareciendo la necesidad de un estudio previo de la proteína objetivo para poder decidir racionalmente qué organismo usar para su

expresión. Concretamente, uno de los principales problemas se presenta en las proteínas con sitios de *N* y/o *O*-glicosilación, las cuales suponen una heterogeneidad en las poblaciones de proteína en solución, dificultando el proceso de cristalización. Generalmente, para la cristalización de proteínas se persigue la obtención de una población homogénea de la proteína bajo estudio para que dicha población sea capaz de ordenarse y formar cristales bajo ciertas condiciones específicas. Por esa razón es necesario el uso de enzimas capaces de homogeneizar las poblaciones, cortando los glicanos que son sintetizados. Especialmente en el caso de la *N*-glicosilación, entra en juego el uso de endoglicosidasas e inhibidores que hacen posible el corte y homogenización de los *N*-glicanos. Toda esta reflexión viene a colación por la dificultad que se encaró al intentar expresar una construcción de C1GalT1 estable, cristizable y con un elevado nivel de expresión. En primer lugar, hubo que realizar el diseño de una construcción sin secuencia señal ni dominio transmembrana para poder expresar la enzima de manera soluble y que además fuese capaz de realizar un correcto plegamiento sin la necesidad de Cosmc, usando construcciones de organismos inferiores tomando ventaja de la inexistencia de una chaperona específica en dichos ortólogos. Gracias a esas dos decisiones se consiguió obtener una proteína activa con un elevado nivel de expresión, con la que se pudieron realizar ensayos biofísicos. Sin embargo, al sufrir cierta degradación el primer constructo realizado, no pudo usarse para llevar a cabo experimentos de cristalización a causa de la heterogeneidad de población en solución. El cambio de sistema de expresión del transiente en mamíferos al de baculovirus se realizó con la esperanza de evitar esa degradación, sin resultados positivos. Merece la pena mencionar cómo la puesta a punto del sistema de expresión en insectos mediante baculovirus fue todo un reto en la presente tesis doctoral y consumió a su vez gran parte del tiempo de la misma para su correcto funcionamiento, fusionando diferentes métodos descritos en la bibliografía para su óptimo funcionamiento. No obstante, mediante la predicción de la estructura secundaria de C1GalT1 de *D.*

melanogaster se pudo obtener un constructo en la que no se observaba degradación y al cual se le podían eliminar la mayor parte de los azúcares de los *N*-glicanos, obteniendo por tanto una población homogénea para su posterior cristalización.

Tras la obtención de los cristales se presentó una nueva problemática, pues a pesar de obtener una buena difracción mediante rayos X de los mismos gracias a la línea *microfocus* I24 de DLS en Oxford, no se consiguió resolver el problema de las fases al no existir un modelo lo suficientemente similar para realizar el reemplazamiento molecular. Además, no se tuvo éxito a la hora de reproducir los cristales para llevar a cabo experimentos de faseado, imposibilitando la obtención de la estructura. Fue solo cuando Alphafold2 se liberó públicamente, cuando se pudo obtener un modelo lo suficientemente bueno como para poder obtener la estructura tridimensional en complejo con los sustratos de *DmC1GalT1* mediante reemplazamiento molecular. A partir de ese momento fue cuando pudieron completarse todos los objetivos que se habían propuesto en esta tesis doctoral.

Con anterioridad se había propuesto que la existencia de *Cosmc* en organismos eucariotas superiores era debido a que *C1GalT1* no se encuentra *N*-glicosilada (ni tiene sitios de *N*-glicosilación) mientras que sí que lo está en organismos eucariotas inferiores e invertebrados (Figura 4.10)(Ju et al., 2006). Sin embargo, esto no significa que la ausencia de *N*-glicosilación en las *C1GalT1* de organismos eucariotas superiores tenga que ser necesariamente la razón de la existencia de *Cosmc*, ya que varias glicosiltransferasas humanas que carecen de sitios de *N*-glicosilación sí que se consiguen plegar correctamente sin la necesidad de una chaperona (de las Rivas et al., 2017; Lira-Navarrete et al., 2014). En base a los análisis estructurales realizados, se podría hipotetizar que *Cosmc* tiene que ser importante en la formación de la interfase del homodímero de *C1GalT1* en eucariotas superiores. La interfase del dímero que forma *DmC1GalT1* se puede definir como un gran entramado de interacciones entre ambos monómeros, estando

tan íntimamente relacionados ambos monómeros que partes de un monómero de adentran en el otro y viceversa, protegiendo las cadenas hidrofóbicas participantes en la interfase del medio acuoso. De hecho, esta interfase es tan compleja que Ju. et al. en 2013 propusieron un mecanismo que da explicación a la necesidad de esta chaperona en la formación del dímero de C1GalT1. Según lo propuesto, en el retículo endoplasmático, Cosmc formaría un dímero, uniéndose cada una de las subunidades a un monómero de C1GalT1 y acercándolos entre ellos para permitir la formación del dímero funcional. Cualquier subunidad que no forme un dímero, en principio, tendería a agregar, a causa de la exposición de las regiones hidrofóbicas de la interfase, y a ser degradadas por el proteosoma tras su ubiquitinación. Por otro lado, todas las unidades que formen homodímeros se transportarían hacia el aparato de Golgi. Esta hipótesis podría ser la posible explicación a por qué cuando no se encuentra Cosmc presente, C1GalT1 no pueda realizar su función al ser degradada debido a la agregación y por no poder trasladarse al aparato de Golgi, donde realiza su función. Aunque se haya comentado que la existencia de sitios de *N*-glicosilación en organismos inferiores no necesariamente sea la causa de que C1GalT1 consiga formar los dímeros en estas especies, los resultados invitan a pensar que esa modificación postraduccional podría ayudar a la formación de los homodímeros y a evitar la agregación y degradación de C1GalT1. Aun así, la inexistencia de Cosmc en organismos invertebrados es un misterio que sigue sin explicación y en el que sería necesario realizar más ensayos para poder llegar a una correcta conclusión.

Dejando de lado la importancia de la chaperona específica y centrando la atención en la vía de glicosilación, los resultados que se han obtenido han podido esclarecer la incógnita que supone que el primer paso de iniciación de la *O*-glicosilación de tipo GalNAc la realice la mayor familia de isoenzimas que catalizan un único enlace glicosídico, siendo las principales reguladoras, según se ha descrito, de esta MPT. Esta familia, las GalNAc-Ts, abarcan una amplia variedad de secuencias de sus

sustratos proteicos, en contraposición al paso inmediato de elongación que lo lleva a cabo exclusivamente C1GalT1. Esto supone que la mayor parte de la regulación proviene del primer paso de glicosilación, constatando sin sorpresa alguna, que C1GalT1 tiene una amplísima especificidad de sustrato aceptor y que las interacciones más fuertes tienen lugar con el residuo de GalNAc del sustrato aceptor azucarado, reafirmando el hecho de su amplia especificidad con respecto a la cadena peptídica del sustrato aceptor. A pesar de ello, la identificación de ciertas interacciones entre la enzima y la cadena peptídica es capaz de mostrar cómo en algunos casos puede existir cierta preferencia por la secuencia, optimizando la correcta transferencia del residuo de galactosa, sin ser crítica para la misma. Uno de los mayores ejemplos que demuestra este hecho a parte de los ensayos biofísicos de STD-RMN y la estructura cristalográfica son los ensayos cinéticos, en los que gracias a la secuencia peptídica de los glicopéptidos puede alcanzarse una óptima transferencia de la azúcar. Las características descritas dejan claro que C1GalT1 es capaz de cubrir todo el espectro de *O*-glicanos del glicoproteoma iniciado por las GalNAc-Ts y sintetizar el *core*-1.

Es indiscutible que C1GalT1 puede tener propiedades cinéticas ligeramente diferentes dependiendo del glicopéptido glicosilado con GalNAc, pero en células normales, la inmensa mayoría, por no decir la totalidad, de los *O*-glicanos son elongados para ocultar la exposición del antígeno Tn, estructura asociada a procesos tumorales. La exposición de glicanos aberrantes en la superficie celular de células cancerígenas como son el antígeno Tn suele encontrarse formando una mezcla heterogénea de antígeno Tn y del *core* 1 en la gran mayoría de los casos (Rømer et al., 2021). Esto puede implicar que no se deba, de manera general, a mutaciones en C1GalT1 que causen la inactivación completa del gen que la codifica (Radhakrishnan et al., 2014) sino que podría sugerir que, la expresión reducida de C1GalT1 desemboca en una incompleta elongación de los *O*-glicanos, trasladándose en una preferencia de la transferencia del residuo de galactosa a sitios

de *O*-glicosilación con mejores parámetros cinéticos para C1GalT1.

En células humanas, la síntesis de *O*-glicanos elongados a partir del *core* 1 depende también de la expresión y equilibrio entre C1GalT1 y otras glicosiltransferasas que usan el antígeno Tn como sustrato para formar otras estructuras, como son el *core* 3 (B3GnT6) y el antígeno STn (ST6GalNAcI) (Figura 5.26). La *core* 3 sintasa se encarga de añadir un residuo de GlcNAc al OH3 del GalNAc inicial como se ha mencionado anteriormente, mientras que la ST6GalNAcI transfiere ácido siálico al OH6 del GalNAc del antígeno Tn para formar el antígeno STn. Aunque en la mayoría de células humanas la base más abundante sobre la que se sintetizan los *O*-glicanos complejos es a partir de la elongación del *core* 1, por ejemplo, en células de colon sanas, la mayor parte de los *O*-glicanos complejos se construyen a partir del *core* 3 (Capon et al., 2001). Además, en células caliciformes es interesante observar cómo se acumulan intracelularmente proteínas glicosiladas con el antígeno STn acetilado (Ogata et al., 1995). En el caso de las células de colon son expresadas tanto la *core* 3 sintasa como la C1GalT1, pero la primera se encuentra regulada de manera que su expresión es mayor con respecto a otras líneas celulares, haciendo que la competición entre ambas enzimas por los residuos de GalNAc iniciales se decante a favor de la síntesis del *core* 3. La acumulación intracelular de glicofomas de STn acetilados en células caliciformes es más complicada de entender. Sin embargo, la ST6GalNAcI se expresa selectivamente en el colon, pudiendo competir con C1GalT1, siendo esta última incapaz de usar como sustrato *O*-glicanos con el antígeno STn debido a impedimentos estéricos respaldados con los resultados obtenidos en esta tesis doctoral. Dichos resultados estructurales proporcionan las bases moleculares por las que se rige la glicosilación que realiza C1GalT1, pudiéndose afirmar que es incapaz de realizar la transferencia al antígeno STn por colisiones con Tyr201(Tyr213^{DmC1GalT1}), Tyr206 (Tyr218^{DmC1GalT1}) y Tyr289 (Tyr304^{DmC1GalT1}) de la enzima humana (Figura 5.20 y en concreto la posición del OH6 del residuo de GalNAc).

Gracias a la resolución de la estructura cristalina de diferentes glicosiltransferasas iniciadoras en complejo con sus sustratos aceptores se han conseguido revelar las diferentes estrategias que estas enzimas usan para reconocer los sustratos proteicos, como fue mencionado en la introducción en el caso de las GalNAc-Ts (García-García, Hicks, et al., 2021; Perrine et al., 2009; Valero-González et al., 2016). Sin embargo, pocas estructuras tridimensionales de glicosiltransferasas de seguimiento o *follow-up* han conseguido resolverse, que son las actúan directamente después de la adición del primer monosacárido al *backbone* de la proteína. Tan solo la estructura cristalina de POMGnT1 ha sido publicada previamente, revelando la unión del residuo de manosa a través de puentes de hidrógeno mientras que el péptido es reconocido exclusivamente mediante interacciones hidrofóbicas con la enzima (Kuwabara et al., 2016). En el caso de la C1GalT1, la suma e integración de los datos cristalográficos de rayos X, los resultados obtenidos de STD-RMN y los modelados y simulaciones moleculares permitieron descifrar el reconocimiento de APDT*RP por *DmC1GalT1*. A partir de la inspección visual de la estructura cristalina del complejo, el glicopéptido es reconocido a través del residuo de GalNAc principalmente, mediante un entramado de puentes de hidrógeno en los que se ven involucrados las posiciones OH3, OH4 y OH6 del residuo de GalNAc. Esta exhaustiva inspección se complementó con el espectro de STD-RMN de APDT*RP en presencia de *DmC1GalT1*, permitiendo obtener información sobre los protones de la molécula de GalNAc y enfatizando que son esos mismos protones del residuo de GalNAc los que se encuentran más cercanos al sitio activo de C1GalT1. Todos estos datos refuerzan que el residuo de GalNAc es la parte más importante que reconoce la enzima y que permite la unión entre ambos para realizar la transferencia del residuo de galactosa y la formación del *core 1*.

Con respecto a las interacciones con la secuencia aminoacídica del glicopéptido hay que resaltar que ambas técnicas indican que los residuos de prolina y el grupo metilo de la treonina se ven involucrados en el reconocimiento, ayudando a estabilizar la

unión con una mezcla de interacciones hidrofóbicas y de puentes de hidrógeno. Esto específicamente permite encontrar cierta correlación con el balance energético que hace posible que esta enzima catalice diferentes sustratos. Concretamente, el hallazgo que se consiguió descubrir gracias a los estudios estructurales fue la imposición de una conformación alternada al enlace de α -GalNAc-Thr que no se encuentra de manera natural en la naturaleza, siendo esta conformación típica de los residuos de GalNAc unidos a serina (α -GalNAc-Ser). Al ocurrir esto, la treonina unida a GalNAc se comporta casi completamente igual que si el aminoácido al que estuviese unido fuese una serina, a excepción del grupo metilo de la Thr. Este grupo metilo, interacciona con los residuos aromáticos cercanos del sitio activo, lo que puede permitir que la penalización de energía debida a la conformación desfavorable que se ve obligada a adoptar α -GalNAc-Thr se vea compensada por dichas interacciones con el sitio activo. La imposición de esta conformación a α -GalNAc-Thr, la cual es esencial para conseguir la glicosilación, es la responsable de que C1GalT1 sea capaz de reconocer y realizar la transferencia indistintamente a ambos sustratos aceptores.

Con la obtención de los resultados expuestos podría pensarse que tanto a un nivel estructural como energético es mucho más favorable para C1GalT1 imponer la conformación alternada a α -GalNAc-Thr, aunque apenas se encuentre en solución, que adaptar su sitio activo a las conformaciones que suelen adoptar en solución: la conformación alternada para el caso de α -GalNAc-Ser y la conformación eclipsada en el caso de α -GalNAc-Thr. Una situación parecida en la que la enzima es capaz de inducir una conformación menos favorable al sustrato aceptor es el caso de FUT8, en el cual se impone la conformación anti- ψ al núcleo de chitobiosa de los residuos de GlcNAc del *N*-glicano para conseguir la fucosilación (García-García et al., 2020). Por tanto, tanto el ejemplo con FUT8 como con el descrito de C1GalT1 certifican que existen casos en los que las enzimas seleccionan las conformaciones más inestables de los sustratos aceptores para conseguir la catálisis y no siempre las

conformaciones más estables para realizar la catálisis, en el caso de C1GalT1 consiguiendo una mayor amplitud de sustratos.

En resumen, se ha propuesto que C1GalT1, una enzima de seguimiento de las más fundamentales en las rutas de *O*-glicosilación de tipo GalNAc sigue el mecanismo típico S_N2 descrito para otras glicosiltransferasas de inversión y se ha conseguido revelar las bases moleculares tras el reconocimiento del glicopéptido. También se ha logrado descubrir que C1GalT1 impone una conformación no favorable y de alta energía a los residuos de α-GalNAc-Thr como requerimiento para la elongación de la glicosilación. Con dichos resultados y conclusiones, esta enzima resulta un extraordinario ejemplo de cómo diversas glicosiltransferasas implementan diferentes estrategias para poder promover diferentes cambios conformacionales en los sustratos aceptores y conseguir realizar la transferencia de los residuos azucarados de manera efectiva.

6.2. Conclusiones.

- Mediante estudios bio-informáticos de *DmC1GalT1*, pudimos conseguir una construcción que se expresa en gran cantidad, que es estable y capaz de cristalizar.
- La obtención de la estructura cristalina de *DmC1GalT1* en complejo con UDP-Mn⁺² y APDT*RP, muestra que esta enzima tiene una amplia especificidad de sustrato, siendo el residuo de GalNAc la fuerza motora de la unión.
- Dependiendo de la secuencia peptídica, se ha podido demostrar que la unión previa a UDP puede ayudar a la unión del glicopéptido (caso del APDT*RP) pero esa unión previa no es estrictamente necesaria.
- Mediante mutagénesis dirigida y tras inspeccionar la estructura tridimensional y simulaciones de dinámica molecular, se ha conseguido demostrar que esta enzima sigue el mecanismo típico S_N2 típico de las

glicosiltransferasas de inversión y que la base catalítica de la reacción es un residuo de Asp (D240^{HsC1GalT1} o D255^{DmC1GalT1}).

- C1GalT1 impone una conformación alternada al residuo de α -GalNAc-Thr, más desfavorable energéticamente, siendo esta la forma natural de α -GalNAc-Ser. La penalización energética debida a la conformación inusual de α -GalNAc-Thr se compensa gracias a las interacciones del metilo de Thr con algunos residuos del sitio activo de la enzima, explicando que C1GalT1 pueda glicosilar residuos de Ser o de Thr unidos a GalNAc indistintamente.

7. BIBLIOGRAFÍA

- Adan, A., Alizada, G., Kiraz, Y., Baran, Y., & Nalbant, A. (2017). Flow cytometry: basic principles and applications. *Critical Reviews in Biotechnology*, 37(2), 163–176. <https://doi.org/10.3109/07388551.2015.1128876>
- Agata, N., Ahmad, R., Kawano, T., Raina, D., Kharbanda, S., & Kufe, D. (2008). MUC1 oncoprotein blocks death receptor-mediated apoptosis by inhibiting recruitment of caspase-8. *Cancer Research*, 68(15), 6136–6144. <https://doi.org/10.1158/0008-5472.CAN-08-0464>
- Akasaka-Manyá, K., Manyá, H., Mizuno, M., Inazu, T., & Endo, T. (2011). Effects of length and amino acid sequence of O-mannosyl peptides on substrate specificity of protein O-linked mannose β 1,2-N-acetylglucosaminyltransferase 1 (POMGnT1). *Biochemical and Biophysical Research Communications*, 410(3), 632–636. <https://doi.org/10.1016/j.bbrc.2011.06.042>
- Allen, A. C., Topham, P. S., Harper, S. J., & Feehally, J. (1997). Leucocyte beta 1,3 galactosyltransferase activity in IgA nephropathy. *Nephrology, Dialysis, Transplantation: Official Publication of the European Dialysis and Transplant Association - European Renal Association*, 12(4), 701–706. <https://doi.org/10.1093/NDT/12.4.701>
- Angulo, J., & Nieto, P. M. (2011). STD-NMR: Application to transient interactions between biomolecules—a quantitative approach. In *European Biophysics Journal* (Vol. 40, Issue 12, pp. 1357–1369). <https://doi.org/10.1007/s00249-011-0749-5>
- Apweiler, R., Hermjakob, H., & Sharon, N. (1999). On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochimica et Biophysica Acta - General Subjects*, 1473(1), 4–8. [https://doi.org/10.1016/S0304-4165\(99\)00165-8](https://doi.org/10.1016/S0304-4165(99)00165-8)
- Aryal, R. P., Ju, T., & Cummings, R. D. (2010). The endoplasmic reticulum chaperone cosmc directly promotes in vitro folding of T-synthase. *Journal of Biological Chemistry*, 285(4), 2456–2462. <https://doi.org/10.1074/jbc.M109.065169>
- Aryal, R. P., Ju, T., & Cummings, R. D. (2012). Tight complex formation between cosmc chaperone and its specific client non-native T-synthase leads to enzyme activity and client-driven dissociation. *Journal of Biological Chemistry*, 287(19), 15317–15329. <https://doi.org/10.1074/jbc.M111.312587>
- Aryal, R. P., Ju, T., & Cummings, R. D. (2014). Identification of a novel protein binding motif within the t-synthase for the molecular chaperone cosmc. *Journal of Biological Chemistry*, 289(17), 11630–11641. <https://doi.org/10.1074/jbc.M114.555870>
- Babin, V., Roland, C., & Sagui, C. (2008). Adaptively biased molecular dynamics for free energy calculations. *Journal of Chemical Physics*, 128(13). <https://doi.org/10.1063/1.2844595>
- Bafna, S., Kaur, S., & Batra, S. K. (2010). Membrane-bound mucins: The mechanistic basis for alterations in the growth and survival of cancer cells. *Oncogene*, 29(20), 2893–2904. <https://doi.org/10.1038/onc.2010.87>
- Bansil, R., Celli, J. P., Hardcastle, J. M., & Turner, B. S. (2013). The influence of mucus microstructure and rheology in *Helicobacter pylori* infection. *Frontiers in Immunology*, 4(OCT), 1–12. <https://doi.org/10.3389/fimmu.2013.00310>
- Bansil, R., Stanley, E., & Thomas LaMont, J. (1995). Mucin biophysics. *Annual Review of Physiology*, 57(February), 635–657. <https://doi.org/10.1146/annurev.ph.57.030195.003223>

- Barry, G. F. (1988). A broad-host-range shuttle system for gene insertion into the chromosomes of Gram-negative bacteria. *Gene*, *71*(1), 75–84. [https://doi.org/10.1016/0378-1119\(88\)90079-0](https://doi.org/10.1016/0378-1119(88)90079-0)
- Beerman, I., Novak, J., Wyatt, R. J., Julian, B. A., & Gharavi, A. G. (2007). The genetics of IgA nephropathy. *Nature Clinical Practice. Nephrology*, *3*(6), 325–338. <https://doi.org/10.1038/NCPNEPH0492>
- Bennett, E. P., Mandel, U., Clausen, H., Gerken, T. A., Fritz, T. A., Tabak, L. A., Fu, C., Zhao, H., Wang, Y., Cai, H., Xiao, Y., Zeng, Y., Chen, H., Hu, Y., Feng, J., Wu, F., Lundström, J., Bojar, D., Palm, G. J., ... Hinrichs, W. (2012). Control of mucin-type O-glycosylation: A classification of the polypeptide GalNAc-transferase gene family. *Glycobiology*, *22*(6), 736–756. <https://doi.org/10.1093/GLYCOB/CWR182>
- Berger, E. G. (1999). Tn-syndrome. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, *1455*(2–3), 255–268. [https://doi.org/10.1016/S0925-4439\(99\)00069-1](https://doi.org/10.1016/S0925-4439(99)00069-1)
- Berger, J., & Hinglais, N. (1968). Intercapillary deposits of IgA-IgG. *Journal d'urologie et de Nephrologie*, *74*(0), 694–695.
- Bergstrom, K., Fu, J., Johansson, M. E. V., Liu, X., Gao, N., Wu, Q., Song, J., McDaniel, J. M., McGee, S., Chen, W., Braun, J., Hansson, G. C., & Xia, L. (2017). Core 1- and 3-derived O-glycans collectively maintain the colonic mucus barrier and protect against spontaneous colitis in mice. *Mucosal Immunology*, *10*(1), 91–103. <https://doi.org/10.1038/mi.2016.45>
- Bermejo, I. A., Usabiaga, I., Compañón, I., Castro-López, J., Insausti, A., Fernández, J. A., Avenzoa, A., Busto, J. H., Jiménez-Barbero, J., Asensio, J. L., Peregrina, J. M., Jiménez-Osés, G., Hurtado-Guerrero, R., Cocinero, E. J., & Corzana, F. (2018). Water Sculpts the Distinctive Shapes and Dynamics of the Tumor-Associated Carbohydrate Tn Antigens: Implications for Their Molecular Recognition. *Journal of the American Chemical Society*, *140*(31), 9952–9960. <https://doi.org/10.1021/jacs.8b04801>
- BioRAD. (2012). *Trans-Blot Turbo™ Transfer System*. Cm, 4–5. www.bio-rad.com.
- Brockhausen, I., Möller, G., Pollex-Krüger, A., Rutz, V., Paulsen, H., & Matta, K. L. (1992). Control of O-glycan synthesis: specificity and inhibition of O-glycan core 1 UDP-galactose:N-acetylgalactosamine- α -R beta 3-galactosyltransferase from rat liver. *Biochemistry and Cell Biology*, *70*(2), 99–108. <https://doi.org/10.1139/o92-015>
- Brockhausen, I., Wandall, H. H., Hagen, K. G. ten, & Stanley, P. (2022). O-GalNAc Glycans. In *Essentials of Glycobiology*. Cold Spring Harbor Laboratory Press. <https://doi.org/10.1101/GLYCOBIOLOGY.4E.10>
- Brooks, S. A., Carter, T. M., Bennett, E. P., Clausen, H., & Mandel, U. (2007). Immunolocalisation of members of the polypeptide N-acetylgalactosaminyl transferase (ppGalNAc-T) family is consistent with biologically relevant altered cell surface glycosylation in breast cancer. *Acta Histochemica*, *109*(4), 273–284. <https://doi.org/10.1016/j.acthis.2007.02.009>
- Brunelle, J. L., & Green, R. (2014). One-dimensional SDS-polyacrylamide gel electrophoresis (1D SDS-PAGE). *Methods in Enzymology*, *541*, 151–159. <https://doi.org/10.1016/B978-0-12-420119-4.00012-4>
- Bujacz, G., Wrzesniewska, B., & Bujacz, A. (2010). Cryoprotection properties of salts of organic acids: A case study for a tetragonal crystal of HEW lysozyme. *Acta Crystallographica Section D: Biological Crystallography*, *66*(7), 789–796. <https://doi.org/10.1107/S0907444910015416>

- Cao, Y., Stosiek, P., Springer, G. F., & Karsten, U. (1996). Thomsen-Friedenreich-related carbohydrate antigens in normal adult human tissues: A systematic and comparative study. *Histochemistry and Cell Biology*, *106*(2), 197–207. <https://doi.org/10.1007/BF02484401>
- Capon, C., Maes, E., Michalski, J. C., Leffler, H., & Kim, Y. S. (2001). Sd(a)-antigen-like structures carried on core 3 are prominent features of glycans from the mucin of normal human descending colon. *Biochemical Journal*, *358*(3), 657–664. <https://doi.org/10.1042/bj3580657>
- Chan, J., Tang, A., & Bennet, A. J. (2012). A stepwise solvent-promoted S_Ni reaction of α -d-glucopyranosyl fluoride: Mechanistic implications for retaining glycosyltransferases. *Journal of the American Chemical Society*, *134*(2), 1212–1220. <https://doi.org/10.1021/ja209339j>
- Chaturvedi, P., Singh, A. P., Chakraborty, S., Chauhan, S. C., Bafna, S., Meza, J. L., Singh, P. K., Hollingsworth, M. A., Mehta, P. P., & Batra, S. K. (2008). MUC4 Mucin Interacts with and Stabilizes the HER2 Oncoprotein in Human Pancreatic Cancer Cells. *Cancer Research*, *68*(7), 2065. <https://doi.org/10.1158/0008-5472.CAN-07-6041>
- Chia, J., Goh, G., & Bard, F. (2016). Short O-GalNAc glycans: Regulation and role in tumor development and clinical perspectives. *Biochimica et Biophysica Acta - General Subjects*, *1860*(8), 1623–1639. <https://doi.org/10.1016/j.bbagen.2016.03.008>
- Chugh, S., Gnanapragassam, V. S., Jain, M., Rachagani, S., Ponnusamy, M. P., & Batra, S. K. (2015). Pathobiological implications of mucin glycans in cancer: Sweet poison and novel targets. *Biochimica et Biophysica Acta - Reviews on Cancer*, *1856*(2), 211–225. <https://doi.org/10.1016/j.bbcan.2015.08.003>
- Clement, E. M., Godfrey, C., Tan, J., Brockington, M., Torelli, S., Feng, L., Brown, S. C., Jimenez-Mallebrera, C., Sewry, C. A., Longman, C., Mein, R., Abbs, S., Vajsar, J., Schachter, H., & Muntoni, F. (2008). Mild POMGnT1 mutations underlie a novel limb-girdle muscular dystrophy variant. *Archives of Neurology*, *65*(1), 137–141. <https://doi.org/10.1001/archneurol.2007.2>
- Collaborative Computational Project, N. 4. (1984). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr*, *50*, 760–763.
- Cong, L., & Zhang, F. (2015). Genome Engineering Using CRISPR-Cas9 System. In S. M. Pruetz-Miller (Ed.), *Chromosomal Mutagenesis* (pp. 197–217). Springer New York. https://doi.org/10.1007/978-1-4939-1862-1_10
- Conibear, A. C. (2020). Deciphering protein post-translational modifications using chemical biology tools. *Nature Reviews Chemistry*, *4*(12), 674–695. <https://doi.org/10.1038/s41570-020-00223-8>
- Corzana, F., Busto, J. H., Jiménez-Osés, G., Asensio, J. L., Jiménez-Barbero, J., Peregrina, J. M., & Avenoza, A. (2006). New insights into α -GalNAc-Ser motif: Influence of hydrogen bonding versus solvent interactions on the preferred conformation. *Journal of the American Chemical Society*, *128*(45), 14640–14648. <https://doi.org/10.1021/ja064539u>
- Corzana, F., Busto, J. H., Jiménez-Osés, G., de Luis, M. G., Asensio, J. L., Jiménez-Barbero, J., Peregrina, J. M., & Avenoza, A. (2007). Serine versus threonine glycosylation: The methyl group causes a drastic alteration on the carbohydrate orientation and on the surrounding water shell. *Journal of the American Chemical Society*, *129*(30), 9458–9467. <https://doi.org/10.1021/ja072181b>

- Creighton, T. E. (1997). *Protein Structure: A Practical Approach*. IRL Press at Oxford University Press. <https://books.google.com/cu/books?id=KE-WlxzOxF0C>
- Crew, V. K., Singleton, B. K., Green, C., Parsons, S. F., Daniels, G., & Anstee, D. J. (2008). New mutations in C1GALT1C1 in individuals with Tn positive phenotype. *British Journal of Haematology*, *142*(4), 657–667. <https://doi.org/10.1111/j.1365-2141.2008.07215.x>
- Cui, Y., Xu, J., Cheng, M., Liao, X., & Peng, S. (2018). Review of CRISPR/Cas9 sgRNA Design Tools. *Interdisciplinary Sciences: Computational Life Sciences* *2018 10:2*, *10*(2), 455–465. <https://doi.org/10.1007/S12539-018-0298-Z>
- Cummings, R. D. (2019). “Stuck on sugars – how carbohydrates regulate cell adhesion, recognition, and signaling.” *Glycoconjugate Journal*, *36*(4), 241–257. <https://doi.org/10.1007/s10719-019-09876-0>
- Cummings, R. D., Chiffolleau, E., Kyook, Y. van, & McEver, R. P. (2022). C-Type Lectins. *Essentials of Glycobiology*. <https://doi.org/10.1101/GLYCOBIOLOGY.4E.34>
- Cummings, R. D., Liu, F.-T., Rabinovich, G. A., Stowell, S. R., & Vasta, G. R. (2022). Galectins. *Essentials of Glycobiology*. <https://doi.org/10.1101/GLYCOBIOLOGY.4E.36>
- D’Amico, G. (2000). Natural history of idiopathic IgA nephropathy: role of clinical and histological prognostic factors. *American Journal of Kidney Diseases: The Official Journal of the National Kidney Foundation*, *36*(2), 227–237. <https://doi.org/10.1053/AJKD.2000.8966>
- Dam, T. K., & Fred Brewer, C. (2007). Applications of Isothermal Titration Calorimetry to Lectin-Carbohydrate Interactions. *Lectins: Analytical Technologies*, 75–101. <https://doi.org/10.1016/B978-044453077-6/50005-3>
- Davies, D. R., & Segal, D. M. B. T. (1971). Protein crystallization: Micro techniques involving vapor diffusion. In *Methods in Enzymology* (Vol. 22, pp. 266–269). Academic Press. [https://doi.org/https://doi.org/10.1016/0076-6879\(71\)22027-9](https://doi.org/https://doi.org/10.1016/0076-6879(71)22027-9)
- Davies, G. J. (2001). Sweet secrets of synthesis. *Nature Structural Biology*, *8*(2), 98–100. <https://doi.org/10.1038/84198>
- de las Rivas, M., Coelho, H., Diniz, A., Lira-Navarrete, E., Compañón, I., Jiménez-Barbero, J., Schjoldager, K. T., Bennett, E. P., Vakhrushev, S. Y., Clausen, H., Corzana, F., Marcelo, F., & Hurtado-Guerrero, R. (2018). Structural Analysis of a GalNAc-T2 Mutant Reveals an Induced-Fit Catalytic Mechanism for GalNAc-Ts. *Chemistry - A European Journal*, *24*(33), 8382–8392. <https://doi.org/10.1002/chem.201800701>
- de las Rivas, M., Lira-Navarrete, E., Daniel, E. J. P., Compañón, I., Coelho, H., Diniz, A., Jiménez-Barbero, J., Peregrina, J. M., Clausen, H., Corzana, F., Marcelo, F., Jiménez-Osés, G., Gerken, T. A., & Hurtado-Guerrero, R. (2017). The interdomain flexible linker of the polypeptide GalNAc transferases dictates their long-range glycosylation preferences. *Nature Communications*, *8*(1). <https://doi.org/10.1038/s41467-017-02006-0>
- de las Rivas, M., Lira-Navarrete, E., Gerken, T. A., & Hurtado-Guerrero, R. (2019). Polypeptide GalNAc-Ts: from redundancy to specificity. In *Current Opinion in Structural Biology* (Vol. 56, pp. 87–96). Elsevier Ltd. <https://doi.org/10.1016/j.sbi.2018.12.007>
- de las Rivas, M., Paul Daniel, E. J., Coelho, H., Lira-Navarrete, E., Raich, L., Compañón, I., Diniz, A., Lagartera, L., Jiménez-Barbero, J., Clausen, H., Rovira, C., Marcelo, F., Corzana, F.,

- Gerken, T. A., & Hurtado-Guerrero, R. (2018). Structural and Mechanistic Insights into the Catalytic-Domain-Mediated Short-Range Glycosylation Preferences of GalNAc-T4. *ACS Central Science*, 4(9), 1274–1290. <https://doi.org/10.1021/acscentsci.8b00488>
- de las Rivas, M., Paul Daniel, E. J., Narimatsu, Y., Compañón, I., Kato, K., Hermosilla, P., Thureau, A., Ceballos-Laita, L., Coelho, H., Bernadó, P., Marcelo, F., Hansen, L., Maeda, R., Lostao, A., Corzana, F., Clausen, H., Gerken, T. A., & Hurtado-Guerrero, R. (2020). Molecular basis for fibroblast growth factor 23 O-glycosylation by GalNAc-T3. *Nature Chemical Biology*, 16(3), 351–360. <https://doi.org/10.1038/s41589-019-0444-x>
- Deng, J., Wang, L., Chen, H., Li, L., Ma, Y., Ni, J., & Li, Y. (2013). The role of tumour-associated MUC1 in epithelial ovarian cancer metastasis and progression. In *Cancer and Metastasis Reviews* (Vol. 32, Issues 3–4, pp. 535–551). *Cancer Metastasis Rev.* <https://doi.org/10.1007/s10555-013-9423-y>
- Drozdetskiy, A., Cole, C., Procter, J., & Barton, G. J. (2015). JPred4: a protein secondary structure prediction server. *Nucleic Acids Research*, 43(W1), W389–W394. <https://doi.org/10.1093/NAR/GKV332>
- Drula, E., Garron, M. L., Dogan, S., Lombard, V., Henrissat, B., & Terrapon, N. (2022). The carbohydrate-active enzyme database: Functions and literature. *Nucleic Acids Research*, 50(D1), D571–D577. <https://doi.org/10.1093/nar/gkab1045>
- Elbein, A. D., Tropea, J. E., Mitchell, M., & Kaushal, G. P. (1990). Kifunensine, a potent inhibitor of the glycoprotein processing mannosidase I. *Journal of Biological Chemistry*, 265(26), 15599–15605. [https://doi.org/10.1016/S0021-9258\(18\)55439-9](https://doi.org/10.1016/S0021-9258(18)55439-9)
- Ellies, L. G., Tsuboi, S., Petryniak, B., Lowe, J. B., Fukuda, M., & Marth, J. D. (1998). Core 2 oligosaccharide biosynthesis distinguishes between selectin ligands essential for leukocyte homing and inflammation. *Immunity*, 9(6), 881–890. [https://doi.org/10.1016/S1074-7613\(00\)80653-6](https://doi.org/10.1016/S1074-7613(00)80653-6)
- Emsley, P., & Cowtan, K. (2004). Coot: Model-building tools for molecular graphics. *Acta Crystallographica Section D: Biological Crystallography*, 60(12 I), 2126–2132. <https://doi.org/10.1107/S0907444904019158>
- Ferentz, A. E., & Wagner, G. (2000). NMR spectroscopy: A multifaceted approach to macromolecular structure. In *Quarterly Reviews of Biophysics* (Vol. 33, Issue 1, pp. 29–65). <https://doi.org/10.1017/S0033583500003589>
- Friedenreich, V. (1931). The Thomsen Hemagglutination Phenomenon: Production of a Specific Receptor Quality in Red Corpuscles by Bacterial Activity. [Book]. *Journal of the American Medical Association*, 97(17), 1249. <https://doi.org/10.1001/jama.1931.02730170061042>
- Fu, C., Zhao, H., Wang, Y., Cai, H., Xiao, Y., Zeng, Y., & Chen, H. (2016). Tumor-associated antigens: Tn antigen, sTn antigen, and T antigen. *Hla*, 88(6), 275–286. <https://doi.org/10.1111/tan.12900>
- Fu, J., Gerhardt, H., McDaniel, J. M., Xia, B., Liu, X., Ivanciu, L., Ny, A., Hermans, K., Silasi-Mansat, R., McGee, S., Nye, E., Ju, T., Ramirez, M. I., Carmeliet, P., Cummings, R. D., Lupu, F., & Xia, L. (2008). Endothelial cell O-glycan deficiency causes blood/lymphatic misconnections and consequent fatty liver disease in mice. *The Journal of Clinical Investigation*, 118(11), 3725–3737. <https://doi.org/10.1172/JCI36077>

- Furukawa, K., & Roth, S. (1985). Co-purification of galactosyltransferases from chick-embryo liver. *Biochemical Journal*, 227(2), 573–582. <https://doi.org/10.1042/bj2270573>
- García-García, A., Ceballos-Laita, L., Serna, S., Artschwager, R., Reichardt, N. C., Corzana, F., & Hurtado-Guerrero, R. (2020). Structural basis for substrate specificity and catalysis of α 1,6-fucosyltransferase. *Nature Communications*, 11(1). <https://doi.org/10.1038/s41467-020-14794-z>
- García-García, A., Hicks, T., el Qaidi, S., Zhu, C., Hardwidge, P. R., Angulo, J., & Hurtado-Guerrero, R. (2021). NleB/SseK-catalyzed arginine-glycosylation and enteropathogen virulence are finely tuned by a single variable position contiguous to the catalytic machinery. *Chemical Science*, 12(36), 12181–12191. <https://doi.org/10.1039/d1sc04065k>
- García-García, A., Serna, S., Yang, Z., Delso, I., Taleb, V., Hicks, T., Artschwager, R., Vakhrushev, S. Y., Clausen, H., Angulo, J., Corzana, F., Reichardt, N. C., & Hurtado-Guerrero, R. (2021). FUT8-Directed Core Fucosylation of N-glycans Is Regulated by the Glycan Structure and Protein Environment. *ACS Catalysis*, 11(15), 9052–9065. <https://doi.org/10.1021/acscatal.1c01698>
- Gerken, T. A., Jamison, O., Perrine, C. L., Collette, J. C., Moinova, H., Ravi, L., Markowitz, S. D., Shen, W., Patel, H., & Tabak, L. A. (2011). Emerging paradigms for the initiation of mucin-type protein O-glycosylation by the polypeptide GalNAc transferase family of glycosyltransferases. *Journal of Biological Chemistry*, 286(16), 14493–14507. <https://doi.org/10.1074/jbc.M111.218701>
- Gerken, T. A., Revoredo, L., Thome, J. J. C., Tabak, L. A., Vester-Christensen, M. B., Clausen, H., Gahlay, G. K., Jarvis, D. L., Johnson, R. W., Moniz, H. A., & Moremen, K. (2013). The lectin domain of the polypeptide GalNAc transferase family of glycosyltransferases (ppGalNAc Ts) acts as a switch directing glycopeptide substrate glycosylation in an N- or C-terminal direction, further controlling mucin type O-Glycosylation. *Journal of Biological Chemistry*, 288(27), 19900–19914. <https://doi.org/10.1074/jbc.M113.477877>
- Gill, D. J., Clausen, H., & Bard, F. (2011). Location, location, location: New insights into O-GalNAc protein glycosylation. *Trends in Cell Biology*, 21(3), 149–158. <https://doi.org/10.1016/j.tcb.2010.11.004>
- Gollamudi, S., Lekhraj, R., Lalezari, S., & Lalezari, P. (2020). COSMC mutations reduce T-synthase activity in advanced Alzheimer’s disease. *Alzheimer’s & Dementia : Translational Research & Clinical Interventions*, 6(1). <https://doi.org/10.1002/TRC2.12040>
- Granovsky, M., Bielfeldt, T., Peters, S., Paulsen, H., Meldal, M., Brockhausen, J., & Brockhausen, I. (1994). UDP-galactose: glycoprotein-N-acetyl-d-galactosamine 3- β -d-galactosyltransferase activity synthesizing O-glycan core 1 is controlled by the amino acid sequence and glycosylation of glycopeptide substrates. *European Journal of Biochemistry*, 221(3), 1039–1046. <https://doi.org/10.1111/j.1432-1033.1994.tb18822.x>
- Gubbels, J. A. A., Felder, M., Horibata, S., Belisle, J. A., Kapur, A., Holden, H., Petrie, S., Migneault, M., Rancourt, C., Connor, J. P., & Patankar, M. S. (2010). MUC16 provides immune protection by inhibiting synapse formation between NK and ovarian tumor cells. *Molecular Cancer*, 9. <https://doi.org/10.1186/1476-4598-9-11>
- Hagen, F. K., Raffo, R., DeSa, D., & Tabak, L. A. (1999). Structure-function analysis of the UDP-N-acetyl-D-galactosamine: Polypeptide N-acetylgalactosaminyltransferase. Essential

- residues lie in a predicted active site cleft resembling a lactose repressor fold. *Journal of Biological Chemistry*, 274(10), 6797–6803. <https://doi.org/10.1074/jbc.274.10.6797>
- Hakomori, S. I. (2001). Tumor-associated carbohydrate antigens defining tumor malignancy: Basis for development of anti-cancer vaccines. *Advances in Experimental Medicine and Biology*, 491, 369–402. https://doi.org/10.1007/978-1-4615-1267-7_24
- Hanes, M. S., Moremen, K. W., & Cummings, R. D. (2017). Biochemical characterization of functional domains of the chaperone Cosmc. *PLoS ONE*, 12(6). <https://doi.org/10.1371/journal.pone.0180242>
- Hang, H. C., & Bertozzi, C. R. (2005). The chemistry and biology of mucin-type O-linked glycosylation. *Bioorganic and Medicinal Chemistry*, 13(17), 5021–5034. <https://doi.org/10.1016/j.bmc.2005.04.085>
- Hanisch, F. G. (2001). O-glycosylation of the mucin type. *Biological Chemistry*, 382(2), 143–149. <https://doi.org/10.1515/BC.2001.022>
- Hanukoglu, I. (2015). Proteopedia: Rossmann fold: A beta-alpha-beta fold at dinucleotide binding sites. *Biochemistry and Molecular Biology Education*, 43(3), 206–209. <https://doi.org/10.1002/bmb.20849>
- Hao, Y., Créquer-Grandhomme, A., Javier, N., Singh, A., Chen, H., Manzanillo, P., Lo, M. C., & Huang, X. (2021). Structures and mechanism of human glycosyltransferase β 1,3-N-acetylglucosaminyltransferase 2 (B3GNT2), an important player in immune homeostasis. *Journal of Biological Chemistry*, 296. <https://doi.org/10.1074/JBC.RA120.015306>
- Harrison, R. L., & Jarvis, D. L. (2006). Protein N-Glycosylation in the Baculovirus-Insect Cell Expression System and Engineering of Insect Cells to Produce “Mammalianized” Recombinant Glycoproteins. In *Advances in Virus Research* (Vol. 68, pp. 159–191). [https://doi.org/10.1016/S0065-3527\(06\)68005-6](https://doi.org/10.1016/S0065-3527(06)68005-6)
- Haugstad, K. E., Gerken, T. A., Stokke, B. T., Dam, T. K., Brewer, C. F., & Sletmoen, M. (2012). Enhanced self-association of mucins possessing the T and Tn carbohydrate cancer antigens at the single-molecule level. *Biomacromolecules*, 13(5), 1400–1409. <https://doi.org/10.1021/bm300135h>
- Haugstad, K. E., Hadjilirezaei, S., Stokke, B. T., Brewer, C. F., Gerken, T. A., Burchell, J., Picco, G., & Sletmoen, M. (2016). Interactions of mucins with the Tn or Sialyl Tn cancer antigens including MUC1 are due to GalNAc–GalNAc interactions. *Glycobiology*, 26(12), 1338. <https://doi.org/10.1093/GLYCOB/CWW065>
- Häuselmann, I., & Borsig, L. (2014). Altered tumor-cell glycosylation promotes metastasis. In *Frontiers in Oncology: Vol. 4 MAR*. Frontiers Research Foundation. <https://doi.org/10.3389/fonc.2014.00028>
- Hazes, B. (1996). The (QxW)₃ domain: A flexible lectin scaffold. *Protein Science*, 5(8), 1490–1501. <https://doi.org/10.1002/pro.5560050805>
- Hiki, Y. (2009). O-linked oligosaccharides of the IgA1 hinge region: roles of its aberrant structure in the occurrence and/or progression of IgA nephropathy. *Clinical and Experimental Nephrology*, 13(5), 415–423. <https://doi.org/10.1007/S10157-009-0173-7>
- Hinoda, Y., Ikematsu, Y., Horinouchi, M., Sato, S., Yamamoto, K., Nakano, T., Fukui, M., Suehiro,

- Y., Hamanaka, Y., Nishikawa, Y., Kida, H., Waki, S., Oka, M., Imai, K., & Yonezawa, S. (2003). Increased expression of MUC1 in advanced pancreatic cancer. *Journal of Gastroenterology*, *38*(12), 1162–1166. <https://doi.org/10.1007/s00535-003-1224-6>
- Hollingsworth, M. A., & Swanson, B. J. (2004). Mucins in cancer: Protection and control of the cell surface. *Nature Reviews Cancer*, *4*(1), 45–60. <https://doi.org/10.1038/nrc1251>
- Holm, L., & Laakso, L. M. (2016). Dali server update. *Nucleic Acids Research*, *44*(1), W351–W355. <https://doi.org/10.1093/NAR/GKW357>
- Holt, G. D., Haltiwanger, R. S., Torres, C. R., & Hart, G. W. (1987). Erythrocytes contain cytoplasmic glycoproteins. O-linked GlcNac on band 4.1. *Journal of Biological Chemistry*, *262*(31), 14847–14850. [https://doi.org/10.1016/s0021-9258\(18\)48100-8](https://doi.org/10.1016/s0021-9258(18)48100-8)
- Homeister, J. W., Thall, A. D., Petryniak, B., Malý, P., Rogers, C. E., Smith, P. L., Kelly, R. J., Gersten, K. M., Askari, S. W., Cheng, G., Smithson, G., Marks, R. M., Misra, A. K., Hindsgaul, O., von Andrian, U. H., & Lowe, J. B. (2001). The alpha(1,3)fucosyltransferases FucT-IV and FucT-VII exert collaborative control over selectin-dependent leukocyte recruitment and lymphocyte homing. *Immunity*, *15*(1), 115–126. [https://doi.org/10.1016/S1074-7613\(01\)00166-2](https://doi.org/10.1016/S1074-7613(01)00166-2)
- Horvath, P., & Barrangou, R. (2010). CRISPR/Cas, the immune system of Bacteria and Archaea. *Science*, *327*(5962), 167–170. <https://doi.org/10.1126/science.1179555>
- Ho, W. L., Hsu, W. M., Huang, M. C., Kadomatsu, K., & Nakagawara, A. (2016). Protein glycosylation in cancers and its potential therapeutic applications in neuroblastoma. *Journal of Hematology and Oncology*, *9*(1), 1–15. <https://doi.org/10.1186/s13045-016-0334-6>
- Hurtado-Guerrero, R. (2016). Recent structural and mechanistic insights into protein O-GalNAc glycosylation. *Biochemical Society Transactions*, *44*, 61–67. <https://doi.org/10.1042/BST20150178>
- Hurtado-Guerrero, R., & Davies, G. J. (2012). Recent structural and mechanistic insights into post-translational enzymatic glycosylation. *Current Opinion in Chemical Biology*, *16*(5–6), 479–487. <https://doi.org/10.1016/j.cbpa.2012.10.013>
- Imberty, A., Piller, V., Piller, F., & Breton, C. (1997). Fold recognition and molecular modeling of a lectin-like domain in UDP-GalNAc:Polypeptide N-acetylgalactosaminyltransferases. *Protein Engineering*, *10*(12), 1353–1356. <https://doi.org/10.1093/protein/10.12.1353>
- Inoue, T., Sugiyama, H., Hiki, Y., Takiue, K., Morinaga, H., Kitagawa, M., Maeshima, Y., Fukushima, K., Nishizaki, K., Akagi, H., Narimatsu, Y., Narimatsu, H., & Makino, H. (2010). Differential expression of glycogenes in tonsillar B lymphocytes in association with proteinuria and renal dysfunction in IgA nephropathy. *Clinical Immunology (Orlando, Fla.)*, *136*(3), 447–455. <https://doi.org/10.1016/J.CLIM.2010.05.009>
- Invitrogen. (2015). Bac-to-Bac® Baculovirus Expression System. *User Guide, August*, 1–78. http://tools.thermofisher.com/content/sfs/manuals/bactobac_man.pdf
- Irimura, T., Denda, K., Iida, S. I., Takeuchi, H., & Kato, K. (1999). Diverse glycosylation of MUC1 and MUC2: potential significance in tumor immunity. *Journal of Biochemistry*, *126*(6), 975–985. <https://doi.org/10.1093/OXFORDJOURNALS.JBCHEM.A022565>
- Iwai, T., Inaba, N., Naundorf, A., Zhang, Y., Gotoh, M., Iwasaki, H., Kudo, T., Togayachi, A.,

- Ishizuka, Y., Nakanishi, H., & Narimatsu, H. (2002). Molecular cloning and characterization of a novel UDP-GlcNAc: GalNAc-peptide β 1,3-N-acetylglucosaminyltransferase (β 3Gn-T6), an enzyme synthesizing the core 3 structure of O-glycans. *Journal of Biological Chemistry*, 277(15), 12802–12809. <https://doi.org/10.1074/jbc.M112457200>
- Jentoft, N. (1990). Why are proteins O-glycosylated? *Trends in Biochemical Sciences*, 15(8), 291–294.
- Jinek, M., Chen, Y. W., Clausen, H., Cohen, S. M., & Conti, E. (2006). Structural insights into the Notch-modifying glycosyltransferase Fringe. *Nature Structural and Molecular Biology*, 13(10), 945–946. <https://doi.org/10.1038/nsmb1144>
- Joshi, H. J., Narimatsu, Y., Schjoldager, K. T., Tytgat, H. L. P., Aebi, M., Clausen, H., & Halim, A. (2018). SnapShot: O-Glycosylation Pathways across Kingdoms. In *Cell* (Vol. 172, Issue 3, pp. 632–632.e2). Cell Press. <https://doi.org/10.1016/j.cell.2018.01.016>
- Julian, B. A., & Novak, J. (2004). IgA nephropathy: an update. *Current Opinion in Nephrology and Hypertension*, 13(2), 171–179. <https://doi.org/10.1097/00041552-200403000-00005>
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Židek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., ... Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583–589. <https://doi.org/10.1038/s41586-021-03819-2>
- Ju, T., Aryal, R. P., Kudelka, M. R., Wang, Y., & Cummings, R. D. (2014). The Cosmc connection to the Tn antigen in cancer. *Cancer Biomarkers*, 14(1), 63–81. <https://doi.org/10.3233/CBM-130375>
- Ju, T., Aryal, R. P., Stowell, C. J., & Cummings, R. D. (2008). Regulation of protein O-glycosylation by the endoplasmic reticulum-localized molecular chaperone Cosmc. *Journal of Cell Biology*, 182(3), 531–542. <https://doi.org/10.1083/jcb.200711151>
- Ju, T., Brewer, K., Souza, A. D., Cummings, R. D., & Canfield, W. M. (2002). Cloning and expression of human core 1 β 1,3-galactosyltransferase. *Journal of Biological Chemistry*, 277(1), 178–186. <https://doi.org/10.1074/jbc.M109060200>
- Ju, T., & Cummings, R. D. (2002). A unique molecular chaperone cosmc required for activity of the mammalian core 1 β -galactosyltransferase. *Proceedings of the National Academy of Sciences of the United States of America*, 99(26), 16613–16618. <https://doi.org/10.1073/pnas.262438199>
- Ju, T., & Cummings, R. D. (2005). Protein Glycosylation: chaperone mutation in Tn syndrome. *Nature*, 437(7063), 1252. <https://doi.org/10.1038/4371252a>
- Ju, T., & Cummings, R. D. (2010). Functional assays for the molecular chaperone cosmc. In *Methods in Enzymology* (Vol. 479, Issue C, pp. 107–122). Academic Press Inc. [https://doi.org/10.1016/S0076-6879\(10\)79006-6](https://doi.org/10.1016/S0076-6879(10)79006-6)
- Ju, T., & Cummings, R. D. (2014). Core 1 β 3Galactosyltransferase (C1GalT1, T-Synthase) and its specific molecular chaperone COSMC (C1GalT1C1). In N. Taniguchi, K. Honke, M. Fukuda, H. Narimatsu, Y. Yamaguchi, & T. Angata (Eds.), *Handbook of Glycosyltransferases and Related Genes* (pp. 149–169). Springer Japan. https://doi.org/10.1007/978-4-431-54240-7_65

- Ju, T., Cummings, R. D., & Canfield, W. M. (2002). Purification, characterization, and subunit structure of rat core 1 β 1,3-galactosyltransferase. *Journal of Biological Chemistry*, 277(1), 169–177. <https://doi.org/10.1074/jbc.M109056200>
- Ju, T., Otto, V. I., & Cummings, R. D. (2011). The Tn antigen-structural simplicity and biological complexity. In *Angewandte Chemie - International Edition* (Vol. 50, Issue 8, pp. 1770–1791). <https://doi.org/10.1002/anie.201002313>
- Ju, T., Wang, Y., Aryal, R. P., Lehoux, S. D., Ding, X., Kudelka, M. R., Cutler, C., Zeng, J., Wang, J., Sun, X., Heimbürg-Molinari, J., Smith, D. F., & Cummings, R. D. (2013). Tn and sialyl-Tn antigens, aberrant O-glycomics as human disease markers. *PROTEOMICS – Clinical Applications*, 7(9–10), 618–631. <https://doi.org/https://doi.org/10.1002/prca.201300024>
- Ju, T., Zheng, Q., & Cummings, R. D. (2006). Identification of core 1 O-glycan T-synthase from *Caenorhabditis elegans*. *Glycobiology*, 16(10), 947–958. <https://doi.org/10.1093/glycob/cw1008>
- Kabsch, W. (2010). XDS. *Acta Crystallographica Section D Biological Crystallography*, 66(2), 125–132. <https://doi.org/10.1107/S09074444909047337>
- Kadirvelraj, R., Yang, J. Y., Kim, H. W., Sanders, J. H., Moremen, K. W., & Wood, Z. A. (2021). Comparison of human poly-N-acetyl-lactosamine synthase structure with GT-A fold glycosyltransferases supports a modular assembly of catalytic subsites. *Journal of Biological Chemistry*, 296, 100110. <https://doi.org/10.1074/jbc.RA120.015305>
- Khoury, G. A., Baliban, R. C., & Floudas, C. A. (2011). Proteome-wide post-translational modification statistics: Frequency analysis and curation of the swiss-prot database. *Scientific Reports*, 1, 1–5. <https://doi.org/10.1038/srep00090>
- Klenk, E., & Uhlenbruck, G. (1960). On the isolation of mucoids containing neuraminic acid from human erythrocyte stroma, a contribution to the chemistry of agglutinogens. *Hoppe-Seyler's Zeitschrift Für Physiologische Chemie*. <https://doi.org/doi:10.1515/bchm2.1960.319.1.151>
- Kudelka, M. R., Ju, T., Heimbürg-Molinari, J., & Cummings, R. D. (2015). Simple sugars to complex disease-mucin-type O-glycans in cancer. In *Advances in Cancer Research* (Vol. 126, pp. 53–135). Academic Press Inc. <https://doi.org/10.1016/bs.acr.2014.11.002>
- Kudelka, M. R., Stowell, S. R., Cummings, R. D., & Neish, A. S. (2020). Intestinal epithelial glycosylation in homeostasis and gut microbiota interactions in IBD. *Nature Reviews. Gastroenterology & Hepatology*, 17(10), 597–617. <https://doi.org/10.1038/S41575-020-0331-7>
- Kudot, T., Iwai, T., Kubota, T., Iwasaki, H., Takayma, Y., Hiruma, T., Inaba, N., Zhang, Y., Gotoh, M., Togayachi, A., & Narimatsu, H. (2002). Molecular Cloning and Characterization of a Novel UDP-Gal:GalNAc α Peptide β 1,3-Galactosyltransferase (C1Gal-T2), an Enzyme Synthesizing a Core 1 Structure of O-Glycan. *Journal of Biological Chemistry*, 277(49), 47724–47731. <https://doi.org/10.1074/JBC.M205839200>
- Kudot, T., Iwai, T., Kubota, T., Iwasaki, H., Takayma, Y., Hiruma, T., Inaba, N., Zhang, Y., Gotoh, M., Togayachi, A., & Narimatsu, H. (2006). Molecular cloning and characterization of a novel UDP-Gal:GalNAc α peptide β 1,3-galactosyltransferase (C1Gal-T2), an enzyme synthesizing a core 1 structure of O-glycan. *Journal of Biological Chemistry*, 277(49), 47724–47731. <https://doi.org/10.1074/jbc.M205839200>

- Kufe, D. W. (2009). Mucins in cancer: Function, prognosis and therapy. *Nature Reviews Cancer*, 9(12), 874–885. <https://doi.org/10.1038/nrc2761>
- Kufe, D. W. (2013). MUC1-C oncoprotein as a target in breast cancer: Activation of signaling pathways and therapeutic approaches. *Oncogene*, 32(9), 1073–1081. <https://doi.org/10.1038/onc.2012.158>
- Kuwabara, N., Manya, H., Yamada, T., Tateno, H., Kanagawa, M., Kobayashi, K., Akasaka-Manya, K., Hirose, Y., Mizuno, M., Ikeguchi, M., Toda, T., Hirabayashi, J., Senda, T., Endo, T., & Kato, R. (2016). Carbohydrate-binding domain of the POMGnT1 stem region modulates O-mannosylation sites of α -dystroglycan. *Proceedings of the National Academy of Sciences of the United States of America*, 113(33), 9280–9285. <https://doi.org/10.1073/pnas.1525545113>
- Kwong, P. D., & Liu, Y. (1999). Use of cryoprotectants in combination with immiscible oils for flash cooling macromolecular crystals. *Journal of Applied Crystallography*, 32(1), 102–105. <https://doi.org/10.1107/S0021889898009686>
- Lairson, L. L., Henrissat, B., Davies, G. J., & Withers, S. G. (2008). Glycosyl transferases: Structures, functions, and mechanisms. *Annual Review of Biochemistry*, 77, 521–555. <https://doi.org/10.1146/annurev.biochem.76.061005.092322>
- Landry, J. R., Mager, D. L., & Wilhelm, B. T. (2003). Complex controls: The role of alternative promoters in mammalian genomes. In *Trends in Genetics* (Vol. 19, Issue 11, pp. 640–648). Elsevier. <https://doi.org/10.1016/j.tig.2003.09.014>
- Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, 26(2), 283–291. <https://doi.org/10.1107/s0021889892009944>
- Lee-Sundlov, M. M., Stowell, S. R., & Hoffmeister, K. M. (2020). Multifaceted role of glycosylation in transfusion medicine, platelets, and red blood cells. *Journal of Thrombosis and Haemostasis : JTH*, 18(7), 1535–1547. <https://doi.org/10.1111/JTH.14874>
- Lehoux, S., Mi, R., Aryal, R. P., Wang, Y., Schjoldager, K. T. B. G., Clausen, H., van Die, I., Han, Y., Chapman, A. B., Cummings, R. D., & Ju, T. (2014). Identification of Distinct Glycoforms of IgA1 in Plasma from Patients with Immunoglobulin A (IgA) Nephropathy and Healthy Individuals. *Molecular & Cellular Proteomics : MCP*, 13(11), 3097. <https://doi.org/10.1074/MCP.M114.039693>
- Leppänen, A., Mehta, P., Ouyang, Y. bin, Ju, T., Helin, J., Moore, K. L., van Die, I., Canfield, W. M., McEver, R. P., & Cummings, R. D. (1999). A novel glycosulfopeptide binds to P-selectin and inhibits leukocyte adhesion to P-selectin. *The Journal of Biological Chemistry*, 274(35), 24838–24848. <https://doi.org/10.1074/JBC.274.35.24838>
- Levy, M., & Berger, J. (1988). Worldwide Perspective of IgA Nephropathy. *American Journal of Kidney Diseases*, 12(5), 340–347. [https://doi.org/10.1016/S0272-6386\(88\)80021-0](https://doi.org/10.1016/S0272-6386(88)80021-0)
- Li, G. S., Zhang, H., Lv, J. C., Shen, Y., & Wang, H. Y. (2007). Variants of C1GALT1 gene are associated with the genetic susceptibility to IgA nephropathy. *Kidney International*, 71(5), 448–453. <https://doi.org/10.1038/SJ.KI.5002088>
- 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. (2015). Dynamic interplay between catalytic

and lectin domains of GalNAc-transferases modulates protein O-glycosylation. *Nature Communications*, 6(1), 1–12. <https://doi.org/10.1038/ncomms7937>

- Lira-Navarrete, E., Iglesias-Fernández, J., Zandberg, W. F., Compañón, I., Kong, Y., Corzana, F., Pinto, B. M., Clausen, H., Peregrina, J. M., Vocadlo, D. J., Rovira, C., & Hurtado-Guerrero, R. (2014). Substrate-guided front-face reaction revealed by combined structural snapshots and metadynamics for the polypeptide N-acetylgalactosaminyltransferase 2. *Angewandte Chemie - International Edition*, 53(31), 8206–8210. <https://doi.org/10.1002/anie.201402781>
- Liu, C. H., Hu, R. H., Huang, M. J., Lai, I. R., Chen, C. H., Lai, H. S., Wu, Y. M., & Huang, M. C. (2014). C1GALT1 Promotes Invasive Phenotypes of Hepatocellular Carcinoma Cells by Modulating Integrin β 1 Glycosylation and Activity. *PLOS ONE*, 9(8), e94995. <https://doi.org/10.1371/JOURNAL.PONE.0094995>
- Loureiro, L. R., Sousa, D. P., Ferreira, D., Chai, W., Lima, L., Pereira, C., Lopes, C. B., Correia, V. G., Silva, L. M., Li, C., Santos, L. L., Ferreira, J. A., Barbas, A., Palma, A. S., Novo, C., & Videira, P. A. (2018). Novel monoclonal antibody L2A5 specifically targeting sialyl-Tn and short glycans terminated by alpha-2–6 sialic acids. *Scientific Reports*, 8(1), 1–16. <https://doi.org/10.1038/s41598-018-30421-w>
- Lowe, J. B., & Marth, J. D. (2003). A genetic approach to mammalian glycan function. *Annual Review of Biochemistry*, 72(1), 643–691. <https://doi.org/10.1146/annurev.biochem.72.121801.161809>
- Lübbers, J., Rodríguez, E., & van Kooyk, Y. (2018). Modulation of Immune Tolerance via Siglec-Sialic Acid Interactions. *Frontiers in Immunology*, 9, 2807. <https://doi.org/10.3389/FIMMU.2018.02807/BIBTEX>
- Luckow, V. A., Lee, S. C., Barry, G. F., & Olins, P. O. (1993). Efficient generation of infectious recombinant baculoviruses by site-specific transposon-mediated insertion of foreign genes into a baculovirus genome propagated in Escherichia coli. *Journal of Virology*, 67(8), 4566–4579. <https://doi.org/10.1128/jvi.67.8.4566-4579.1993>
- Luong, J. H. T., Mahmoud, K. A., & Male, K. B. (2011). Instrumentation and Analytical Methods. *Comprehensive Biotechnology, Second Edition*, 2, 829–838. <https://doi.org/10.1016/B978-0-08-088504-9.00139-2>
- Macías-León, J., Bermejo, I. A., Asín, A., García-García, A., Compañón, I., Jiménez-Moreno, E., Coelho, H., Mangini, V., Albuquerque, I. S., Marcelo, F., Asensio, J. L., Bernardes, G. J. L., Joshi, H. J., Fiammengo, R., Blixt, O., Hurtado-Guerrero, R., & Corzana, F. (2020). Structural characterization of an unprecedented lectin-like antitumoral anti-MUC1 antibody. *Chemical Communications*, 56(96), 15137–15140. <https://doi.org/10.1039/d0cc06349e>
- 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. (2014). Serine versus threonine glycosylation with α -o-GalNAc: Unexpected selectivity in their molecular recognition with lectins. *Chemistry - A European Journal*, 20(39), 12616–12627. <https://doi.org/10.1002/chem.201403700>
- Madsen, C. B., Lavrsen, K., Steentoft, C., Vester-Christensen, M. B., Clausen, H., Wandall, H. H., & Pedersen, A. E. (2013). Glycan Elongation Beyond the Mucin Associated Tn Antigen Protects Tumor Cells from Immune-Mediated Killing. *PLoS ONE*, 8(9), e72413.

<https://doi.org/10.1371/journal.pone.0072413>

- Mahajan, V. S., & Pillai, S. (2016). SIALIC ACIDS AND AUTOIMMUNE DISEASE. *Immunological Reviews*, 269(1), 145. <https://doi.org/10.1111/IMR.12344>
- Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., & Simmerling, C. (2015). ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation*, 11(8), 3696–3713. <https://doi.org/10.1021/acs.jctc.5b00255>
- Malycha, F., Eggermann, T., Hristov, M., Schena, F. P., Mertens, P. R., Zerres, K., Floege, J., & Eitner, F. (2009). No evidence for a role of cosmc-chaperone mutations in European IgA nephropathy patients. *Nephrology Dialysis Transplantation*, 24(1), 321–324. <https://doi.org/10.1093/ndt/gfn538>
- Mandel, U., Hassan, H., Therkildsen, M. H., Rygaard, J., Jakobsen, M. H., Juhl, B. R., Dabelsteen, E., & Clausen, H. (1999). Expression of polypeptide GalNAc-transferases in stratified epithelia and squamous cell carcinomas: Immunohistological evaluation using monoclonal antibodies to three members of the GalNAc-transferase family. *Glycobiology*, 9(1), 43–52. <https://doi.org/10.1093/glycob/9.1.43>
- Maresca, M., Lin, V. G., Guo, N., & Yang, Y. (2013). Obligate ligation-gated recombination (ObLiGaRe): Custom-designed nuclease-mediated targeted integration through nonhomologous end joining. *Genome Research*, 23(3), 539–546. <https://doi.org/10.1101/gr.145441.112>
- Marth, J. D., & Grewal, P. K. (2008). Mammalian glycosylation in immunity. *Nature Reviews Immunology* 2008 8:11, 8(11), 874–887. <https://doi.org/10.1038/nri2417>
- 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. (2015). Deciphering the Non-Equivalence of Serine and Threonine O -Glycosylation Points: Implications for Molecular Recognition of the Tn Antigen by an anti-MUC1 Antibody . *Angewandte Chemie*, 127(34), 9968–9972. <https://doi.org/10.1002/ange.201502813>
- Martínez-Sáez, N., Peregrina, J. M., & Corzana, F. (2017). Principles of mucin structure: implications for the rational design of cancer vaccines derived from MUC1-glycopeptides. *Chem. Soc. Rev.*, 46(23), 7154–7175. <https://doi.org/10.1039/C6CS00858E>
- Matsukita, S., Nomoto, M., Kitajima, S., Tanaka, S., Goto, M., Irimura, T., Kim, Y. S., Sato, E., & Yonezawa, S. (2003). Expression of mucins (MUC1, MUC2, MUC5AC and MUC6) in mucinous carcinoma of the breast: comparison with invasive ductal carcinoma. *Histopathology*, 42(1), 26–36. <https://doi.org/10.1046/J.1365-2559.2003.01530.X>
- Mattu, T. S., Pleass, R. J., Willis, A. C., Kilian, M., Wormald, M. R., Lellouch, A. C., Rudd, P. M., Woof, J. M., & Dwek, R. A. (1998). The glycosylation and structure of human serum IgA1, Fab, and Fc regions and the role of N-glycosylation on Fc α receptor interactions. *The Journal of Biological Chemistry*, 273(4), 2260–2272. <https://doi.org/10.1074/JBC.273.4.2260>
- Maverakis, E., Kim, K., Shimoda, M., Gershwin, M. E., Patel, F., Wilken, R., Raychaudhuri, S., Ruhaak, L. R., & Lebrilla, C. B. (2015). Glycans in the immune system and The Altered Glycan Theory of Autoimmunity: a critical review. *Journal of Autoimmunity*, 57, 1–13. <https://doi.org/10.1016/J.JAUT.2014.12.002>

- Mayer, M., & Meyer, B. (1999). Characterization of ligand binding by saturation transfer difference NMR spectroscopy. *Angewandte Chemie - International Edition*, 38(12), 1784–1788. [https://doi.org/10.1002/\(SICI\)1521-3773\(19990614\)38:12<1784::AID-ANIE1784>3.0.CO;2-Q](https://doi.org/10.1002/(SICI)1521-3773(19990614)38:12<1784::AID-ANIE1784>3.0.CO;2-Q)
- Mayer, M., & Meyer, B. (2001). Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor. *Journal of the American Chemical Society*, 123(25), 6108–6117. <https://doi.org/10.1021/ja0100120>
- McEver, R. P., Moore, K. L., & Cummings, R. D. (1995). Leukocyte trafficking mediated by selectin-carbohydrate interactions. *The Journal of Biological Chemistry*, 270(19), 11025–11028. <https://doi.org/10.1074/JBC.270.19.11025>
- Mestecky, J., Raska, M., Julian, B. A., Gharavi, A. G., Renfrow, M. B., Moldoveanu, Z., Novak, L., Matousovic, K., & Novak, J. (2013). IgA Nephropathy: Molecular Mechanisms of the Disease. *Annual Review of Pathology: Mechanisms of Disease*, 8(1), 217–240. <https://doi.org/10.1146/annurev-pathol-011110-130216>
- Mestecky, J., Tomana, M., Moldoveanu, Z., Julian, B. A., Suzuki, H., Matousovic, K., Renfrow, M. B., Novak, L., Wyatt, R. J., & Novak, J. (2008). Role of Aberrant Glycosylation of IgA1 Molecules in the Pathogenesis of IgA Nephropathy. *Kidney & Blood Pressure Research*, 31(1), 29. <https://doi.org/10.1159/000112922>
- Moniaux, N., Escande, F., Porchet, N., Aubert, J. P., & Batra, S. K. (2001). Structural organization and classification of the human mucin genes. *Frontiers in Bioscience : A Journal and Virtual Library*, 6, 1192–1206. <https://doi.org/10.2741/moniaux>
- Moreau, R., Dausset, J., Bernard, J., & Moullec, J. (1957). Acquired hemolytic anemia with polyagglutinability of erythrocytes by a new factor present in normal blood. *Bulletins et Memoires de La Societe Medicale Des Hopitaux de Paris*, 73, 569–587.
- Moremen, K. W., & Haltiwanger, R. S. (2019). Emerging structural insights into glycosyltransferase-mediated synthesis of glycans. In *Nature Chemical Biology* (Vol. 15, Issue 9, pp. 853–864). Nature Publishing Group. <https://doi.org/10.1038/s41589-019-0350-2>
- Mukhopadhyay, P., Lakshmanan, I., Ponnusamy, M. P., Chakraborty, S., Jain, M., Pai, P., Smith, L. M., Lele, S. M., & Batra, S. K. (2013). MUC4 Overexpression Augments Cell Migration and Metastasis through EGFR Family Proteins in Triple Negative Breast Cancer Cells. *PLoS ONE*, 8(2). <https://doi.org/10.1371/journal.pone.0054455>
- Müller, I. (2017). Guidelines for the successful generation of protein-ligand complex crystals. In *Acta Crystallographica Section D: Structural Biology* (Vol. 73, Issue 2, pp. 79–92). International Union of Crystallography. <https://doi.org/10.1107/S2059798316020271>
- Müller, R., Hülsmeier, A. J., Altmann, F., ten Hagen, K., Tiemeyer, M., & Hennet, T. (2005). Characterization of mucin-type core-1 β 1-3 galactosyltransferase homologous enzymes in *Drosophila melanogaster*. *FEBS Journal*, 272(17), 4295–4305. <https://doi.org/10.1111/j.1742-4658.2005.04838.x>
- Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., & Vagin, A. A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D: Biological Crystallography*, 67(4), 355–367. <https://doi.org/10.1107/S0907444911001314>

- Narimatsu, Y., Joshi, H. J., Yang, Z., Gomes, C., Chen, Y. H., Lorenzetti, F. C., Furukawa, S., Schjoldager, K. T., Hansen, L., Clausen, H., Bennett, E. P., & Wandall, H. H. (2018). A validated gRNA library for CRISPR/Cas9 targeting of the human glycosyltransferase genome. *Glycobiology*, *28*(5), 295–305. <https://doi.org/10.1093/glycob/cwx101>
- Nason, R., Büll, C., Konstantinidi, A., Sun, L., Ye, Z., Halim, A., Du, W., Sørensen, D. M., Durbesson, F., Furukawa, S., Mandel, U., Joshi, H. J., Dworkin, L. A., Hansen, L., David, L., Iverson, T. M., Bensing, B. A., Sullam, P. M., Varki, A., ... Narimatsu, Y. (2021). Display of the human mucinome with defined O-glycans by gene engineered cells. *Nature Communications*, *12*(1). <https://doi.org/10.1038/s41467-021-24366-4>
- Neelamegham, S., Aoki-Kinoshita, K., Bolton, E., Frank, M., Lisacek, F., Lütke, T., O'Boyle, N., Packer, N. H., Stanley, P., Toukach, P., Varki, A., & Woods, R. J. (2019). Updates to the Symbol Nomenclature for Glycans guidelines. *Glycobiology*, *29*(9), 620–624. <https://doi.org/10.1093/glycob/cwz045>
- Nguyen, D. C., Joyner, C. J., Sanz, I., & Lee, F. E. H. (2019). Factors Affecting Early Antibody Secreting Cell Maturation Into Long-Lived Plasma Cells. *Frontiers in Immunology*, *10*, 2138. <https://doi.org/10.3389/FIMMU.2019.02138>
- Nihiro, H., & Clark, E. A. (2002). Regulation of B-cell fate by antigen-receptor signals. *Nature Reviews. Immunology*, *2*(12), 945–956. <https://doi.org/10.1038/NRI955>
- Ogata, S., Ho, I., Chen, A., Dubois, D., Maklansky, J., Itzkowitz, S. H., Singhal, A., & Sen-itiroh, H. (1995). Tumor-associated Sialylated Antigens Are Constitutively Expressed in Normal Human Colonic Mucosa. *Cancer Research*, *55*(9), 1869–1874.
- Ohtsubo, K., & Marth, J. D. (2006). Glycosylation in Cellular Mechanisms of Health and Disease. In *Cell* (Vol. 126, Issue 5, pp. 855–867). Elsevier B.V. <https://doi.org/10.1016/j.cell.2006.08.019>
- Olson, F. J., Bäckström, M., Karlsson, H., Burchell, J., & Hansson, G. C. (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*, *15*(2), 177–191. <https://doi.org/10.1093/glycob/cwh158>
- Padler-Karavani, V. (2014). Aiming at the sweet side of cancer: Aberrant glycosylation as possible target for personalized-medicine. In *Cancer Letters* (Vol. 352, Issue 1, pp. 102–112). Elsevier Ireland Ltd. <https://doi.org/10.1016/j.canlet.2013.10.005>
- Paone, P., & Cani, P. D. (2020). Mucus barrier, mucins and gut microbiota: The expected slimy partners? *Gut*, *69*(12), 2232–2243. <https://doi.org/10.1136/gutjnl-2020-322260>
- Parkin, S., & Hope, H. (1998). Macromolecular Cryocrystallography: Cooling, Mounting, Storage and Transportation of Crystals. *Journal of Applied Crystallography*, *31*(6), 945–953. <https://doi.org/10.1107/S0021889898005305>
- Párraga Martínez, J., Delgado Calvo-flores, G., Delgado Calvo-flores, R., & Martín García, J. M. (2010). Práctica de cristalización de proteínas por técnicas de contradifusión en geles , para estudiantes de Farmacia Protein crystallization practice by gels counter-diffusion techniques for Pharmacy students . *ARS Pharmaceutica*, *51*(2), 316–323.
- Paulson, J. C., & Colley, K. J. (1989). Glycosyltransferases. Structure, localization, and control of cell type-specific glycosylation. *Journal of Biological Chemistry*, *264*(30), 17615–17618.

[https://doi.org/10.1016/s0021-9258\(19\)84610-0](https://doi.org/10.1016/s0021-9258(19)84610-0)

- Pedersen, J. W., Bennett, E. P., Schjoldager, K. T. B. G., Meldal, M., Holmér, A. P., Blixt, O., Cló, E., Levery, S. B., Clausen, H., & Wandall, H. H. (2011). Lectin domains of polypeptide GalNAc transferases exhibit glycopeptide binding specificity. *Journal of Biological Chemistry*, *286*(37), 32684–32696. <https://doi.org/10.1074/jbc.M111.273722>
- Pedersen, N. B., Wang, S., Narimatsu, Y., Yang, Z., Halim, A., Schjoldager, K. T. B. G., Madsen, T. D., Seidah, N. G., Bennett, E. P., Levery, S. B., & Clausen, H. (2014). Low density lipoprotein receptor class A repeats are O-glycosylated in linker regions. *Journal of Biological Chemistry*, *289*(25), 17312–17324. <https://doi.org/10.1074/JBC.M113.545053>
- Pei, J., & Grishin, N. v. (2017). Expansion of divergent SEA domains in cell surface proteins and nucleoporin 54. *Protein Science: A Publication of the Protein Society*, *26*(3), 617–630. <https://doi.org/10.1002/pro.3096>
- Pelaseyed, T., Zäch, M., Petersson, Å. C., Svensson, F., Johansson, D. G. A., & Hansson, G. C. (2013). Unfolding dynamics of the mucin SEA domain probed by force spectroscopy suggest that it acts as a cell-protective device. *FEBS Journal*, *280*(6), 1491–1501. <https://doi.org/10.1111/febs.12144>
- Pérez-Vilar, J., & Hill, R. L. (1999). The structure and assembly of secreted mucins. *Journal of Biological Chemistry*, *274*(45), 31751–31754. <https://doi.org/10.1074/jbc.274.45.31751>
- Perrine, C., Ju, T., Cummings, R. D., & Gerken, T. A. (2009). Systematic determination of the peptide acceptor preferences for the human UDP-Gal:glycoprotein- α -GalNAc β 3 galactosyltransferase (T-synthase). *Glycobiology*, *19*(3), 321–328. <https://doi.org/10.1093/glycob/cwn143>
- Piller, V., Piller, F., & Fukuda, M. (1990). Biosynthesis of truncated O-glycans in the T cell line Jurkat. Localization of O-glycan initiation. *Journal of Biological Chemistry*, *265*(16), 9264–9271. [https://doi.org/10.1016/S0021-9258\(19\)38842-8](https://doi.org/10.1016/S0021-9258(19)38842-8)
- Pirulli, D., Crovella, S., Ulivi, S., Zadro, C., Bertok, S., Rendine, S., Scolari, F., Foramitti, M., Ravani, P., Roccatello, D., Savoldi, S., Cerullo, G., Lanzilotta, S. G., Bisceglia, L., Zelante, L., Floege, J., Alexopoulos, E., Kirmizis, D., Ghiggeri, G. M., ... Amoroso, A. (2009). Genetic variant of C1GalT1 contributes to the susceptibility to IgA nephropathy. *Journal of Nephrology*, *22*(1), 152–159.
- Ponnusamy, M. P., Seshacharyulu, P., Vaz, A., Dey, P., & Batra, S. K. (2011). MUC4 stabilizes HER2 expression and maintains the cancer stem cell population in ovarian cancer cells. *Journal of Ovarian Research*, *4*(1). <https://ovarianresearch.biomedcentral.com/articles/10.1186/1757-2215-4-7>
- Pratt, M. R., Hang, H. C., ten Hagen, K. G., Rarick, J., Gerken, T. A., Tabak, L. A., & Bertozzi, C. R. (2004). Deconvoluting the functions of polypeptide N- α -acetylgalactosaminyltransferase family members by glycopeptide substrate profiling. *Chemistry and Biology*, *11*(7), 1009–1016. <https://doi.org/10.1016/J.CHEMBIOL.2004.05.009>
- Promega. (2020). UDP-Glo™ Glycosyltransferase Assay. *User Guide*. www.promega.com
- Qin, W., Zhong, X., Fan, J. M., Zhang, Y. J., Liu, X. R., & Ma, X. Y. (2008). External suppression causes the low expression of the Cosmc gene in IgA nephropathy. *Nephrology, Dialysis, Transplantation: Official Publication of the European Dialysis and Transplant Association -*

European Renal Association, 23(5), 1608–1614. <https://doi.org/10.1093/NDT/GFM781>

- Qin, W., Zhou, Q., Yang, L. C., Li, Z., Su, B. H., Luo, H., & Fan, J. M. (2005). Peripheral B lymphocyte beta1,3-galactosyltransferase and chaperone expression in immunoglobulin A nephropathy. *Journal of Internal Medicine*, 258(5), 467–477. <https://doi.org/10.1111/J.1365-2796.2005.01558.X>
- Rachagani, S., Torres, M. P., Moniaux, N., & Batra, S. K. (2009). Current status of mucins in the diagnosis and therapy of cancer. *BioFactors*, 35(6), 509–527. <https://doi.org/10.1002/biof.64>
- Radhakrishnan, P., Dabelsteen, S., Madsen, F. B., Francavilla, C., Kopp, K. L., Steentoft, C., Vakhrushev, S. Y., Olsen, J. v., Hansen, L., Bennett, E. P., Woetmann, A., Yin, G., Chen, L., Song, H., Bak, M., Hlady, R. A., Peters, S. L., Opavsky, R., Thode, C., ... Wandall, H. H. (2014). Immature truncated O-glycophenotype of cancer directly induces oncogenic features. *Proceedings of the National Academy of Sciences of the United States of America*, 111(39), E4066–E4075. <https://doi.org/10.1073/pnas.1406619111>
- Rajewsky, K. (1996). Clonal selection and learning in the antibody system. *Nature*, 381(6585), 751–758. <https://doi.org/10.1038/381751A0>
- Raman, J., Guan, Y., Perrine, C. L., Gerken, T. A., & Tabak, L. A. (2012). UDP-N-acetyl α -d-galactosamine: Polypeptide N- acetylgalactosaminyltransferases: Completion of the family tree. *Glycobiology*, 22(6), 768–777. <https://doi.org/10.1093/glycob/cwr183>
- Ramazi, S., & Zahiri, J. (2021). Post-translational modifications in proteins: Resources, tools and prediction methods. *Database*, 2021(7), 1–20. <https://doi.org/10.1093/database/baab012>
- Rao, S. T., & Rossmann, M. G. (1973). Comparison of super-secondary structures in proteins. *Journal of Molecular Biology*, 76(2), 241–256. [https://doi.org/10.1016/0022-2836\(73\)90388-4](https://doi.org/10.1016/0022-2836(73)90388-4)
- Rawlings, D. J., Metzler, G., Wray-Dutra, M., & Jackson, S. W. (2017). Altered B cell signalling in autoimmunity. *Nature Reviews. Immunology*, 17(7), 421–436. <https://doi.org/10.1038/NRI.2017.24>
- Reily, C., Stewart, T. J., Renfrow, M. B., & Novak, J. (2019). Glycosylation in health and disease. *Nature Reviews Nephrology* 2019 15:6, 15(6), 346–366. <https://doi.org/10.1038/s41581-019-0129-4>
- Revoredo, L., Wang, S., Bennett, E. P., Clausen, H., Moremen, K. W., Jarvis, D. L., ten Hagen, K. G., Tabak, L. A., & Gerken, T. A. (2016). Mucin-type o-glycosylation is controlled by short- and long-range glycopeptide substrate recognition that varies among members of the polypeptide GalNAc transferase family. *Glycobiology*, 26(4), 360–376. <https://doi.org/10.1093/glycob/cwv108>
- Rømer, T. B., Aasted, M. K. M., Dabelsteen, S., Groen, A., Schnabel, J., Tan, E., Pedersen, J. W., Haue, A. D., & Wandall, H. H. (2021). Mapping of truncated O-glycans in cancers of epithelial and non-epithelial origin. *British Journal of Cancer*, 125(9), 1239–1250. <https://doi.org/10.1038/s41416-021-01530-7>
- Röttger, S., White, J., Wandall, H. H., Olivo, J. C., Stark, A., Bennett, E. P., Whitehouse, C., Berger, E. G., Clausen, H., & Nilsson, T. (1998). Localization of three human polypeptide GalNAc-transferases in HeLa cells suggests initiation of O-linked glycosylation throughout the Golgi apparatus. *Journal of Cell Science*, 111(1), 45–60. <https://doi.org/10.1242/jcs.111.1.45>

- Schachter, H., McGuire, E. J., & Roseman, S. (1971). Sialic acids. 13. A uridine diphosphate D-galactose: mucin galactosyltransferase from porcine submaxillary gland. *The Journal of Biological Chemistry*, 246(17), 5321–5328.
- Schjoldager, K. T. B. G., & Clausen, H. (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*, 1820(12), 2079–2094. <https://doi.org/10.1016/j.bbagen.2012.09.014>
- Schjoldager, K. T. B. G., Joshi, H. J., Kong, Y., Goth, C. K., King, S. L., Wandall, H. H., Bennett, E. P., Vakhrushev, S. Y., & Clausen, H. (2015). Deconstruction of O-glycosylation—Gal NA c-T isoforms direct distinct subsets of the O-glycoproteome. *EMBO Reports*, 16(12), 1713–1722. <https://doi.org/10.15252/embr.201540796>
- Schjoldager, K. T. B. G., Vakhrushev, S. Y., Kong, Y., Steentoft, C., Nudelman, A. S., Pedersen, N. B., Wandall, H. H., Mandel, U., Bennett, E. P., Levery, S. B., & Clausen, H. (2012). Probing isoform-specific functions of polypeptide GalNAc-transferases using zinc finger nuclease glycoengineered SimpleCells. *Proceedings of the National Academy of Sciences of the United States of America*, 109(25), 9893–9898. <https://doi.org/10.1073/pnas.1203563109>
- Scholz, J., & Suppmann, S. (2017). A new single-step protocol for rapid baculovirus-driven protein production in insect cells. *BMC Biotechnology*, 17(1). <https://doi.org/10.1186/s12896-017-0400-3>
- Schwientek, T., Bennett, E. P., Flores, C., Thacker, J., Hollmann, M., Reis, C. A., Behrens, J., Mandel, U., Keck, B., Mireille, A., Schäfer, Haselmann, K., Zubarev, R., Roepstorff, P., Burchell, J. M., Taylor-Papadimitriou, J., Hollingsworth, M. A., & Clausena, H. (2002). Functional conservation of subfamilies of putative UDP-N-acetylgalactosamine:Polypeptide N-acetylgalactosaminyltransferases in *Drosophila*, *Caenorhabditis elegans*, and mammals. One subfamily composed of I(2)35Aa is essential in *Drosophila*. *Journal of Biological Chemistry*, 277(25), 22623–22638. <https://doi.org/10.1074/jbc.M202684200>
- Schwientek, T., Nomoto, M., Levery, S. B., Merckx, G., van Kessel, A. G., Bennett, E. P., Hollingsworth, M. A., & Clausen, H. (1999). Control of O-glycan branch formation: Molecular cloning of human cDNA encoding a novel β 1,6-N-acetylglucosaminyltransferase forming core 2 and core 4. *Journal of Biological Chemistry*, 274(8), 4504–4512. <https://doi.org/10.1074/jbc.274.8.4504>
- Senapati, S., Sharma, P., Bafna, S., Roy, H. K., & Batra, S. K. (2008). The MUC gene family: Their role in the diagnosis and prognosis of gastric cancer. *Histology and Histopathology*, 23(12), 1541–1552.
- Shendure, J., Balasubramanian, S., Church, G. M., Gilbert, W., Rogers, J., Schloss, J. A., & Waterston, R. H. (2017). DNA sequencing at 40: Past, present and future. *Nature*, 550(7676). <https://doi.org/10.1038/nature24286>
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., & Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, 7(1), 539. <https://doi.org/10.1038/MSB.2011.75>
- Sigma-Aldrich. (2014). Enzymatic Deglycosylation. *User Guide*, 1–2.
- Sinnott, M. L. (1990). Catalytic Mechanisms of Enzymic Glycosyl Transfer. *Chemical Reviews*,

90(7), 1171–1202. <https://doi.org/10.1021/cr00105a006>

- Skoog, L., & Tani, E. (2011). Immunocytochemistry: An indispensable technique in routine cytology. *Cytopathology*, 22(4), 215–229. <https://doi.org/10.1111/j.1365-2303.2011.00887.x>
- Sletmoen, M., Gerken, T. A., Stokke, B. T., Burchell, J., & Brewer, C. F. (2018). Tn and STn are members of a family of carbohydrate tumor antigens that possess carbohydrate–carbohydrate interactions. *Glycobiology*, 28(7), 437–442. <https://doi.org/10.1093/GLYCOB/CWY032>
- Stanley, P. (2011). Golgi glycosylation. *Cold Spring Harbor Perspectives in Biology*, 3(4), 1–13. <https://doi.org/10.1101/cshperspect.a005199>
- States, D. J., & Gish, W. (1994). Combined use of sequence similarity and codon bias for coding region identification. *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology*, 1(1), 39–50. <https://doi.org/10.1089/CMB.1994.1.39>
- Steenftoft, C., Vakhrushev, S. Y., Joshi, H. J., Kong, Y., Vester-Christensen, M. B., Schjoldager, K. T. B. G., Lavrsen, K., Dabelsteen, S., Pedersen, N. B., Marcos-Silva, L., Gupta, R., Paul Bennett, E., Mandel, U., Brunak, S., Wandall, H. H., Lavery, S. B., & Clausen, H. (2013). Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *EMBO Journal*, 32(10), 1478–1488. <https://doi.org/10.1038/emboj.2013.79>
- Steenftoft, C., Vakhrushev, S. Y., Vester-Christensen, M. B., Schjoldager, K. T. B. G., Kong, Y., Bennett, E. P., Mandel, U., Wandall, H., Lavery, S. B., & Clausen, H. (2011). Mining the O-glycoproteome using zinc-finger nuclease-glycoengineered SimpleCell lines. *Nature Methods*, 8(11), 977–982. <https://doi.org/10.1038/nmeth.1731>
- Steenftoft, C., Yang, Z., Wang, S., Ju, T., Vester-Christensen, M. B., Festari, M. F., King, S. L., Moremen, K., Larsen, I. S. B., Goth, C. K., Schjoldager, K. T., Hansen, L., Bennett, E. P., Mandel, U., & Narimatsu, Y. (2019). A validated collection of mouse monoclonal antibodies to human glycosyltransferases functioning in mucin-type O-glycosylation. *Glycobiology*, 29(9), 645–656. <https://doi.org/10.1093/glycob/cwz041>
- Stojanovic, D., Vischer, P., & Hughes, R. C. (1984). Glycosyl transferases of baby hamster kidney cells and ricin-resistant mutants: O-Glycan biosynthesis. *European Journal of Biochemistry*, 138(3), 551–562. <https://doi.org/10.1111/j.1432-1033.1984.tb07950.x>
- Sun, Q., Ju, T., & Cummings, R. D. (2011). The transmembrane domain of the molecular chaperone cosmc directs its localization to the endoplasmic reticulum. *Journal of Biological Chemistry*, 286(13), 11529–11542. <https://doi.org/10.1074/JBC.M110.173591>
- Sun, Q., Zhang, J., Zhou, N., Liu, X., & Shen, Y. (2015). DNA Methylation in Cosmc Promoter Region and Aberrantly Glycosylated IgA1 Associated with Pediatric IgA Nephropathy. *PLOS ONE*, 10(2), e0112305. <https://doi.org/10.1371/JOURNAL.PONE.0112305>
- Suzuki, H., Moldoveanu, Z., Hall, S., Brown, R., Vu, H. L., Novak, L., Julian, B. A., Tomana, M., Wyatt, R. J., Edberg, J. C., Alarcón, G. S., Kimberly, R. P., Tomino, Y., Mestecky, J., & Novak, J. (2008). IgA1-secreting cell lines from patients with IgA nephropathy produce aberrantly glycosylated IgA1. *The Journal of Clinical Investigation*, 118(2), 629–639. <https://doi.org/10.1172/JCI33189>
- Swaigood, H. E. (2005). The importance of disulfide bridging. *Biotechnology Advances*, 23(1), 71–73. <https://doi.org/10.1016/j.biotechadv.2004.09.004>

- Swinehart, D. F. (1962). The Beer-Lambert law. *Journal of Chemical Education*, 39(7), 333–335. <https://doi.org/10.1021/ed039p333>
- Taujale, R., Zhou, Z., Yeung, W., Moremen, K. W., Li, S., & Kannan, N. (2021). Mapping the glycosyltransferase fold landscape using interpretable deep learning. *Nature Communications*, 12(1). <https://doi.org/10.1038/s41467-021-25975-9>
- Taylor-Papadimitriou, J., Burchell, J., Miles, D. W., & Dalziel, M. (1999). MUC1 and cancer. In *Biochimica et Biophysica Acta - Molecular Basis of Disease* (Vol. 1455, Issues 2–3, pp. 301–313). Biochim Biophys Acta. [https://doi.org/10.1016/S0925-4439\(99\)00055-1](https://doi.org/10.1016/S0925-4439(99)00055-1)
- Tellinghuisen, J. (2008). Isothermal titration calorimetry at very low c. *Analytical Biochemistry*, 373(2), 395–397. <https://doi.org/10.1016/j.ab.2007.08.039>
- ten Hagen, K. G., Fritz, T. A., & Tabak, L. A. (2003a). All in the family: The UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferases. *Glycobiology*, 13(1), 1–16. <https://doi.org/10.1093/glycob/cwg007>
- ten Hagen, K. G., Fritz, T. A., & Tabak, L. A. (2003b). All in the family: the UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferases. *Glycobiology*, 13(1), 1R-16R. <https://doi.org/10.1093/GLYCOB/CWG007>
- Tenno, M., Ohtsubo, K., Hagen, F. K., Ditto, D., Zarbock, A., Schaerli, P., Andrian, U. H. von, Ley, K., Le, D., Tabak, L. A., & Marth, J. D. (2007). Initiation of Protein O Glycosylation by the Polypeptide GalNAcT-1 in Vascular Biology and Humoral Immunity. *Molecular and Cellular Biology*, 27(24), 8783. <https://doi.org/10.1128/MCB.01204-07>
- Tessier, D. C., Thomas, D. Y., Khouri, H. E., Laliberié, F., & Vernet, T. (1991). Enhanced secretion from insect cells of a foreign protein fused to the honeybee melittin signal peptide. *Gene*, 98(2), 177–183. [https://doi.org/10.1016/0378-1119\(91\)90171-7](https://doi.org/10.1016/0378-1119(91)90171-7)
- Thériault, C., Pinard, M., Comamala, M., Migneault, M., Beaudin, J., Matte, I., Boivin, M., Piché, A., & Rancourt, C. (2011). MUC16 (CA125) regulates epithelial ovarian cancer cell growth, tumorigenesis and metastasis. *Gynecologic Oncology*, 121(3), 434–443. <https://doi.org/10.1016/J.YGYNO.2011.02.020>
- Thurnher, M., Berger, E. G., Clausen, H., Fierz, W., & Lanzavecchia, A. (1992). T cell clones with normal or defective O-galactosylation from a patient with permanent mixed-field polyagglutinability. *European Journal of Immunology*, 22(7), 1835–1842. <https://doi.org/10.1002/EJL1830220724>
- Turnbull, W. B., & Daranas, A. H. (2003). On the Value of c: Can Low Affinity Systems Be Studied by Isothermal Titration Calorimetry? *Journal of the American Chemical Society*, 125(48), 14859–14866. <https://doi.org/10.1021/ja036166s>
- Valero-González, J., Leonhard-Melief, C., Lira-Navarrete, E., Jiménez-Osés, G., Hernández-Ruiz, C., Pallarés, M. C., Yruela, I., Vasudevan, D., Lostao, A., Corzana, F., Takeuchi, H., Haltiwanger, R. S., & Hurtado-Guerrero, R. (2016). A proactive role of water molecules in acceptor recognition by protein O-fucosyltransferase 2. *Nature Chemical Biology*, 12(4), 240–246. <https://doi.org/10.1038/nchembio.2019>
- van der Laan, J. M., Swarte, M. B. A., Groendijk, H., Hol, W. G. J., & Drenth, J. (1989). The influence of purification and protein heterogeneity on the crystallization of p-hydroxybenzoate hydroxylase. *European Journal of Biochemistry*, 179(3), 715–724.

<https://doi.org/10.1111/j.1432-1033.1989.tb14605.x>

- van Putten, J. P. M., & Strijbis, K. (2017). Transmembrane Mucins: Signaling Receptors at the Intersection of Inflammation and Cancer. *Journal of Innate Immunity*, 9(3), 281–299. <https://doi.org/10.1159/000453594>
- Varki, A., Cummings, R. D., Aebi, M., Packer, N. H., Seeberger, P. H., Esko, J. D., Stanley, P., Hart, G., Darvill, A., Kinoshita, T., Prestegard, J. J., Schnaar, R. L., Freeze, H. H., Marth, J. D., Bertozzi, C. R., Etzler, M. E., Frank, M., Vliegenthart, J. F. G., Lütke, T., ... Kornfeld, S. (2015). Symbol nomenclature for graphical representations of glycans. *Glycobiology*, 25(12), 1323–1324. <https://doi.org/10.1093/glycob/cwv091>
- Varki, A., Cummings, R. D., Esko, J. D., Freeze, H. H., Hart, G. W., & Marth, J. D. (2022). Essentials of Glycobiology. In *Essentials of Glycobiology (4th Ed.)*. Cold Spring Harbor Laboratory Press. <https://doi.org/10.1101/9781621824213>
- Wagner, G. K., & Pesnot, T. (2010). Glycosyltransferases and their assays. *ChemBioChem*, 11(14), 1939–1949. <https://doi.org/10.1002/cbic.201000201>
- Walsh, C. T., Garneau-Tsodikova, S., & Gatto, G. J. (2005). Protein posttranslational modifications: The chemistry of proteome diversifications. *Angewandte Chemie - International Edition*, 44(45), 7342–7372. <https://doi.org/10.1002/anie.200501023>
- Wandall, H. H., Irazoqui, F., Tarp, M. A., Bennett, E. P., Mandel, U., Takeuchi, H., Kato, K., Irimura, T., Suryanarayanan, G., Hollingsworth, M. A., & Clausen, H. (2007). The lectin domains of polypeptide GalNAc-transferases exhibit carbohydrate-binding specificity for GalNAc: Lectin binding to GalNAc-glycopeptide substrates is required for high density GalNAc-O-glycosylation. *Glycobiology*, 17(4), 374–387. <https://doi.org/10.1093/glycob/cwl082>
- Wandall, H. H., Nielsen, M. A. I., King-Smith, S., de Haan, N., & Bagdonaite, I. (2021). Global functions of O-glycosylation: promises and challenges in O-glycobiology. In *FEBS Journal* (Vol. 288, Issue 24, pp. 7183–7212). John Wiley and Sons Inc. <https://doi.org/10.1111/febs.16148>
- Wang, R. Q., & Fang, D. C. (2003). Alterations of MUC1 and MUC3 expression in gastric carcinoma: Relevance to patient clinicopathological features. *Journal of Clinical Pathology*, 56(5), 378–384. <https://doi.org/10.1136/jcp.56.5.378>
- Wang, S., Mao, Y., Narimatsu, Y., Ye, Z., Tian, W., Goth, C. K., Lira-Navarrete, E., Pedersen, N. B., Benito-Vicente, A., Martin, C., Uribe, K. B., Hurtado-Guerrero, R., Christoffersen, C., Seidah, N. G., Nielsen, R., Christensen, E. I., Hansen, L., Bennett, E. P., Vakhrushev, S. Y., ... Clausen, H. (2018). Site-specific O-glycosylation of members of the low-density lipoprotein receptor superfamily enhances ligand interactions. *Journal of Biological Chemistry*, 293(19), 7408–7422. <https://doi.org/10.1074/JBC.M117.817981>
- Wang, Y., Jobe, S. M., Ding, X., Choo, H., Archer, D. R., Mi, R., Ju, T., & Cummings, R. D. (2012). Platelet biogenesis and functions require correct protein O-glycosylation. *Proceedings of the National Academy of Sciences of the United States of America*, 109(40), 16143–16148. <https://doi.org/10.1073/pnas.1208253109>
- Wang, Y., Ju, T., Ding, X., Xia, B., Wang, W., Xia, L., He, M., & Cummings, R. D. (2010). Cosmc is an essential chaperone for correct protein O-glycosylation. *Proceedings of the National Academy of Sciences of the United States of America*, 107(20), 9228–9233.

<https://doi.org/10.1073/pnas.0914004107>

- Wei, X., Xu, H., & Kufe, D. (2005). Human MUC1 oncoprotein regulates p53-responsive gene transcription in the genotoxic stress response. *Cancer Cell*, 7(2), 167–178. <https://doi.org/10.1016/j.ccr.2005.01.008>
- Wilde, M., Klausberger, M., Palmberger, D., Ernst, W., & Grabherr, R. (2014). Tnao38, high five and Sf9-evaluation of host-virus interactions in three different insect cell lines: Baculovirus production and recombinant protein expression. *Biotechnology Letters*, 36(4), 743–749. <https://doi.org/10.1007/s10529-013-1429-6>
- Wildt, S., & Gerngross, T. U. (2005). The humanization of N-glycosylation pathways in yeast. *Nature Reviews Microbiology*, 3(2), 119–128. <https://doi.org/10.1038/nrmicro1087>
- Wilkinson, H., & Saldova, R. (2020). Current methods for the characterization of o-glycans. *Journal of Proteome Research*, 19(10), 3890–3905. <https://doi.org/10.1021/acs.jproteome.0c00435>
- Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., & Wilson, K. S. (2011). Overview of the CCP4 suite and current developments. In *Acta Crystallographica Section D: Biological Crystallography* (Vol. 67, Issue 4, pp. 235–242). <https://doi.org/10.1107/S0907444910045749>
- Wyman, C., & Kanaar, R. (2006). DNA double-strand break repair: All's well that ends well. *Annual Review of Genetics*, 40, 363–383. <https://doi.org/10.1146/annurev.genet.40.110405.090451>
- Xia, L., Ju, T., Westmuckett, A., An, G., Ivanciu, L., McDaniel, J. M., Lupu, F., Cummings, R. D., & McEver, R. P. (2004). Defective angiogenesis and fatal embryonic hemorrhage in mice lacking core 1-derived O-glycans. *Journal of Cell Biology*, 164(3), 451–459. <https://doi.org/10.1083/jcb.200311112>
- Xia, L., & McEver, R. P. (2006). Targeted disruption of the gene encoding core 1 beta1-3-galactosyltransferase (T-synthase) causes embryonic lethality and defective angiogenesis in mice. *Methods in Enzymology*, 416, 314–331. [https://doi.org/10.1016/S0076-6879\(06\)16021-8](https://doi.org/10.1016/S0076-6879(06)16021-8)
- Xie, L. S., Qin, W., Fan, J. M., Huang, J., & Xie, X. S. (2010). The role of C1GALT1C1 in lipopolysaccharide-induced IgA1 aberrant O-glycosylation in IgA nephropathy. *Clinical and Investigative Medicine. Medecine Clinique et Experimentale*, 33(1). <https://doi.org/10.25011/CIM.V33I1.11832>
- Yamada, K., Kobayashi, N., Ikeda, T., Suzuki, Y., Tsuge, T., Horikoshi, S., Emancipator, S. N., & Tomino, Y. (2010). Down-regulation of core 1 beta1,3-galactosyltransferase and Cosmc by Th2 cytokine alters O-glycosylation of IgA1. *Nephrology, Dialysis, Transplantation: Official Publication of the European Dialysis and Transplant Association - European Renal Association*, 25(12), 3890–3897. <https://doi.org/10.1093/NDT/GFQ325>
- Yang, Z., Halim, A., Narimatsu, Y., Joshi, H. J., Steentoft, C., Schjoldager, K. T. B. G., Schulz, M. A., Sealover, N. R., Kayser, K. J., Bennett, E. P., Lavery, S. B., Vakhrushev, S. Y., & Clausen, H. (2014). The GalNAc-type O-glycoproteome of CHO cells characterized by the simplecell strategy. *Molecular and Cellular Proteomics*, 13(12), 3224–3235. <https://doi.org/10.1074/mcp.M114.041541>

- Yang, Z., Steentoft, C., Hauge, C., Hansen, L., Thomsen, A. L., Niola, F., Vester-Christensen, M. B., Frödin, M., Clausen, H., Wandall, H. H., & Bennett, E. P. (2015). Fast and sensitive detection of indels induced by precise gene targeting. *Nucleic Acids Research*, *43*(9). <https://doi.org/10.1093/nar/gkv126>
- Yang, Z., Wang, S., Halim, A., Schulz, M. A., Frodin, M., Rahman, S. H., Vester-Christensen, M. B., Behrens, C., Kristensen, C., Vakhrushev, S. Y., Bennett, E. P., Wandall, H. H., & Clausen, H. (2015). Engineered CHO cells for production of diverse, homogeneous glycoproteins. *Nature Biotechnology*, *33*(8), 842–844. <https://doi.org/10.1038/nbt.3280>
- Yeh, J. C., Hiraoka, N., Petryniak, B., Nakayama, J., Ellies, L. G., Rabuka, D., Hindsgaul, O., Marth, J. D., Lowe, J. B., & Fukuda, M. (2001). Novel sulfated lymphocyte homing receptors and their control by a Core1 extension beta 1,3-N-acetylglucosaminyltransferase. *Cell*, *105*(7), 957–969. [https://doi.org/10.1016/S0092-8674\(01\)00394-4](https://doi.org/10.1016/S0092-8674(01)00394-4)
- Yoshida, H., Fuwa, T. J., Arima, M., Hamamoto, H., Sasaki, N., Ichimiya, T., Osawa, K. I., Ueda, R., & Nishihara, S. (2008). Identification of the *Drosophila* core 1 β 1,3-galactosyltransferase gene that synthesizes T antigen in the embryonic central nervous system and hemocytes. *Glycobiology*, *18*(12), 1094–1104. <https://doi.org/10.1093/glycob/cwn094>
- Zeng, J., Aryal, R. P., Stavenhagen, K., Luo, C., Liu, R., Wang, X., Chen, J., Li, H., Matsumoto, Y., Wang, Y., Wang, J., Ju, T., & Cummings, R. D. (2021). Cosmc deficiency causes spontaneous autoimmunity by breaking B cell tolerance. *Science Advances*, *7*(41), 1–14. <https://doi.org/10.1126/sciadv.abg9118>
- Zeng, J., Eljalby, M., Aryal, R. P., Lehoux, S., Stavenhagen, K., Kudelka, M. R., Wang, Y., Wang, J., Ju, T., von Andrian, U. H., & Cummings, R. D. (2020). Cosmc controls B cell homing. *Nature Communications*, *11*(1). <https://doi.org/10.1038/s41467-020-17765-6>

8. APÉNDICES

8.1. Lista de tablas y figuras.

Tabla 1.1. MPTs más importantes y sus principales funciones.	6
Tabla 2.1. Núcleos o <i>cores</i> de los glicanos de tipo GalNAc y epítomos antigénicos de las mucinas.	28
Tabla 5.1. Secuencia aminoacídica de la serie de glicopéptidos diseñados racionalmente.	104
Tabla 5.2. Parámetros cinéticos de los (glico)péptidos usados en esta tesis doctoral tanto con <i>DmC1GalT1</i> ^{T43-Q388} como con <i>DmC1GalT1</i> ^{S73-Q388}	108
Tabla 5.3. Tabla de asignación de los picos del espectro RMN-H ¹	115 – 116
Tabla 5.4. Parámetros termodinámicos de la unión de los ligandos a <i>DmC1GalT1</i> ^{T43-Q388}	119
Tabla 5.5. Recolección de datos de la difracción cristalográfica y datos estadísticos del refinamiento de la estructura de <i>DmC1GalT1</i> ^{S73-Q388}	126 – 127
Tabla 5.6. Interacciones observadas durante la simulación de DM del complejo formado por APDT*RP y <i>DmC1GalT1</i> en presencia de UDP-Gal	152 – 153
Tabla 5.7. Interacciones observadas durante la simulación de MD del complejo formado por P4 y <i>DmC1GalT1</i> en presencia de UDP-Gal	154 – 155
Tabla 8.1. Código genético	209
Figura 1.1. Vías de O-glicosilación en los diferentes reinos	9
Figura 2.1 Mucinas de transmembrana y secretadas.....	14
Figura 2.2. Familia de mucinas de transmembrana	16
Figura 2.3. Proteínas transmembrana del tipo II.....	19

Figura 2.4. Representación gráfica de la asistencia de los dominios catalíticos y de lectina de las GalNAc-Ts sobre péptidos o glicopéptidos	22
Figura 2.5. Estructuras y vías de síntesis de los O-glicanos	26
Figura 2.6. Modelo del funcionamiento de Cosmc como chaperona molecular en el retículo endoplasmático para asistir al plegamiento de C1GalT1.....	33
Figura 2.7. Biosíntesis de los precursores de las cadenas de O-glicanos.....	37
Figura 2.8. Glicosilaciones aberrantes de las mucinas en tumores y condiciones cancerosas	38
Figura 2.9. Las dos glicofomas de la inmunoglobulina A1 presentes en el plasma humano.....	43
Figura 2.10. Antígenos Tn y STn como TACAs en células tumorales.....	45
Figura 2.11. Funciones de las mucinas transmembrana en las células de colon....	46
Figura 2.12. Papel de los glicanos de tipo O-GalNAc en la invasión de células cancerígenas	48
Figura 4.1. Estados larvarios y adultos de <i>Spodoptera frugiperda</i> (a) y <i>Trichoplusia ni</i> (b).....	54
Figura 4.2. Mapa del vector pHLSec	55
Figura 4.3. Mapa del vector EPB71	56
Figura 4.4. Mapa del vector pFastBac TM 1	57
Figura 4.5. Sistema de edición genómica CRISPR/Cas9.....	59
Figura 4.6. Esquema simplificado del sistema de expresión de baculovirus.....	64
Figura 4.7. Protocolo de aislamiento de ADN plasmídico	66

Figura 4.8. Esquema de la correcta colocación de todos los componentes para la electrotransferencia mediante el Trans-Blot Turbo	72
Figura 4.9. Sitio de corte y requerimientos estructurales para que se produzca el corte de los glicanos por la endoglicosidasa-H.....	75
Figura 4.10. Gel SDS-PAGE que muestra la pureza de <i>DmC1GalT1</i> ^{S73-Q388} y el cambio de peso molecular en el gel tras el tratamiento de la proteína (overnight) con Endo-H	76
Figura 4.11. Análisis de <i>DmC1GalT1</i> ^{S73-Q388} mediante cromatografía de exclusión molecular en la columna HiLoad 26/60 Superdex.....	77
Figura 4.12. Esquema del proceso de resolución estructural de proteínas por cristalización	82
Figura 4.13. Ejemplo de cristales de proteína.....	84
Figura 4.14. Esquema de un experimento de STD-RMN	88
Figura 4.15. Esquema general del funcionamiento de la calorimetría de titulación isoterma.....	92
Figura 4.16. Principios del funcionamiento del kit UDP-Glo para las cinéticas de C1GalT1	94
Figura 5.1. Alineamiento múltiple de las secuencias de C1GalT1	99 – 100
Figura 5.2. Alineamiento múltiple de las secuencias de C1GalT1 y predicción de la estructura secundaria.....	101
Figura 5.3. Pruebas de expresión de construcciones de <i>DmC1GalT1</i>	102
Figura 5.4. Cinéticas de <i>DmC1GalT1</i> ^{T43-Q388} con concentraciones variables de UDP-Gal.....	105

Figura 5.5. Cinéticas de <i>DmC1GalT1</i> ^{T43-Q388} con concentraciones variables de APDT*RP y APDS*RP	106
Figura 5.6. Ensayos cinéticos de <i>DmC1GalT1</i> ^{T43-Q388} con varios (glico)péptidos y α -O-metil-GalNAc	107
Figura 5.7. Cinéticas de <i>DmC1GalT1</i> ^{T43-Q388} con concentraciones variables de α -O-metil-GalNAc.....	109
Figura 5.8. Espectro STD-RMN de APDTRP	111
Figura 5.9. Espectro STD-RMN de α -O-methyl-GalNAc (a) y de APDT*RP (b)	113 – 114
Figura 5.10. Espectros STD-RMN y mapas de epítipo de los glicopéptidos APDT*RP y P4 en presencia y ausencia de UDP.....	117 – 118
Figura 5.11. Datos de ITC de la unión entre UDP, APDTRP y P7 en presencia y ausencia de exceso de UDP a <i>DmC1GalT1</i> ^{T43-Q388}	120
Figura 5.12. Datos de ITC de la unión entre APDT*RP y P4 a <i>DmC1GalT1</i> ^{T43-Q388}	121
Figura 5.13. Parámetros termodinámicos de la unión a <i>DmC1GalT1</i> ^{T43-Q388} de los glicopéptidos en presencia o ausencia de UDP.....	123
Figura 5.14. Cinéticas de <i>DmC1GalT1</i> ^{T43-Q388} y <i>DmC1GalT1</i> ^{S73-Q388} con concentraciones variables de APDT*RP	124
Figura 5.15. Estructura cristalina de la forma dimérica del complejo <i>DmC1GalT1</i> ^{S73-Q388} -UDP-Mn ²⁺ -APDT*RP en formato <i>ribbon</i>	125 – 126
Figura 5.16. Vista en primer plano del sitio activo de <i>DmC1GalT1</i> ^{S73-Q388} mostrando Mn ²⁺ , UDP y APDT*RP unidos a la proteína.....	128

Figura 5.17. Representación superficial de homología de C1GalT1 y de la superficie electrostática de <i>DmC1GalT1</i> ^{S73-Q388}	129
Figura 5.18. Homólogos estructurales de <i>DmC1GalT1</i> , los ligandos y los cofactores	130
Figura 5.19. Superposición de los ligandos de <i>DmC1GalT1</i> , B3GNT2 y Mfng....	131
Figura 5.20. Vista estereoscópica del sitio activo de <i>DmC1GalT1</i> ^{S73-Q388} en complejo ternario con UDP-Mn ²⁺ y APDT*RP	132
Figura 5.21. Geometría del enlace glicosídico del glicopéptido APDT*RP en solución obtenida mediante simulaciones de dinámica molecular de 0,5 μ s (conformación eclipsada).....	134
Figura 5.22. Vista en primer plano del complejo <i>DmC1GalT1</i> -APDT*RP en presencia de UDP-Gal-Mn ⁺²	135
Figura 5.23. Mapas de energía libre (ϕ , ψ) de los ángulos diédricos glicosídicos.	136
Figura 5.24. Simulación y vista detallada del sitio activo de <i>DmC1GalT1</i> con los sustratos de la reacción	138
Figura 5.25. Actividades relativas de los mutantes de <i>DmC1GalT1</i> (a) y mecanismo de reacción propuesto (b).....	139
Figura 5.26. La vía de <i>O</i> -glicosilación se indica con el nombre de las enzimas involucradas en la síntesis de las estructuras de los <i>O</i> -glicanos	140
Figura 5.27. PCR de la inserción de los constructos y análisis inmunocitológicos con el mAb 3C9 específico para el antígeno T en los clones <i>knockins</i>	141 – 142
Figura 5.28. Análisis por citometría de flujo de la reinsertión del antígeno T con	

los mutantes de <i>HsC1GalT1</i>	143
Figura 5.29. Fluctuaciones atómicas ($C\alpha$) obtenidas de las simulaciones de MD durante 0,5 μ s de <i>DmC1GalT1</i> y dos complejos binarios	144
Figura 5.30. Relación entre los ángulos ϕ/ψ del enlace glicosídico obtenida mediante simulaciones de MD de diferentes complejos de <i>DmC1GalT1</i>	145
Figura 5.31. Simulaciones de MD de complejos ternarios de <i>DmC1GalT1</i>	
.....	146 – 147
Figura 5.32. Representación de la distancia entre la enzima y el glicopéptido (Å) en diferentes complejos obtenidas mediante MDs durante 0,5 μ s.....	148
Figura 5.33. Representación de la distancia entre la enzima y el glicopéptido (Å) en diferentes complejos obtenidas mediante MDs durante 0,5 μ s.....	149
Figura 5.34. Gráfica de la RMSD derivadas de las simulaciones de MD (0,5 μ s) de la forma apo y los diferentes complejos.....	150
Figura 5.35. Simulaciones de DM de los complejos de <i>DmC1GalT1</i>	151
Figura 8.1. Estructura de los aminoácidos	210

8.2. Lista de abreviaturas.

aa. Aminoácido

ADN. Ácido desoxirribonucleico

ADNc. Ácido desoxirribonucleico complementario

Antígeno ST. Neu5Aca2-3Gal β 1-3GalNAca1-O-Ser/Thr

Antígeno STn. Neu5Aca2-6GalNAca1-O-Ser/Thr

Antígeno T. Gal β 1-3NAca1-O-Ser/Thr

Antígeno Tn. GalNAca1-O-Ser/Thr

ARN. Ácido ribonucleico

ARNm. ARN mensajero

ATP. Adenosín trifosfato

Bf. Bos Taurus

C1GalT1. N-acetilgalactosamina β -1,3-galactosiltransferasa

CAZy. Enzimas activas hacia carbohidratos

Ce. Caenorhabditis elegans

Cf. Canis familiaris

Da. Daltons

Dm. Drosophila melanogaster

Dr. Danio rerio

E. coli. Escherichia coli

EGFR. Factor de crecimiento epidérmico

EGFR. Receptor del factor de crecimiento epidérmico

GalNAc. N-Acetilgalactosamina

Gg. Gallus gallus

GlcNAc. N-Acetilglucosamina

GT. Glicosiltransferasa

Hi5. HighFive™

Hs. Homo sapiens

Ig. Inmunoglobulina

IMAC. Cromatografía de afinidad de metales inmovilizados

IPTG. Isopropil-β-D-1-tiogalactopiranosido

ITC. Calorimetría de Titulación Isotérmica

K_d. Constante de disociación

LB. Caldo de Lisogenia (Lysogeny broth)

MALDI-TOF-TOF. Desorción/ionización laser asistida por matriz -tiempo de vuelo-tiempo de vuelo (Matrix assisted laser desorption/ionization-time of flight-time of flight)

MD. Dinámica molecular

Mm. Mus musculus

MPTs. Modificaciones postraduccionales

MUC. Mucina

NOE. Efecto Overhauser nuclear

PAGE. Electroforesis en gel de poliacrilamida

PBS. Tampón fosfato salino

PCR. Reacción en cadena de la polimerasa (Polymerase chain reaction)

PDB. Banco de datos de proteínas (Protein Data Bank)

PEG. Polietilenglicol

ppGalNAc-T(s) o GalNAc-T(s). N-Acetilgalactosaminiltransferasa(s) polipeptídica(s)

Pt. Pan troglodytes

RMN. Resonancia Magnética Nuclear

Rn. Rattus norvegicus

SDS. Dodecilsulfato sódico (Sodium dodecyl sulfate)

Sf9. *Spodoptera frugiperda*

SN1. Sustitución nucleófila unimolecular

SN2. Sustitución nucleófila bimolecular

SNi. Sustitución nucleófila intramolecular

SOB. Medio súper óptimo (Super optimal broth)

SOC. Medio súper óptimo con represión de catabolitos (Super optimal broth with catabolite repression)

SPPS. Síntesis de péptidos en fase sólida

STD. Diferencia de transferencia de saturación

STD-RMN. Espectroscopia de resonancia magnética nuclear de diferencia de transferencia de saturación (Saturation transfer difference – Nuclear magnetic resonance)

S-Tn. Antígeno de Thomsen Friedenreich sialilado

Tn. Antígeno de Thomsen-Friedenreich

TRIS. Tris(hidroxietil)aminometano

UDP. Uridín difosfato

UDP-galactosa. Uridina difosfato galactosa

UDP-glucosa. Uridina difosfato glucosa

UV. Ultravioleta

VNTR. *Variable number tandem repeat* / Repeticiones en tándem de número variable

WT. *Wild type* o silvestre

Xt. Xenopus tropicalis

8.3. Recetario.

LB

- 1% (w/v) triptona
- 0.5% (w/v) extracto de levadura
- 1% (w/v) NaCl

LB/Agar

- 1% (w/v) triptona
- 0.5% (w/v) extracto de levadura
- 1% (w/v) NaCl
- 1.5% (w/v) agar

LB/Agar – BluoGal

- 1% (w/v) triptona
- 0.5% (w/v) extracto de levadura
- 1% (w/v) NaCl
- 1.5% (w/v) agar
- 0,01% (w/v) BluoGal

SOC

- 2% (w/v) triptona
- 0.5% (w/v) extracto de levadura
- 0.05% (w/v) NaCl
- 10 mM MgCl₂
- 20 mM glucosa

Tampón de carga de proteínas (Laemmli)

- 24% (v/v) TRIS 1M pH 6.8
- 8% (w/v) SDS
- 40% (v/v) glicerol
- 10% (v/v) β-mercaptoetanol
- 0.1% (w/v) azul de bromofenol

PBS (1 l)

- 0,8% (w/v) NaCl
- 0,02% (w/v) KCl
- 0,144% (w/v) Na₂HPO₄
- 0,024% (w/v) KH₂PO₄

8.4. Secuencia y propiedades de las proteínas recombinantes.

Se recogen a continuación las propiedades de la proteína recombinante silvestre, las diferentes construcciones y los mutantes con las que se ha trabajado a lo largo de esta tesis. Todas las proteínas expresadas se basan en la secuencia aminoacídica original C1GalT1 de *D. melanogaster*, la cual es la siguiente:

DmC1GalT1 (Full lenght) - wt

MTANSLGRSILNEGRSNKRSFVSLIVGLIVGFCLAELFVYSTPE
RSEFMPYDGHHRHGDVNDAAHSHDMMEMSGPEQDVGGEHVH
ENSTIAERLYSEVRVLCWIMTNPSNHQKKARHVKRTWGKRCN
KLIFMSSAKDDELDAVALPVGEGRNNLWGKTKEAYKYIYEHHI
NDADWFLKADDDTYTIVENMRYMLYPYSPETPVYFGCKFKPY
VKQGYMSGGAGYVLSREAVRRFVVEALPNPKLCKSDNSGAED
VEIGKCLQNVNVLGDSRDSNGRGRFFPFVPEHHLIPSHTDKKF
WYWQYIFYKTDEGLDCCSDNAISFHYVSPNQMYVLDYLIYHLR
PYGIINTPDALPNKLAVGELMPEIKEQATESTSDGVSKRSAETKT
Q

Número de aminoácidos: 388

Peso Molecular: 44323,01 Da

pI teórico: 6,19

Número de residuos cargados negativamente (Asp + Glu): 51

Número de residuos cargados positivamente (Arg + Lys): 44

Coefficiente de extinción molar: 65780 M⁻¹cm⁻¹(a 280 nm en agua, asumiendo que todas las cisteínas están reducidas).

Las construcciones que se insertaron para su expresión en **mamíferos** fueron las siguientes (en azul se señala la cola de histidinas en el extremo C-terminal y en rojo las mutaciones que se llevaron a cabo):

DmC1GalT1 (T43-Q388) - wt

TPERSEFMPYDGHRHGDVNDAAHSHDMMEMSGPEQDVGGE
HVHENSTIAERLYSEVRVLCWIMTNPSNHQKKARHVKRTWGK
RCNKLIFMSSAKDDELDAVALPVGEGRNLLWGKTKEAYKYIY
EHHINDADWFLKADDDTYTIVENMRYMLYPYSPETPVYFGCKF
KPYVKQGYMSGGAGYVLSREAVRRFVVEALPNPKLCKSDNSG
AEDVEIGKCLQNVNVLGDSRDSNGRGRFFPFVPEHHLIPSHTD
KKFWYWQYIFYKTDEGLDCCSDNAISFHYVSPNQMYVLDYLIY
HLRPYGIINTPDALPNKLA VGELMPEIKEQATESTSDGVSKRSAE
TKTQ**GTKHHHHHHH**

Número de aminoácidos: 355

Peso Molecular: 40873,81 Da

pI teórico: 6,28

Número de residuos cargados negativamente (Asp + Glu): 49

Número de residuos cargados positivamente (Arg + Lys): 41

Coefficiente de extinción molar: 64290 M⁻¹cm⁻¹(a 280 nm en agua, asumiendo que todas las cisteínas están reducidas).

DmC1GALT1(T43-Q388) posiciones de los mutantes realizados (R152A, Y213A, Y218A, D255A, W300A y Y325A):

TPERSEFMPYDGHRHGDVNDAAHSHDMMEMSGPEQDVGGE
HVHENSTIAERLYSEVRVLCWIMTNPSNHQKKARHVKRTWGK
RCNKLIFMSSAKDDELDAVALPVGEG**A**NNLWGKTKEAYKYIY
EHHINDADWFLKADDDTYTIVENMRYMLYPYSPETPVYFGCKF
K**P****A**VKQ**G****A**MSGGAGYVLSREAVRRFVVEALPNPKLCKSDNSG
AE**A**VEIGKCLQNVNVLGDSRDSNGRGRFFPFVPEHHLIPSHTD

KKFAYWQYIFYKTDEGLDCCSDNAISFHAVSPNQMYVLDYLIY
HLRPYGIINTPDALPNKLAVGELMPEIKEQATESTSDGVSKRSAE
TKTQGTKHHHHHHH

Número de aminoácidos: 356

Peso Molecular:

- R152A: 40788,7 Da
- Y213A, Y218A e Y325A: 40781,71 Da
- D255A: 40829,8 Da
- W300A: 40758,67 Da

pI teórico:

- R152A: 6,2
- Y213A, Y218A e Y325A: 6,28
- D255A: 6,35
- W300A: 6,28

Número de residuos cargados negativamente (Asp + Glu):

- R152A: 49
- Y213A, Y218A e Y325A: 49
- D255A: 48
- W300A: 49

Número de residuos cargados positivamente (Arg + Lys):

- R152A: 40
- Y213A, Y218A e Y325A: 41
- D255A: 41
- W300A: 41

Coefficiente de extinción molar (a 280 nm en agua, asumiendo que todas las cisteínas están reducidas):

- R152A: 64290 M⁻¹cm⁻¹
- Y213A, Y218A e Y325A: 62800 M⁻¹cm⁻¹
- D255A: 64290 M⁻¹cm⁻¹
- W300A: 58790 M⁻¹cm⁻¹

Las construcciones que se insertaron para su expresión en **células de insecto** en el sistema de expresión de baculovirus fueron las siguientes

junto a una señal de “*mellitin honey bee*” (en azul se señala la cola de histidinas en el extremo C-terminal):

DmC1GalT1 (T43-Q388)

TPERSEFMPYDGHRHGDVNDAAHSHDMMEMSGPEQDVGGHE
HVHENSTIAERLYSEVRVLCWIMTNPSNHQKKARHVKRTWGK
RCNKLIFMSSAKDDELDAVALPVGEGRNNLWGKTKEAYKYIY
EHHINDADWFLKADDDTYTIVENMRYMLYPYSPETPVYFGCKF
KPYVKQGYMSGGAGYVLSREAVRRFVVEALPNPKLCKSDNSG
AEDVEIGKCLQNVNVLGDSRDSNGRGRFFPFVPEHHLIPSHTD
KKFWYWQYIFYKTDEGLDCCSDNAISFHYVSPNQMYVLDYLIY
HLRPYGIINTPDALPNKLA VGELMPEIKEQATESTSDGVSKRSAE
TKTQGGHHHHHH

Número de aminoácidos: 354

Peso Molecular: 40701,58 Da

pI teórico: 6,2

Número de residuos cargados negativamente (Asp + Glu): 49

Número de residuos cargados positivamente (Arg + Lys): 40

Coefficiente de extinción molar: 64290 M⁻¹cm⁻¹(a 280 nm en agua, asumiendo que todas las cisteínas están reducidas).

DmC1GalT1 (H55-Q388)

HRHGDVNDAAHSHDMMEMSGPEQDVGGHEHVHENSTIAERL
YSEVRVLCWIMTNPSNHQKKARHVKRTWGKRCNKLIFMSSAK
DDELDAVALPVGEGRNNLWGKTKEAYKYIY EHHINDADWFLK
ADDDTYTIVENMRYMLYPYSPETPVYFGCKFKPYVKQGYMSG
GAGYVLSREAVRRFVVEALPNPKLCKSDNSGAEDVEIGKCLQN

VNVLGDSRDSNGRGRFFPFVPEHHLIPSHTDKKFWYWQYIFY
KTDEGLDCCSDNAISFHVYVSPNQMYVLDYLIYHLRPYGIINTPD
ALPNKLAVGELMPEIKEQATESTSDGVSKRSAETKTQGGHHHH
HH

Número de aminoácidos: 342

Peso Molecular: 39291,06 Da

pI teórico: 6,37

Número de residuos cargados negativamente (Asp + Glu): 46

Número de residuos cargados positivamente (Arg + Lys): 39

Coefficiente de extinción molar: 62800 M⁻¹cm⁻¹ (a 280 nm en agua, asumiendo que todas las cisteínas están reducidas).

DmC1GalT1 (S73-Q388)

SGPEQDVGGEHVHENSTIAERLYSEVRVLCWIMTNPSNHQKK
ARHVKRTWGKRCNKLIFMSSAKDDELDAVALPVGEGRNWLW
GKTKEAYKYIYEHINDADWFLKADDDTYTIVENMRYMLYPY
SPETPVYFGCKFKPYVKQGYMSGGAGYVLSREA VRRFVVEALP
NPKLCKSDNSGAEDVEIGKCLQNVNVLGDSRDSNGRGRFFPF
VPEHHLIPSHTDKKFWYWQYIFYKTDEGLDCCSDNAISFHVYVSP
NQMYVLDYLIYHLRPYGIINTPDALPNKLAVGELMPEIKEQATE
STSDGVSKRSAETKTQGGHHHHHHH

Número de aminoácidos: 324

Peso Molecular: 37152,76 Da

pI teórico: 6,52

Número de residuos cargados negativamente (Asp + Glu): 42

Número de residuos cargados positivamente (Arg + Lys): 38

Coefficiente de extinción molar: 62800 M⁻¹cm⁻¹(a 280 nm en agua, asumiendo que todas las cisteínas están reducidas).

DmC1GalT1 (T43-P343)

TPERSEFMPYDGHRHGDVNDAAHSHDMMEMSGPEQDVGGE
HVHENSTIAERLYSEVRVLCWIMTNPSNHQKKARHVKRTWGK
RCNKLIFMSSAKDDELDAVALPVGEGRNLLWGKTKEAYKYIY
EHHINDADWFLKADDDTYTIVENMRYMLYPYSPETPVYFGCKF
KPYVKQGYMSGGAGYVLSREAVRRFVVEALPNPKLCKSDNSG
AEDVEIGKCLQNVNVLGDSRDSNGRGRFFPFVPEHHLIPSHTD
KKFWYWQYIFYKTDEGLDCCSDNAISFHYVSPNQMYVLDYLIY
HLRPGGHHHHHHH

Número de aminoácidos: 309

Peso Molecular: 35900,22 Da

pI teórico: 6,35

Número de residuos cargados negativamente (Asp + Glu): 42

Número de residuos cargados positivamente (Arg + Lys): 35

Coefficiente de extinción molar: 62800 M⁻¹cm⁻¹(a 280 nm en agua, asumiendo que todas las cisteínas están reducidas).

DmC1GalT1 (T43-P354)

TPERSEFMPYDGHRHGDVNDAAHSHDMMEMSGPEQDVGGE
HVHENSTIAERLYSEVRVLCWIMTNPSNHQKKARHVKRTWGK
RCNKLIFMSSAKDDELDAVALPVGEGRNLLWGKTKEAYKYIY
EHHINDADWFLKADDDTYTIVENMRYMLYPYSPETPVYFGCKF
KPYVKQGYMSGGAGYVLSREAVRRFVVEALPNPKLCKSDNSG
AEDVEIGKCLQNVNVLGDSRDSNGRGRFFPFVPEHHLIPSHTD

KKFWYWQYIFYKTDEGLDCCSDNAISFHYVSPNQMYVLDYLIY
HLRPYGIINTPDALP**GGHHHHHH**

Número de aminoácidos: 320

Peso Molecular: 37055,54 Da

pI teórico: 6,27

Número de residuos cargados negativamente (Asp + Glu): 43

Número de residuos cargados positivamente (Arg + Lys): 35

Coefficiente de extinción molar: 64290 M⁻¹cm⁻¹(a 280 nm en agua, asumiendo que todas las cisteínas están reducidas).

DmC1GalT1 (T43-S373)

TPERSEFMPYDGHRHGDVNDAAHSHDMMEMSGPEQDVGGE
HVHENSTIAERLYSEVRVLCWIMTNPSNHQKKARHVKRTWGK
RCNKLIFMSSAKDDELDAVALPVGEGRNLLWGKTKEAYKYIY
EHHINDADWFLKADDDTYTIVENMRYMLYPYSPETPVYFGCKF
KPYVKQGYMSGGAGYVLSREAVRRFVVEALPNPKLCKSDNSG
AEDVEIGKCLQNVNVLGDSRDSNGRGRFFPFVPEHHLIPSHTD
KKFWYWQYIFYKTDEGLDCCSDNAISFHYVSPNQMYVLDYLIY
HLRPYGIINTPDALPNKLAVGELMPEIKEQATES**GGHHHHHH**

Número de aminoácidos: 338

Peso Molecular: 39023,79 Da

pI teórico: 6,14

Número de residuos cargados negativamente (Asp + Glu): 47

Número de residuos cargados positivamente (Arg + Lys): 37

Coefficiente de extinción molar: 64290 M⁻¹cm⁻¹(a 280 nm en agua, asumiendo que todas las cisteínas están reducidas).

8.5. Información general.

8.5.1. El código genético.

Tabla 8.1. Código genético. Tabla de correspondencias entre la transcripción de los distintos codones a aminoácidos en humanos. Dependiendo de la especie pueden existir diferencias en la transcripción.

		SEGUNDO NUCLEÓTIDO									
		U		C		A		G			
PRIMER NUCLEÓTIDO	U	UUU	Fenilalanina	UCU	Serina	UAU	Tirosina	UGU	Cisteína	U	
		UUC		UCC		UAC		UGC		C	
		UUA	Leucina	UCA		UAA	Stop	UGA	Stop	A	
		UUG		UCG		UAG	Stop	UGG	Triptófano	G	
	C	CUU	Leucina	CCU	Prolina	CAU	Histidina	CGU	Arginina	U	
		CUC		CCC		CAC		CGC		C	
		CUA		CCA		CAA	Glutamina	CGA		A	
		CUG		CCG		CAG		CGG		G	
	A	AUU	Isoleucina	ACU	Treonina	AAU	Asparagina	AGU	Serina	U	
		AUC		ACC		AAC		AGC		C	
		AUA	Metionina	ACG		AAA	Lisina	AGA	Arginina	A	
		AUG		ACG		AAG		AGG		G	
	G	GUU	Valina	GCU	Alanina	GAU	Aspártico	GCU	Glicina	U	
		GUC		GCC		GAC		GGC		C	
		GUA		GCA		GAA	Glutámico	GGA		A	
		GUG		GGG		GAG		GGG		G	

8.5.2. Los aminoácidos.

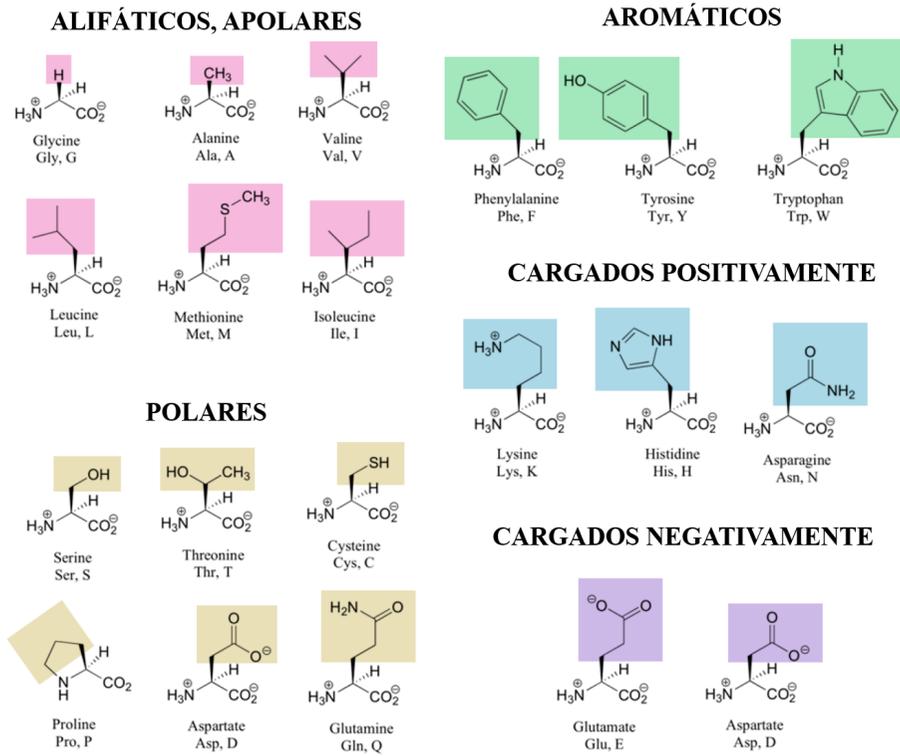


Figura 8.1. Estructura de los aminoácidos. En diferentes colores se resaltan las cadenas laterales de los diferentes aminoácidos agrupándose dependiendo de sus características bioquímicas.

*Dentro del puño encontró sin saber cómo una esfera dorada. Un
glorispren solitario. "El paso más importante que puede dar alguien.
No es el primero, ¿verdad?" "Es el próximo. Siempre el
próximo paso Dalinar"*

Juramentada. Brandon Sanderson.

