

Víctor Taleb Seral

Bases moleculares y mecanismo
de hidrólisis de mucinasas que
dependen de clústeres de O-
glicanos

Director/es

Hurtado Guerrero, Ramón

EXTRACTO

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

El presente documento es un extracto de la tesis original depositada en el Archivo Universitario.

En cumplimiento del artículo 14.6 del Real Decreto 99/2011, de 28 de enero, por el que se regulan las enseñanzas oficiales de doctorado, los autores que puedan verse afectados por alguna de las excepciones contempladas en la normativa citada deberán solicitar explícitamente la no publicación del contenido íntegro de su tesis doctoral en el repositorio de la Universidad de Zaragoza. Las situaciones excepcionales contempladas son:

- Que la tesis se haya desarrollado en los términos de un convenio de confidencialidad con una o más empresas o instituciones.
- Que la tesis recoja resultados susceptibles de ser patentados.
- Alguna otra circunstancia legal que impida su difusión completa en abierto.



Universidad de Zaragoza
Servicio de Publicaciones

ISSN 2254-7606

Tesis Doctoral [Extracto]

**BASES MOLECULARES Y MECANISMO DE
HIDRÓLISIS DE MUCINASAS QUE DEPENDEN DE
CLÚSTERES DE O-GLICANOS**

Autor

Víctor Taleb Seral

Director/es

Hurtado Guerrero, Ramón

UNIVERSIDAD DE ZARAGOZA
Escuela de Doctorado

Programa de Doctorado en Bioquímica y Biología Molecular

2023

Bases moleculares y mecanismo de hidrólisis de mucinasas que dependen de clústeres de O-glicanos

Memoria presentada por **Víctor Taleb Seral**, graduado en Química,
para optar al grado de Doctor en Ciencias.

Zaragoza 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 **Bases moleculares y mecanismo de hidrólisis de mucinas que dependen de clústeres de O-glicanos** ha sido realizada, bajo su dirección por el graduado VÍCTOR TALEB SERAL, en el Instituto de Biocomputación y Física de Sistemas Complejos y que reúne, a su juicio, las condiciones requeridas para optar al grado de Doctor en Ciencias por la Universidad de Zaragoza.

Zaragoza, octubre de 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 (ARAIID), 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 ALBA en Barcelona (línea BL13 XALOC). Durante dos de los cuatro años de realización de esta tesis (desde agosto de 2021 hasta octubre de 2023, 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 (PRE2020-094130) (BOE del 05/10/2020) y cofinanciada por el Fondo Social Europeo.

Después de cuatro años parece mentira haber llegado al final de esta etapa. Ha sido un camino lleno de baches, experimentos fallidos, noches en vela y mucha cafeína. Pero echando la vista atrás, ha merecido la pena.

Tengo que agradecer a Ramón por acogerme para hacer la tesis y darme la opción de descubrir lo interesante que es el mundo de los azúcares. A mis compañeros de grupo Andrés y Javi que han sufrido a mi lado las noches en el ALBA y demás contratiempos pero que gracias a ellos esta experiencia ha sido mucho más entretenida con nuestras charletas de rol. A los nuevos cadetes David, Irene, María y Billy, que llevan menos tiempo, pero que cogen las riendas del laboratorio con fuerza. También quiero agradecer al resto de compañeros del BIFI por hacer más ameno y agradable el camino.

A Dani E., David V. y Darío F., nos vemos poco pero cada vez que quedamos me recargáis las pilas de energía para seguir a tope. Mis compañeros de aventuras, Jesús, Cris, Julio, Javi, Aitor, Andrés y Ros que me habéis dado la oportunidad de disfrutar de un mundo de fantasía con vosotros cada semana en el que las risas y el buen rollo son constantes.

Por supuesto agradezco a mi familia, que me apoyaron desde el principio, que me dieron fuerzas para seguir y para perseguir mis sueños durante todos estos años. A mi familia adoptiva que me apoyan en la distancia y aunque nos vemos cuando podemos os llevo en el corazón. A Odín, la bola de pelo que ya es parte de nuestra familia y que solo con sus ronroneos ya me hace sentir mejor

Y como no, Ana, la que más ha sufrido mis desvaríos y mis frustraciones, sin quien esto hubiera sido solo una etapa más en la vida y ella lo ha convertido en mis mejores años, no sé hasta donde habría llegado sin tener el apoyo que me has dado todo este tiempo.

A todos los que me habéis seguido en este sendero, Gracias.

CONTENIDO

ÍNDICE DE FIGURAS	7
ÍNDICE DE TABLAS.....	15
CAPÍTULO 1. INTRODUCCIÓN	17
1.1 VISIÓN GENERAL.....	17
1.2. MUCINAS.....	20
1.2.1. Características y tipos de mucinas	20
1.2.2. O-glicosilación de tipo GalNAc	25
1.2.3. Funciones de las mucinas	26
1.3. IMPLICACIÓN EN LA PROTECCIÓN FRENTE A INFECCIONES POR PARTE DE LA MUCOSA.....	28
1.3.1. Mucinas presentes en el sistema respiratorio.....	28
1.3.2. Mucinas presentes en el sistema digestivo	29
1.4. MECANISMOS DE DEGRADACIÓN DE MUCINAS	31
1.4.1. Visión general de la degradación.....	31
1.4.2. Mucinasas: Tipos y características	33
1.5. ORGANISMOS ADAPTADOS A LA MUCOSA INTESTINAL CON MUCINASAS QUE RECONOCEN CLÚSTERES DE O-GLICANOS	35
1.5.1. <i>Akkermansia Muciniphilia</i> y su interacción con la mucosa intestinal	35
CAPÍTULO 2. OBJETIVOS	39
CAPÍTULO 3. MATERIALES Y MÉTODOS	41
3.1. AM0627.....	41
3.1.1. Técnicas de biología molecular	41
3.1.1.1. Síntesis de los plásmidos de las construcciones AM0627 ²¹⁻⁵⁰⁶ , AM0627 ⁷¹⁻⁵⁰⁶ y sus mutantes	41
3.1.1.2. Transformación en células competentes <i>E. coli</i> BL21(DE3).....	43

3.1.1.3. Conservación de las colonias transformadas	44
3.1.2. Expresión de AM0627 y sus mutantes	44
3.1.3. Purificación de proteínas AM0627 y sus mutantes	45
3.1.3.1. Lisis celular y separación de la fracción insoluble	45
3.1.3.2. Purificación por cromatografía de afinidad con HisTrap (IMAC)	46
3.1.3.3. Electroforesis desnaturante en gel de acrilamida (SDS-PAGE)	48
3.1.3.4. Intercambio de tampón mediante HiPrep 26/10 Desalting	50
3.1.3.5. Proteólisis de la HisTag y eliminación	52
3.1.3.6. Concentración y cuantificación	53
3.1.3.7. Eliminación de impurezas y agregados mediante cromatografía de exclusión molecular	55
3.1.4. Preparación y síntesis de los glicopéptidos empleados con AM0627	57
3.1.4.1. Síntesis de glicopéptidos en fase sólida (SPPS)	57
3.1.4.2. Síntesis de glicopéptidos con antígenos T	60
3.1.5. Técnicas estructurales: Cristalografía de rayos X.....	61
3.1.5.1. Generación de cristales	61
3.1.5.2. Difracción de rayos X y toma de datos	65
3.1.5.3. Procesamiento y refinamiento de la estructura cristalográfica	67
3.1.6. Ensayos de proteólisis.....	72
3.1.6.1. Ensayos de proteólisis de los glicopéptidos <i>in vitro</i>	72
3.1.6.2. Ensayos de proteólisis con <i>reporters</i>	72
3.1.7. Técnicas computacionales	74
3.1.7.1. Simulaciones de dinámica molecular de AM0627 ⁷¹⁻⁵⁰⁶ y BT4224	74
3.1.7.2. Simulaciones QM/MM del complejo AM0627 ⁷¹⁻⁵⁰⁶ -P1-Zn ²⁺	76
CAPÍTULO 4. RESULTADOS	81
4.1. AM0627.....	81

4.1.1. Características generales de la estructura del complejo AM0627 ⁷¹⁻ ⁵⁰⁶ E326A-P1-Zn ²⁺	81
4.1.2. Detalles de la interacción entre AM0627 y el bis T O-glicopéptido P1	88
4.1.3. Determinación de los residuos clave para la actividad de AM0627	92
4.1.4. Caracterización del epítipo mínimo de interacción en P1 para el reconocimiento por parte de AM0627	97
4.1.5. Estudio de la estabilidad de los complejos de AM0627 con P1, P2 y P9.	103
4.1.6. Estudio del mecanismo de reacción de AM0627 a nivel computacional .	106
4.1.6.1. Mecanismo guiado por el ataque nucleofílico directo de Glu326	107
4.1.6.2. Mecanismo guiado por el ataque nucleofílico indirecto mediado por una molécula de H ₂ O.....	109
4.1.7. Comparación estructural de AM0627 frente a otras mucinasas y O- glicoproteasas	111
4.1.7.1. AM0627 y las mucinasas BT4244, ZmpB, ZmpC y IMPa.....	111
4.1.7.2. AM0627 y la O-glicoproteasa OgpA.....	114
4.1.8. Ensayos de actividad de BT4244 por O-glicopéptidos con bis Tn y bis T O- glicanos.....	116
4.1.9. Estudio de la capacidad de AM0627 de acomodar residuos de ácido siálico en el sitio activo.....	120
CAPÍTULO 5. DISCUSIÓN Y CONCLUSIONES	125
5.1. AM06227.....	125
CAPÍTULO 6. REFERENCIAS	131
CAPÍTULO 7. ABREVIATURAS	145
CAPÍTULO 8. APENDICES	151
8.1. MEDIOS DE CULTIVO Y RECETAS	151
8.2 PROPIEDADES DE LAS PROTEÍNAS RECOMBINANTES	152
8.2.1 AM0627 ²¹⁻⁵⁰⁶	153

8.2.2. AM0627 ⁷¹⁻⁵⁰⁶	154
8.3. LOS AMINOÁCIDOS	155

ÍNDICE DE FIGURAS

Figura 1. Representación del epitelio intestinal.....	18
Figura 2. Diagrama de la mucosa intestinal durante una infección	20
Figura 3. Representación de los distintos dominios que componen las mucinas que forman geles extracelulares.....	22
Figura 4. Representación de una red tridimensional formada por mucinas...	23
Figura 5. Representación de las mucinas transmembrana MUC1, MUC3, MUC4, MUC12, MUC13, MUC16 y MUC17	24
Figura 6. Representación simplificada de las vías de O-glicosilación de tipo GalNAc	26
Figura 7. Representación de las mucinas y su implicación en la defensa del sistema respiratorio	29
Figura 8. Representación de los distintos estratos de la mucosa en el intestino grueso.....	30
Figura 9. Representación de los distintos dominios que componen MUC2..	31
Figura 10. Representación de las hidrolasas implicadas en la degradación de proteínas tipo mucina.....	32
Figura 11. Ejemplos de mucinasas diferenciadas y clasificadas.....	34
Figura 12. Mapa del vector pMALc2x	422
Figura 13. Secuencia de aminoácidos de la construcción AM0627 ²¹⁻⁵⁰⁶	433
Figura 14. Esquema de la expresión de proteína en <i>E. coli</i>	455
Figura 15. Esquema del funcionamiento de una columna IMAC.....	477
Figura 16. Cromatograma de la purificación de AM0627 ⁷¹⁻⁵⁰⁶	4848
Figura 17. SDS-PAGE ejemplo del corte del sitio TEV para AM0627	500
Figura 18. Esquema de la separación de proteína y tampón.....	511

Figura 19. Esquema de la separación de la proteasa TEV y la MBP-His de AM0627.....	533
Figura 20. Filtración en gel de AM0627 ⁷¹⁻⁵⁰⁶	577
Figura 21. Representación de los glicopéptidos P1 a P9	5858
Figura 22. Reacción de C1GalT1.....	600
Figura 23. Diagrama de fases.....	622
Figura 24. Cristales de AM0627	655
Figura 25. Representación del fenómeno de la difracción.....	666
Figura 31. Secuencia y estructura secundaria de AM0627	82
Figura 32. Análisis SDS-PAGE de la actividad de la enzima <i>wildtype</i> AM0627 ²¹⁻⁵⁰⁶ y el truncado AM0627 ⁷¹⁻⁵⁰⁶	83
Figura 33. Estructura cristalina del complejo AM0627 ⁷¹⁻⁵⁰⁶ E326A-P1-Zn ²⁺	85
Figura 34. Vista en detalle del sitio activo de AM0627	87
Figura 35. Detalle de las interacciones en el sitio activo del complejo AM0627 ⁷¹⁻⁵⁰⁶ E326A-P1-Zn ²⁺	89
Figura 36. Detalle de las interacciones de Tyr470 con el glicopéptido	90
Figura 37. Evolución temporal de los ensayos de corte de <i>wildtype</i> AM0627 ²¹⁻⁵⁰⁶ y los diferentes mutantes a alanina frente a P1	93
Figura 38. Espectros de MALDI-TOF de los ensayos de corte <i>in vitro</i> del AM0627 ²¹⁻⁵⁰⁶ y sus mutantes frente a P1	94
Figura 39. Evolución temporal de los ensayos de corte de AM0627 ²¹⁻⁵⁰⁶ frente a los glicopéptidos P1 a P9	99
Figura 40. Espectros de MALDI-TOF de los ensayos de corte <i>in vitro</i> de AM0627 ²¹⁻⁵⁰⁶ frente a los glicopéptidos P1 a P9	100

Figura 41. Evolución de las distancias (His324-Nε···Zn ²⁺ , His329-Nε···Zn ²⁺ y NR291···OQ4 y NR291···NQ4) en las simulaciones de dinámica molecular de AM0627 ⁷¹⁻⁵⁰⁶ en complejo con P1, P2 y P9, así como la estructura general de estos complejos	104
Figura 42. Conformaciones representativas del clúster más poblado de los complejos AM0627 ⁷¹⁻⁵⁰⁶ -P2-Zn ²⁺ y AM0627 ⁷¹⁻⁵⁰⁶ -P1-Zn ²⁺	105
Figura 43. Esquema de los dos mecanismos propuestos para la hidrólisis de enlaces peptídicos catalizada por AM0627.....	106
Figura 44. Perfil de energía libre y estados del mecanismo de hidrólisis de AM0627 obtenido de simulaciones de QM/MM para el ataque nucleofílico directo.	108
Figura 45. Perfil de energía libre y estados del mecanismo de hidrólisis de AM0627 obtenido de simulaciones de QM/MM para el ataque nucleofílico asistido por una molécula de agua	110
Figura 46. Análisis de los sitios activos de BT4244, ZmpB, IMPa y ZmpC.	113
Figura 47. Análisis del sitio activo de OgpA.....	115
Figura 48. Análisis SDS-PAGE de la actividad de AM0627 y BT4244 frente a <i>reporters</i> de MUC1	118
Figura 49. Conjunto de estructuras obtenidas de 0,5 μs de las simulaciones de MD junto con los valores de RMSD.....	120
Figura 50. Detalle de las distancias entre los posibles glicanos sialilados y AM0627 ⁷¹⁻⁵⁰⁶ E326A así como la representación de la superficie electrostática de AM0627 ⁷¹⁻⁵⁰⁶ E326A	121
Figura 74. Clasificación de los aminoácidos según sus propiedades.....	155155

ÍNDICE DE TABLAS

Tabla 1. Coeficientes de extinción y pesos moleculares para cada construcción de AM0627.	555
Tabla 2. Recopilación de datos y estadísticas del refinamiento de AM0627 ⁷¹⁻⁵⁰⁶ E326A en complejo con P1 y Zn ²⁺	700
Tabla 6. Resumen de las interacciones entre el Zn ²⁺ y la molécula de agua con los residuos de AM0627 y P1.	88
Tabla 7. Resumen de las interacciones entre los residuos de AM0627 y el O-glicopéptido P1.	91

CAPÍTULO 1. INTRODUCCIÓN

1.1 VISIÓN GENERAL

La supervivencia es una de las mayores prioridades de un organismo, así que debe protegerse frente a otros organismos que podrían amenazar su supervivencia (Hansson, 2020).

El cuerpo humano está expuesto a numerosas infecciones y microorganismos tanto simbióticos como patogénicos. La piel ofrece una gran protección frente a infecciones, pero las zonas más expuestas de nuestro organismo son aquellas que necesitan hacer un intercambio de nutrientes, gases o residuos con el exterior, por lo que no pueden utilizarla como defensa (Linden et al., 2008). Es por ello que se necesitan sistemas dinámicos que mantengan a los patógenos a raya.

Todos los órganos más expuestos de nuestro organismo, presentan una capa de mucinas unidas a la membrana de las células epiteliales y además secretan otras mucinas que forman una capa externa a las células (**Figura 1**). Estas mucinas son proteínas altamente O-glicosiladas, siendo esta O-glicosilación una modificación postraducciona que les permite hacer de barrera física y de señuelo para evitar que los microorganismos penetren en los tejidos que protegen (Linden et al., 2008).

La biosíntesis de la O-glicosilación está bien estudiada (Schjoldager et al., 2020) así como el papel de la misma. En particular, este tipo de modificación

postraduccional les da a las mucinas funciones especiales que les permiten formar geles y además les proporcionan resistencia a la degradación mediada por proteasas convencionales (Wagner et al., 2018).

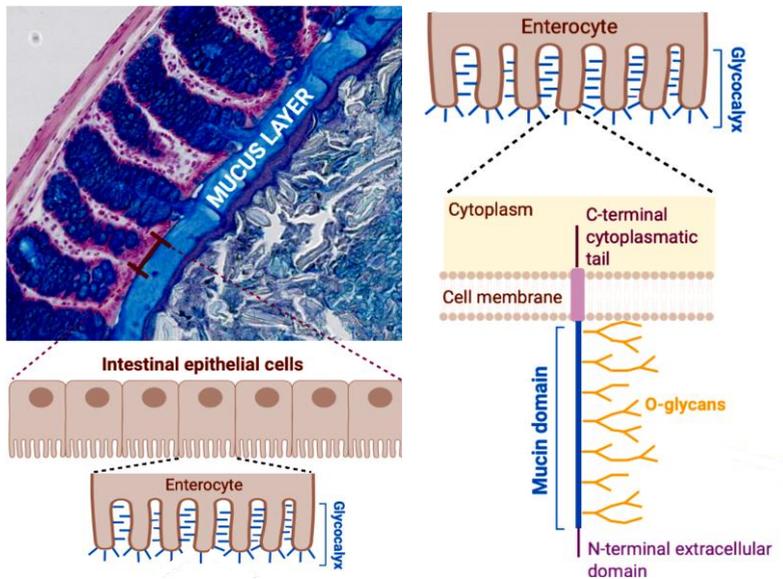


Figura 1. Representación del epitelio intestinal donde se encuentran las mucinas transmembrana y mucinas secretadas que conforman una capa externa que está expuesta a los microorganismos. Modificada de (Paone y Cani, 2020).

Por ello, varios organismos se han adaptado para cortar y degradar específicamente este tipo de estructuras, usando una batería de O-glicoproteasas específicas de dominios tipo mucinas o mucinasas (Shon et al., 2021).

En este momento se conocen muy pocas mucinasas, y todas son de origen bacteriano, la mayoría de las cuales son capaces de cortar solo en presencia de

un solo sitio de O-glicosilación. Cabe mencionar que además de mucinasas, las bacterias secretan otro tipo de enzimas para degradar los O-glicanos como glicosidasas o sulfatasas (M. E. V. Johansson, Ambort, et al., 2011; Raba & Luis, 2023).

Dado que las mucinas contienen clústeres de O-glicanos contiguos y en mucha menor medida O-glicanos simples, es difícil de entender que no se hayan descubierto mucinasas que reconozcan estos clústeres. Por ello sería de gran interés descubrir este tipo de mucinasas y entender cómo funcionan a nivel molecular. Esta información nos permitirá ahondar en cómo se renueva y recicla la mucosa a la vez que dará pistas de como los organismos patógenos son capaces de atravesar esta barrera protectora e infectar al huésped, tal como se puede ver en la **Figura 2** (Linden et al., 2008). En esta tesis se presentan algunas de estas mucinasas y cómo llevan a cabo su actividad enzimática.

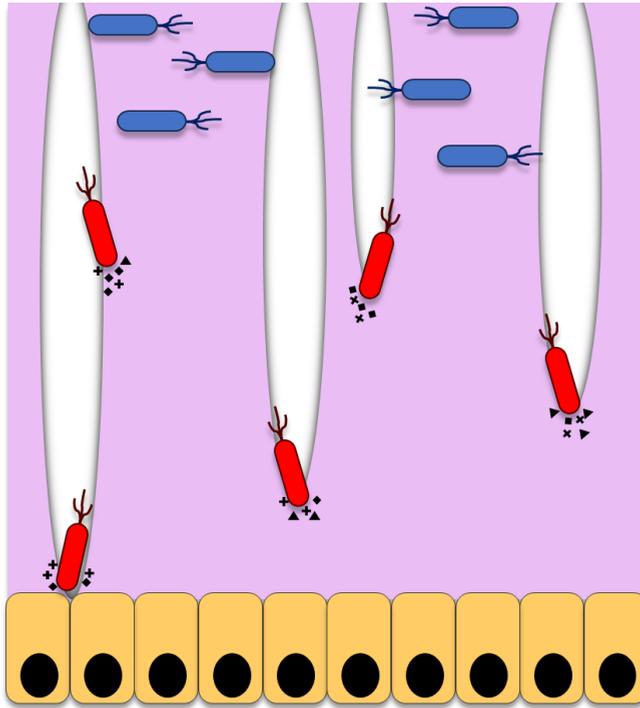


Figura 2. Diagrama de la mucosa intestinal durante una infección donde las bacterias comensales en color azul viven en la capa más externa de la mucosa, mientras las bacterias patógenas en color rojo se abren paso a través de la mucosa para llegar al epitelio utilizando una batería de mucinasas, glicosidasas y sulfatasas. Modificada de (Linden et al., 2008).

1.2. MUCINAS

1.2.1. Características y tipos de mucinas

Las mucinas son unas O-glicoproteínas de gran peso molecular producidas en las células epiteliales, cuya principal característica es que están altamente

glicosiladas, correspondiendo más de un 50% de su peso principalmente a los O-glicanos de su estructura. Estas son las proteínas más abundantes en los fluidos corporales y se encuentran localizadas en todas las mucosas del cuerpo como las que se encuentran en los pulmones o el intestino.

Su secuencia aminoacídica contiene unas regiones denominadas secuencias repetidas en tándem (TR) que se repiten numerosas veces y tienen la característica de contener una alta proporción de serinas y treoninas junto con prolinas, denominando a estas secuencias como PTS (Hansson, 2020). Los grupos hidroxilo de las serinas y treoninas sirven a las GalNAc-transferasas del aparato de Golgi como puntos de inicio de la O-glicosilación de tipo GalNAc llevada a cabo por un gran número de enzimas de esta familia (de las Rivas et al., 2019), mientras las prolinas aseguran una estructura desplegada para que estas GalNAc-transferasas puedan acceder a los sitios de O-glicosilación. Estos O-glicanos pueden estar separados o adyacentes generando clústeres, confiriendo a las mucinas una estructura tipo escobilla rígida compuesta por una cadena polipeptídica desplegada de donde se ramifican las O-glicosilaciones confiriéndoles protección frente a las proteasas convencionales (Shon et al., 2021).

Hoy en día se conocen numerosas mucinas presentes en humanos las cuales se pueden clasificar en dos grupos diferenciados según sean secretadas o ancladas a la membrana celular.

Entre las mucinas secretadas se encuentran MUC7 y MUC20, que son proteínas solubles mientras que otras como MUC2, MUC5AC, MUC5B, y MUC6 en humanos, forman geles extracelulares. Estas mucinas que forman

estratos gelificados contienen dominios tipo D del factor von Willebrand (VWD) en el extremo N-terminal, seguido del dominio PTS los cuales contienen insertados dominios ricos en cisteína (CysD) y pueden contener dominios tipo C del factor von Willebrand (VWC) seguidos finalmente del dominio C-terminal rico en cisteína (Hansson, 2020). Estos dominios se pueden ver representados en la **Figura 3**. Previamente a la O-glicosilación, todas las mucinas gelificantes forman dímeros en el retículo endoplasmático mediante puentes disulfuros generados a través de sus dominios C-terminal.

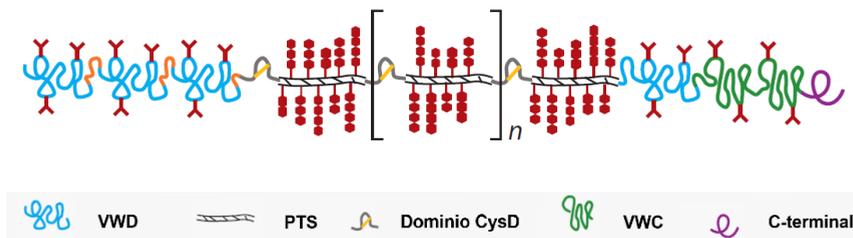


Figura 3. Representación de los distintos dominios que componen las mucinas que forman geles extracelulares. Los dominios se encuentran representados en los siguientes colores: la región VWF-D en azul, el dominio PTS en blanco, los dominios CysD en amarillo, el dominio VWF-C en verde y el dominio C-terminal en morado. Modificado de (Wagner et al., 2018).

Los geles extracelulares formados por las mucinas son estructuras oligoméricas o redes tridimensionales capaces de actuar como un filtro protector. En la **Figura 4** se puede ver una representación de las interacciones entre las mucinas que las lleva a formar este tipo de estructuras. Entre algunas de ellas se encuentran interacciones de tipo puente disulfuro previamente mencionados, interacciones hidrofóbicas e interacciones electrostáticas de repulsión entre las cargas negativas de los O-glicanos (Wagner et al., 2018).

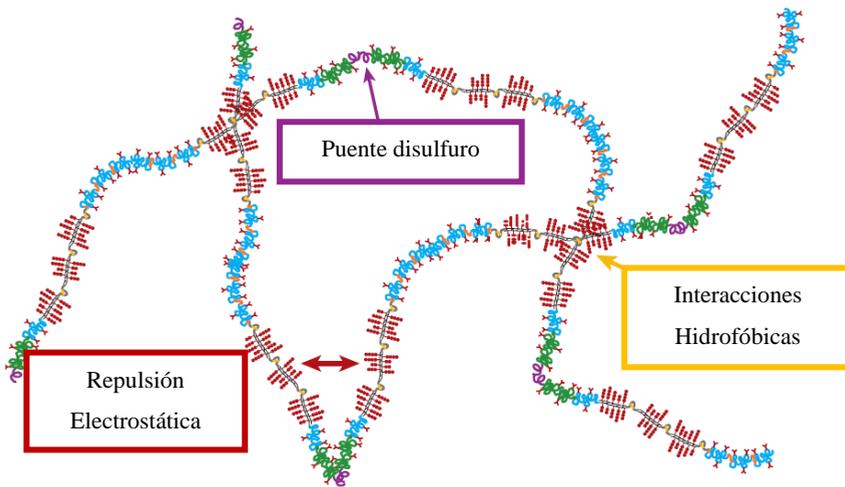


Figura 4. Representación de una red tridimensional formada por mucinas donde se muestran las distintas interacciones entre los dominios de los monómeros. Modificado de (Wagner et al., 2018).

En el grupo de mucinas transmembrana encontramos a MUC1, MUC3, MUC4, MUC12, MUC13, MUC15, MUC16, MUC17, MUC21 y MUC22. Estas poseen un dominio transmembrana, una cola citoplasmática, pueden tener un dominio de auto proteólisis y el dominio N-terminal extracelular donde se encuentran los TRs como se puede observar en la **Figura 5** (Supruniuk y Radziejewska, 2021).

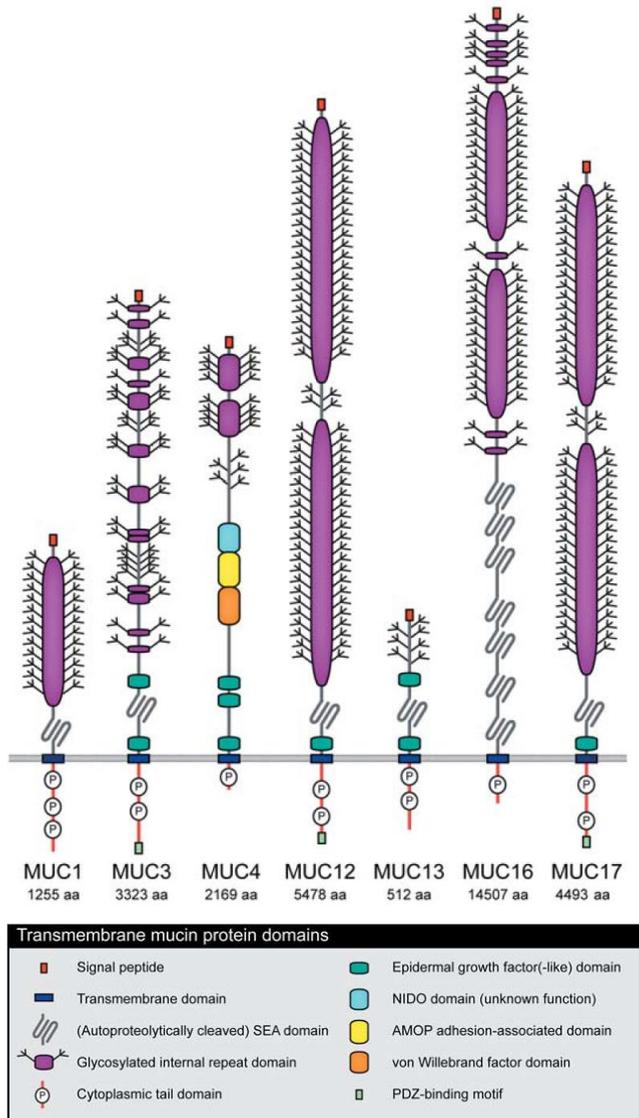


Figura 5. Representación de las mucinas transmembrana MUC1, MUC3, MUC4, MUC12, MUC13, MUC16 y MUC17 donde se pueden observar los distintos dominios que las componen. Modificado de (Van Putten & Strijbis, 2017).

1.2.2. O-glicosilación de tipo GalNAc

La O-glicosilación empieza con la adición de una molécula de N-Acetil galactosamina (GalNAc) a aminoácidos tipo serina o treonina (GalNAc α Ser/Thr; también llamado antígeno Tn). A partir de esta estructura se derivan las cuatro O-glicosilaciones de tipo GalNAc más frecuentes producidas por el cuerpo humano.

El primer tipo de O-glicosilación o *Core 1*, se forma al añadir una galactosa (Gal) al antígeno Tn (Gal β 1-3GalNAc α Ser/Thr), el *Core 2* se forma al añadir un grupo de N-Acetil glucosamina (GlcNAc) al *Core 1* (GlcNAc β 1-6(Gal β 1-3) GalNAc α Ser/Thr), el *Core 3* se forma al añadir GlcNAc al antígeno Tn (GlcNAc β 1-3GalNAc α Ser/Thr) y, por último, el *Core 4* se forma al añadir GlcNAc al *Core 3* (GlcNAc β 1-6(GlcNAc β 1-3) GalNAc α Ser/Thr). Estas estructuras son posteriormente modificadas por la adición de diferentes azúcares para formar los O-glicanos complejos.

Además, se encuentran otro tipo de estructuras conocidas como antígenos de carbohidratos asociados a cáncer o TACAs por sus siglas en inglés. Se les denomina de esta manera ya que en tejidos cancerígenos la O-glicosilación puede estar afectada y estas estructuras forman parte de la glicosilación aberrante que se observa en células tumorales (Schjoldager et al., 2020).

Entre ellas se encuentran los antígenos T, Tn y STn. Los antígenos T y Tn corresponden respectivamente a las estructuras Gal β 1-3GalNAc α Ser/Thr o *Core 1* y GalNAc α Ser/Thr. Mientras que el antígeno STn se forma al añadir un ácido siálico (Sia) sobre la base de GalNAc para la formación de un Sia α 2-

6GalNAc α Ser/Thr. Todas estas estructuras se encuentran representadas en la **Figura 6**.

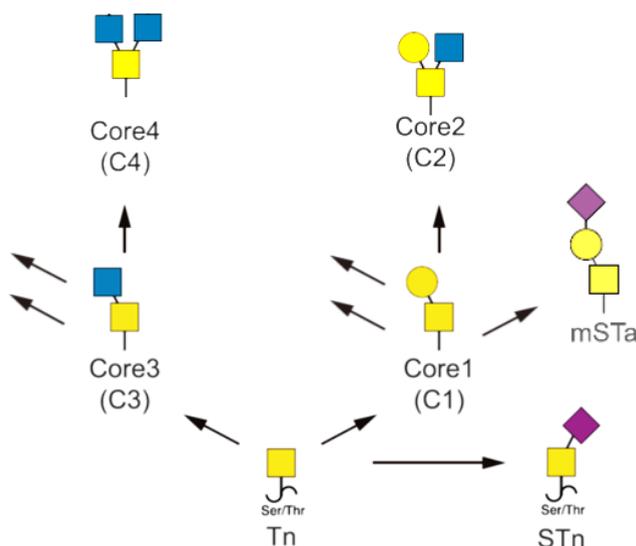


Figura 6. Representación simplificada de las vías de O-glicosilación de tipo GalNAc donde se pueden encontrar señaladas las estructuras de los *Core* 1-4 así como de los TACAs y otras estructuras sialiladas. Modificado de (Taleb et al., 2022).

1.2.3. Funciones de las mucinas

Las mucinas presentan distintos tipos de funciones gracias a su estructura altamente O-glicosilada y están localizadas en la superficie y el exterior celular.

Entre dichas funciones se encuentran la capacidad de regular la señalización célula-célula, orientar la parte apical de la célula, su implicación en el proceso

de adhesión celular, la formación de la mucosa y la limpieza de la misma (Wagner et al., 2018). Por último, una de sus funciones prioritarias es servir de barrera protectora frente a microorganismos patógenos, así como de hogar para la microbiota (Paone y Cani, 2020).

Están a su vez fuertemente relacionadas con cáncer ya que la O-glicosilación suele estar alterada en esta enfermedad dando lugar a mucinas con O-glicosilaciones aberrantes que modifican sus propiedades de adhesión facilitando la metástasis, modificando su señalización celular y protegiendo a las células cancerígenas frente al sistema inmune (Nath y Mukherjee, 2014; Van Putten y Strijbis, 2017).

También pueden controlar la fisiología y patogenización de organismos comensales. Este es el caso de organismos como *Streptococcus mutans* o *Candida albicans* cuyos biofilms se ven inhibidos por la presencia de MUC5B (Frenkel y Ribbeck, 2015) o MUC5AC respectivamente (Kavanaugh et al., 2014). También son capaces de disminuir la competitividad entre organismos comensales para favorecer la formación de entornos multiespecie cuando resulta positivo para el huésped (Frenkel y Ribbeck, 2017).

1.3. IMPLICACIÓN EN LA PROTECCIÓN FRENTE A INFECCIONES POR PARTE DE LA MUCOSA

1.3.1. Mucinas presentes en el sistema respiratorio

Los órganos más expuestos al exterior y por lo tanto a cualquier posible infección son aquellos que componen el sistema respiratorio y el digestivo.

El sistema respiratorio está expuesto a millones de bacterias constantemente y una de las maneras que tiene para protegerse es mediante los cilios que tienen las células de su epitelio.

La membrana de estas células ciliadas está cubierta de mucinas como MUC1, MUC4 y MUC16 que mantienen el agua alrededor de los cilios gracias a su alta densidad de O-glicanos. Esto les permite reducir el rozamiento, ayudando al movimiento continuo de los cilios que elimina microorganismos de las vías respiratorias. Además, actúan como un cepillo, barriendo los microorganismos y protegen al epitelio de las posibles infecciones actuando como barrera (Hansson, 2020).

También se encuentran glándulas submucosales (**Figura 7 a**) que generan un polímero de MUC5B con hasta 25 μm de diámetro (**Figura 7 b**). Este limpia la superficie del epitelio y además elimina a los microorganismos. Cómo se generan estos polímeros y el control de sus movimientos todavía es desconocido (Hansson, 2019; Wagner et al., 2018). En caso de una infección, se genera una capa de mucosa formada por MUC5AC y MUC5B (Hansson,

2020) capaz de separar las bacterias del epitelio (**Figura 7 c**) parecida a la que se encuentra en el intestino grueso (Fernández-Blanco et al., 2018).

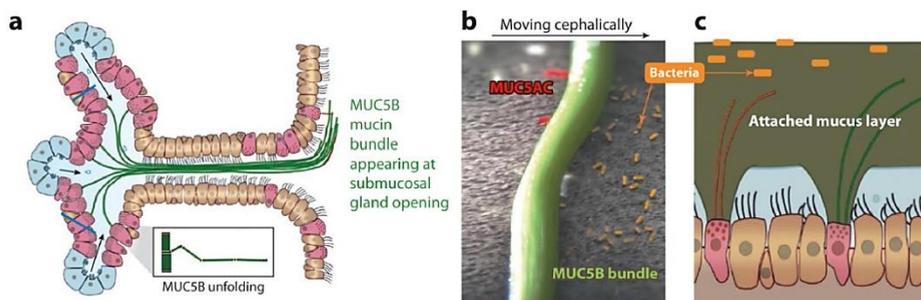


Figura 7. Representación de las mucinas y su implicación en la defensa del sistema respiratorio. a. Glándula submucosal donde se observa la formación del polímero de MUC5B a partir de monómeros, b. Imagen de microscopía electrónica de superficie donde se ve el polímero de MUC5B y las bacterias que son eliminadas de la superficie del epitelio, y c. generación de la capa de mucosa frente a la infección. Modificada de (Hansson, 2020).

1.3.2. Mucinas presentes en el sistema digestivo

En el sistema digestivo y concretamente en los intestinos delgado y grueso se pueden encontrar diferentes maneras de protección, ya que el intestino delgado solo tiene una capa de MUC2 en la membrana de las células mientras que el intestino grueso genera una mucosa de gran espesor con dos zonas diferenciadas, una capa interna con una red densa de mucinas y una capa externa con una red menos densa que le confiere más movilidad, representadas en la **Figura 8** (Paone y Cani, 2020).

En el intestino grueso, la MUC2 localizada en la capa interna de la mucosa forma redes oligoméricas mediante puentes disulfuro, formando una película

protectora de tal manera que en raras ocasiones hay contacto entre las bacterias y el epitelio, ya que estas habitan en la capa externa de la mucosa (M. E. V. Johansson, Holmén Larsson, et al., 2011; M. E. V Johansson et al., 2008).

La capa externa está formada por MUC2, pero forma una red menos densa (Paone y Cani, 2020) donde las mucinas ya han sido digeridas parcialmente (Kufe, 2009) y donde habitan las bacterias comensales. Estas se dedican a degradar y regenerar las mucinas que llegan a las capas más externas contribuyendo así a la homeostasis y regeneración de la mucosa.

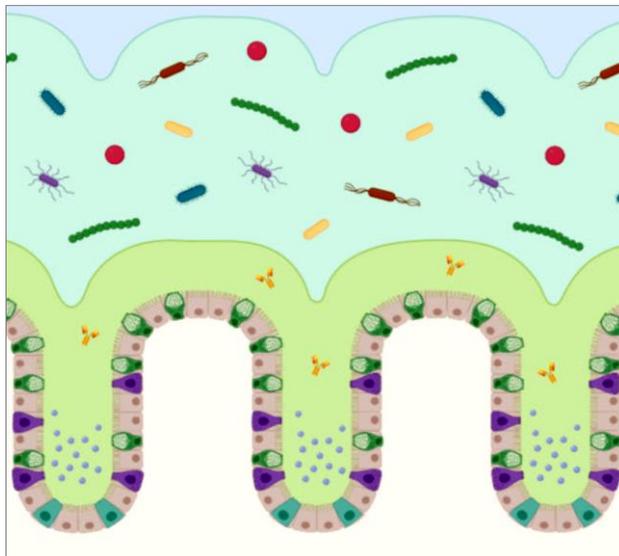


Figura 8. Representación de los distintos estratos de la mucosa en el intestino grueso donde se puede observar una capa externa (azul) donde habita la microbiota y una capa interna (verde) que las separa del epitelio. Modificado de (Paone y Cani, 2020).

Por otro lado, los patógenos intentan penetrar de la capa externa a la capa más interna de la mucosa para llegar al epitelio y así poder infectar al huésped

(Linden et al., 2008). Cuando llegan a la superficie celular, los patógenos cortan las mucinas transmembrana en su intento de infectar al huésped, lo que lleva a una cascada de señalización generando una respuesta a la infección (Birchenough et al., 2016; Linden et al., 2008).

Debido al alto contenido de O-glicanos contiguos en la secuencia consenso de los TRs de MUC2 (Svensson et al., 2018), estos microorganismos tienen que ser capaces de degradar clústeres de glicanos de diferentes tamaños como los que se muestran en la **Figura 9**.

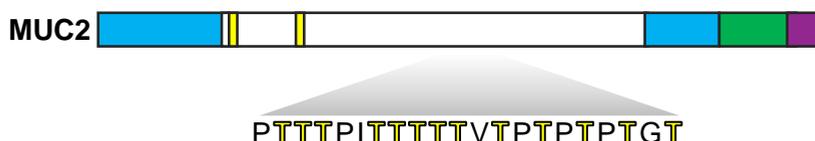


Figura 9. Representación de los distintos dominios que componen MUC2 representados en los siguientes colores: la región VWD en azul, el dominio PTS en blanco, los dominios CysD en amarillo, el dominio VWC en verde y el dominio C-terminal en morado. Se encuentra representada la secuencia consenso de los TR en la región PTS con los puntos de O-glicosilación indicados en amarillo.

1.4. MECANISMOS DE DEGRADACIÓN DE MUCINAS

1.4.1. Visión general de la degradación

Los microorganismos que habitan el intestino utilizan principalmente la fibra de la dieta como fuente de energía. Cuando una persona sigue una dieta baja en fibra, la microbiota es capaz de adaptarse para consumir las mucinas

presentes en el organismo huésped, así como hacen los microorganismos patógenos para llegar al epitelio (Desai et al., 2016; Paone y Cani, 2020).

Como ya ha sido indicado, se pensaba que las mucinas estaban protegidas frente a la degradación gracias a su estructura de escobilla rígida, que impide el acceso de las proteasas a la cadena polipeptídica escondida bajo los densos O-glicanos. Sin embargo, se ha visto que hay varios tipos de hidrolasas que ayudan a su degradación: glicosidasas que atacan los enlaces entre azúcares, proteasas que degradan la parte no O-glicosilada de las mucinas y O-glicoproteasas, enzimas que degradan proteínas con O-glicanos (**Figura 10**). Dentro de esta familia se encuentran las mucinasas, enzimas solo encontradas en bacterias que están diseñadas específicamente contra los dominios protegidos tipo mucina (Laville et al., 2019; Shon et al., 2021).

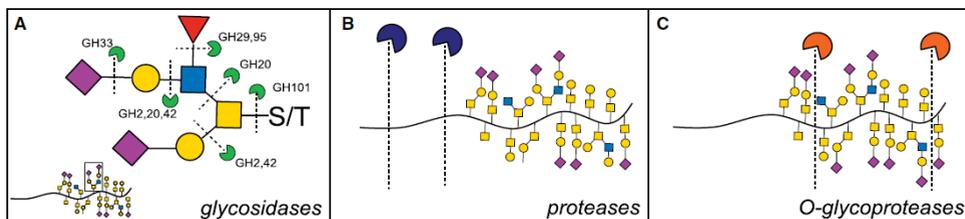


Figura 10. Representación de las hidrolasas implicadas en la degradación de proteínas tipo mucina con grandes clústeres de O-glicanos. Modificado de (Shon et al., 2021).

Estas enzimas son las que usan organismos patógenos como StcE de *Escherichia coli* (*E. coli*) o ZmpC de *Streptococcus pneumoniae* para romper la capa de mucosa y acceder al epitelio para infectar al huésped. Sin embargo,

las bacterias comensales del intestino también tienen este tipo de enzimas, que les ayuda a digerir las mucinas de la capa externa de la mucosa, reciclándola y contribuyendo a su renovación (Shon et al., 2021).

1.4.2. Mucinasas: Tipos y características

Las mucinasas son un conjunto de enzimas recientemente descubiertas clasificadas en familias MEROPS (Rawlings et al., 2016) según su secuencia y maquinaria catalítica. También se pueden diferenciar según sus motivos de corte y su tolerancia a los patrones de O-glicosilación. Como se puede observar en la **Figura 11** aquellas mucinasas que provienen de bacterias patógenas tienen una tolerancia más amplia a los O-glicanos siendo capaces de reconocer también O-glicanos complejos ya que tienen que abrirse camino a través de la capa interna de mucosa mientras que aquellas mucinasas que provienen de bacterias comensales están más restringidas en los patrones de glicanos que pueden cortar, ya que las usan para alimentarse de las mucinas de la capa superior de la mucosa ya predigeridas y con O-glicanos más simples (Shon et al., 2021).

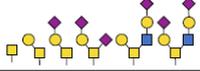
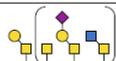
Protease	Organism	MEROPS	Cleavage motif	Glycan
BT4244	<i>Bacteroides thetaiotaomicron</i>	M60	X T/S*	
AM0627	<i>Akkermansia muciniphila</i>	M60	T/S* T/S*	
AM0908	<i>Akkermansia muciniphila</i>	M60	T/S T/S*	
AM1514	<i>Akkermansia muciniphila</i>	M98	T/S T/S*	
ZmpB	<i>Clostridium perfringens</i>	M60	X T/S*	
IMPa	<i>Pseudomonas aeruginosa</i>	M88	X T/S*	
ZmpC	<i>Streptococcus pneumoniae</i>	M26	T/S* X X X X	
StcE	<i>Escherichia coli</i> O157:H7	M66	T/S* X T/S	
CpaA	<i>Acinetobacter</i>	M72	P T/S*	n.d.
OgpA	<i>Akkermansia muciniphila</i>	M11	X T/S*	
Pic	Enteroaggregative <i>Escherichia coli</i>	S6	s T/S*	

Figura 11. Ejemplos de mucinasas diferenciadas y clasificadas según su organismo de origen, familia MEROPS, sitio de corte y tolerancia a estructuras O-glicosiladas con distinto grado de complejidad. Modificado de (Shon et al., 2021).

La mayoría de estas mucinasas son metalopeptidasas dependientes de zinc, lo que las clasifica como zincin (Cerdà-Costa y Gomis-Rüth, 2014). Estas están caracterizadas por tener una hélice en el sitio activo con una secuencia corta de unión a metal HEXXH. Las histidinas proporcionan sitios de unión al metal mientras que el glutamato actúa como la base catalítica (Cerdà-Costa y Gomis-Rüth, 2014; Shon et al., 2021). Actualmente se conocen dos mecanismos

posibles para metalopeptidasas dependientes de zinc, dependiendo si el ataque nucleofílico al enlace peptídico lo realiza un glutamato o es intermediado a través de una molécula de agua (Cerdà-Costa y Gomis-Rüth, 2014).

El dominio C-terminal tiene grandes variaciones entre las zincins y está principalmente caracterizado por la presencia de un aminoácido que es capaz de proporcionar un tercer sitio para la unión del metal. Uno de los clanes principales de las zincin son las gluzincins, las cuales contienen un glutamato que aporta este tercer residuo de unión al metal (Cerdà-Costa y Gomis-Rüth, 2014).

1.5. ORGANISMOS ADAPTADOS A LA MUCOSA INTESTINAL CON MUCINASAS QUE RECONOCEN CLÚSTERES DE O-GLICANOS

1.5.1. *Akkermansia Muciniphilia* y su interacción con la mucosa intestinal

Akkermansia muciniphila es una bacteria comensal que habita en el intestino (Derrien et al., 2004), capaz de degradar las mucinas de la capa externa de la mucosa donde habita. Es un organismo simbiote presente desde el primer mes de vida, alcanzando cantidades comparables a la etapa adulta en tan solo un año (Collado et al., 2007), y llega a representar hasta el 4% de bacterias localizadas en el colon (Derrien et al., 2008). Su presencia se ha relacionado inversamente con la obesidad y el autismo (Belzer y De Vos, 2012).

Contiene varias mucinasas descritas hoy en día en la literatura, como son AM0908, AM1514 y AM0627 (Shon et al., 2021). Esta última pertenece a la familia MEROPS M60 de metaloproteasas (Shon et al., 2021), las cuales suelen pertenecer a organismos de la microbiota capaces de degradar las mucinas del huésped como fuente de energía como BT4244 de *Bacteroides thetaiotaomicron* (Shon et al., 2021) aunque hay algunos casos de bacterias patogénicas que contienen este tipo de mucinasas como ZmpB de *Clostridium perfringens* como se ve indicado en la **Figura 11, apartado 1.4.2.**

AM0627 forma parte de las gluzincin, ya descritas anteriormente, y es capaz de cortar entre dos serinas o treoninas O-glicosiladas y es la única descrita hasta la fecha que reconoce y proteoliza entre dos residuos O-glicosilados contiguos (Shon et al., 2020), lo cual resulta de un gran interés ya que conseguir entender su mecanismo ayudaría a entender la degradación de los grandes clústeres de O-glicanos presentes en las mucinas.

CAPÍTULO 2. OBJETIVOS

En base a toda la información introducida previamente se plantearon una serie de objetivos generales y específicos para esta tesis doctoral.

Puesto que el conocimiento sobre como es el proceso de degradación de las mucinas está muy limitado hoy en día, uno de los objetivos generales de esta tesis es el de elucidar el mecanismo por el que se degradan los clústeres de O-glicanos y por ello es de suma importancia estudiar enzimas que puedan ofrecer información acerca del reconocimiento de más de un solo sitio de glicosilación como AM0627, la única mucinasa descrita hasta la fecha capaz de reconocer y cortar dos O-glicanos contiguos.

Por ello, los objetivos específicos planteados durante el transcurso de la tesis respecto a la enzima AM0627 han sido los siguientes:

- Obtención de la estructura cristalina del complejo formado por AM0627 con un sustrato O-diglicopeptídico.
- Determinación de los residuos clave para la actividad de la AM0627.
- Caracterizar el epítipo mínimo de interacción con los O-glicanos para llevar a cabo su actividad.
- Elucidación del mecanismo de hidrólisis de AM0627.
- Comparación estructural de AM0627 frente a otras mucinasas dependientes de un solo O-glicano.

CAPÍTULO 3. MATERIALES Y MÉTODOS

3.1. AM0627

3.1.1. Técnicas de biología molecular

3.1.1.1. Síntesis de los plásmidos de las construcciones AM0627²¹⁻⁵⁰⁶, AM0627⁷¹⁻⁵⁰⁶ y sus mutantes

La secuencia de ADN que codifica los aminoácidos 21 a 506 de la enzima AM0627 fue optimizada y sintetizada por GenScript (EEUU) para su expresión en *E. coli*. En el extremo 5' la construcción contenía un sitio de corte para *EcoRI*, una cola de 12 histidinas y un sitio de corte de la proteasa del virus del mosaico del tabaco (TEV). Además, en el extremo 3' contenía un sitio de corte para *Sall*. La construcción fue clonada en el plásmido pMALc2x (**Figura 12**) el cual tiene un sitio de resistencia a ampicilina y aparte genera una proteína de fusión entre la proteína de unión a maltosa (MBP) y nuestra proteína de interés (**Figura 13**) expresándolas en el citoplasma, dando lugar al vector pMALc2x-12His-TEV-AM0627²¹⁻⁵⁰⁶. El sitio de corte TEV está colocado entre la MBP-12His y la enzima AM0627²¹⁻⁵⁰⁶. Todos los mutantes de AM0627²¹⁻⁵⁰⁶ (Y288A, D318A, W321A, E326A, N347A, R362A, F390A, Y470A y V389A) fueron generados mediante mutagénesis dirigida por GenScript (EEUU) usando el vector pMALc2x-12His-TEV-AM0627²¹⁻⁵⁰⁶ previamente generado.

Siguiendo el mismo protocolo se optimizó y sintetizó la secuencia de ADN que codifica los aminoácidos 71 a 506 de la enzima AM0627 junto con la cola de histidinas y el sitio de corte TEV que fueron clonados en el plásmido pMALc2x, dando lugar al vector pMALc2x-12His-TEV-AM0627⁷¹⁻⁵⁰⁶. De igual manera el mutante AM0627⁷¹⁻⁵⁰⁶E326A fue generado mediante mutagénesis dirigida por GenScript (EEUU) usando el vector pMALc2x-12His-TEV-AM0627⁷¹⁻⁵⁰⁶.

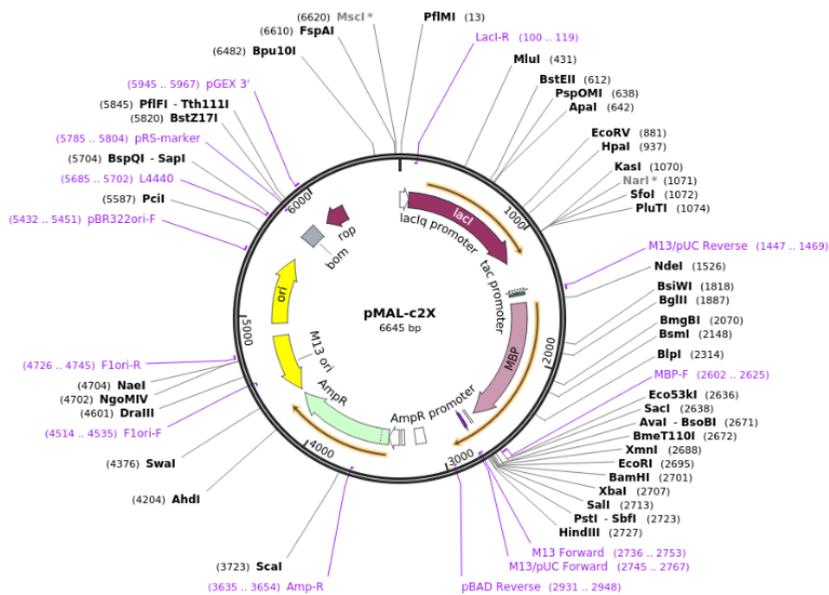


Figura 12. Mapa del vector pMALc2x usado por GenScript para la creación de los plásmidos.

Fuente: <https://www.addgene.org>.

MKIKTGARILALSALTTMMFSASALAKIEEGKLVWINGDKGYNGLAEVGKKFEKDTGIKVTVEHP
 DKLEEKFPQVAATGDGPDIIHWAHDFRGGYQAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLI
 AYPPIAVEALSIIYKDLLPNPPKTWEEIPALDKELKAKGKSALMFLQEPYFTWPLIAADGGYAF
 KYENGYDIKDVGVNAGAKAGLTFVLVLIKNKHMNADTDYSIAEAFNKGETAMTINGPWAW
 SNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENYLLTDEGLEAVNKD
 KPLGAVALKSYEEELAKDPRIAATMENAQKGEIMPNIQMSAFWYAVRTAVINAASGRQTVDEA
 LKDAQTRITKGGGGGGHHHHHHHHHHHHGGGGGGENLYFQGGGGGANTPEHIGNDLKLFK
 DSSCTSLKPDVKNTSAFQSDAMKELATKILAGHYKPDYLYAEYRALPSRQGTGKNLRIGDGFSSK
 YDNMTGVVYLEKGRHVVLVGKTEGQEISLLLPLNMRKPAEGVQPTKDPNGWGLHKKQIPLKEGI
 NIIDVETPANAYISYFTEDAGKAPKIPVHFVTGKANGYFDTTRGDTNKDWVRLDQAVSPIMDAR
 GKYIQVAYPVEFLKFKTKDRGTELINAYDKLIGIQYQLMGLDKYGKIPENRVLARVNFNYMFRD
 GDGVAYLGNMGTMRMVTDPENVLKGDACWGFSGHEVGHVMQMRPMTWGGMTEVSNNIFSLQ
 AAAKTGNESRLKRQGSYDKARKEIIEGEIAYLQSKDVFNKLVPWLWQLHLYFTKNGHPDFYPDVM
 EYLRNAGNYGGNDTVKYQFEFVKACCDVTKTDLTDFFEKWGFFKPGKFHIGDYAQYDFNVT
 PEMVEETKKWIAGKGYPKPETDITELSE

Figura 13. Secuencia de aminoácidos para la construcción de AM0627²¹⁻⁵⁰⁶ utilizada durante el desarrollo de esta tesis doctoral. La MBP se encuentra coloreada en cian, la cola de histidinas en rosa, el sitio de corte TEV en verde y AM0627²¹⁻⁵⁰⁶ en gris.

3.1.1.2. Transformación en células competentes *E. coli* BL21(DE3)

Cada plásmido fue transformado en células de la cepa *E. coli* BL21(DE3) mediante choque térmico.

Se añadieron 200 ng de plásmido a 300 µL de células competentes BL21(DE3), incubando la mezcla en hielo durante 30 minutos y dándole después un choque térmico a 42 °C durante un minuto. Después se dejaron reposar en hielo dos minutos tras lo cual se añadieron 500 µL de medio LB (1% (w/v) triptona, 0,5% (w/v) extracto de levadura, 1% (w/v) NaCl) y se incubaron a 37 °C agitando a 180-220 rpm entre una y tres horas. El último paso fue centrifugar las células 5 minutos a 2000 rpm, retirar los 500 µL de medio de recuperación añadido y sembrar las células en placas de LB/Agar (1% (w/v) triptona, 0,5% (w/v) extracto de levadura, 1% (w/v) NaCl y 1,5% (w/v) agar) añadiendo 100 µg/mL de ampicilina (Carbosynth) de tal manera

que se seleccionaron solo las células que han integrado el plásmido y se dejaron incubar *overnight* (o/n) a 37 °C para que aparecieran las colonias.

3.1.1.3. Conservación de las colonias transformadas

Las colonias que integraron correctamente los plásmidos se añadieron a 10 mL de medio LB con 100 µg/mL de ampicilina y se incubaron a 37 °C o/n agitando a 180 rpm. Al día siguiente en esterilidad, se generaron stocks de las células transformadas mezclando 700 µL de cultivo junto con 300 µL de glicerol al 80% (v/v) y se almacenaron a -80 °C para su posterior uso.

3.1.2. Expresión de AM0627 y sus mutantes

Para el crecimiento a gran escala de AM0627 y sus mutantes se inocularon células en un matraz de 100 mL de medio 2XTY (1,6% (w/v) triptona, 1% (w/v) extracto de levadura y 0,5% (w/v) NaCl) con 100 µg/mL de ampicilina, y se incubaron a 37 °C o/n agitando a 180 rpm. Al día siguiente en esterilidad, se inoculó un matraz con 2 L de 2XTY con el precultivo de 100 mL y se dejó incubar a 37 °C y 180 rpm hasta alcanzar una densidad óptica a 600 nm (O.D.) de 0,6-0,8. En ese momento se indujo la expresión de la proteína añadiendo 1 mM de isopropil-β-D-1-tiogalactopiranosido o IPTG (Carbosynth) y se redujo la temperatura de incubación a 18 °C, manteniendo la expresión durante 20 horas. Pasado este tiempo las células fueron recogidas mediante centrifugación a 6500 g y 4 °C durante 10 minutos tras lo cual fueron reservadas a -20 °C para su posterior uso, descartando el sobrenadante (**Figura 14**).

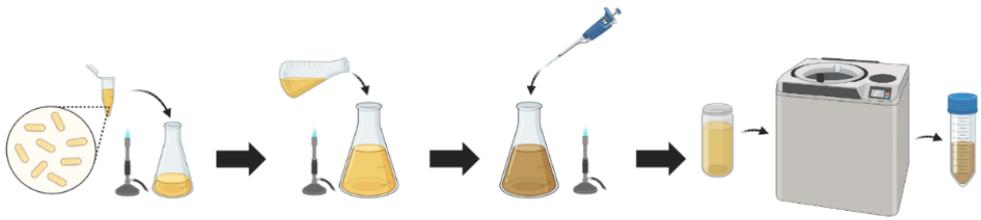


Figura 14. Esquema de la expresión de proteína en *E. coli*. De izquierda a derecha: Preparación de precultivos, inoculación de los cultivos, inducción con IPTG y centrifugación de los cultivos para obtener el pellet. Imagen generada en <https://www.biorender.com>.

3.1.3. Purificación de proteínas AM0627 y sus mutantes

3.1.3.1. Lisis celular y separación de la fracción insoluble

Cuando se trabaja con expresión en el citoplasma de *E. coli* el primer paso es romper las células para liberar la proteína del citoplasma. Las células se resuspendieron en tampón A (25 mM Tris pH 8, 500 mM NaCl y 10 mM imidazol) en una relación de 15 mL de tampón A por litro crecido de cultivo. Por cada 50 mL de células resuspendidas se añadió 500 μ L de una mezcla de inhibidores de proteasas (20 mM Fluoruro de fenilmetilsulfonilo (PMSF), 1 mM Leupeptina y 200 mM Benzamidina) para evitar la degradación de las proteínas y posteriormente se procedió a lisar las células.

Las células fueron lisadas mediante ultrasonidos usando un sonicador (Sonics VibraCell™) aplicando 10 ciclos con 30 segundos de sonicación y 30 segundos de descanso a un 80% de amplitud en hielo. Tras esto se centrifugó el lisado a 19500 rpm y 4 °C durante 30 minutos en una centrifuga Avanti J-

26XP (Beckman Coulter) con un rotor JA-20 para separar la parte insoluble del extracto crudo, el cual se pasó por un filtro de 0,45 μm (Scharlau).

3.1.3.2. Purificación por cromatografía de afinidad con HisTrap (IMAC)

La cromatografía de afinidad es una de las técnicas más empleadas para la purificación de proteínas, en la que se expone una muestra heterogénea a una matriz con afinidad específica por la proteína de interés. En este caso se usó la cromatografía de afinidad por metales inmovilizados o IMAC, donde en la columna se encuentra Ni^{2+} inmovilizado en la matriz que interacciona con la cola de histidinas que tiene la proteína de manera específica. De esta manera se consigue retener la proteína de interés en la columna mientras las impurezas son descartadas (**Figura 15**). Este tipo de purificaciones se suelen hacer en presencia de NaCl para evitar posibles interacciones inespecíficas de la matriz de la columna con las impurezas de la muestra.

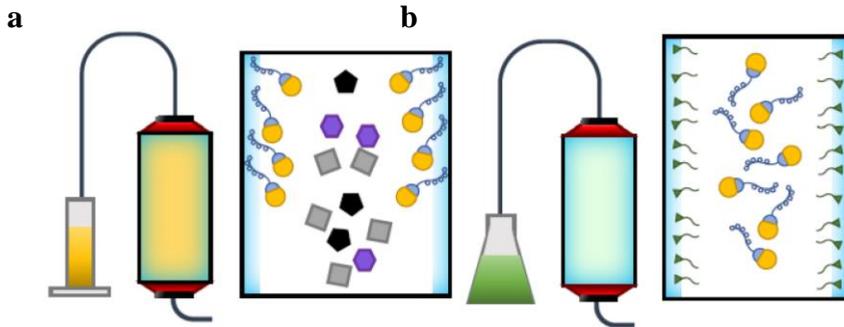


Figura 15. Esquema del funcionamiento de una columna IMAC donde a. la proteína de interés (amarillo) interacciona con la matriz a través de la cola de histidinas (azul) y las impurezas (negro, gris y morado) son eliminadas de la muestra b. El imidazol inyectado en la columna compite con la cola de histidinas separándola de la matriz.

Una vez filtrado el extracto crudo, se inyectó a 3 mL/min en una columna HisTrap HP (Cytiva) de afinidad a Ni^{2+} previamente equilibrada con tampón A. Una vez inyectada toda la muestra se pasaron 70-80 mL de tampón A para limpiar la columna de impurezas tras lo que se eluyó en gradiente de imidazol a tampón B (25 mM Tris pH 8, 500 mM NaCl y 500 mM imidazol) donde el imidazol compite por la proteína y la libera de la matriz mostrándose como un pico en el cromatograma (**Figura 16**).

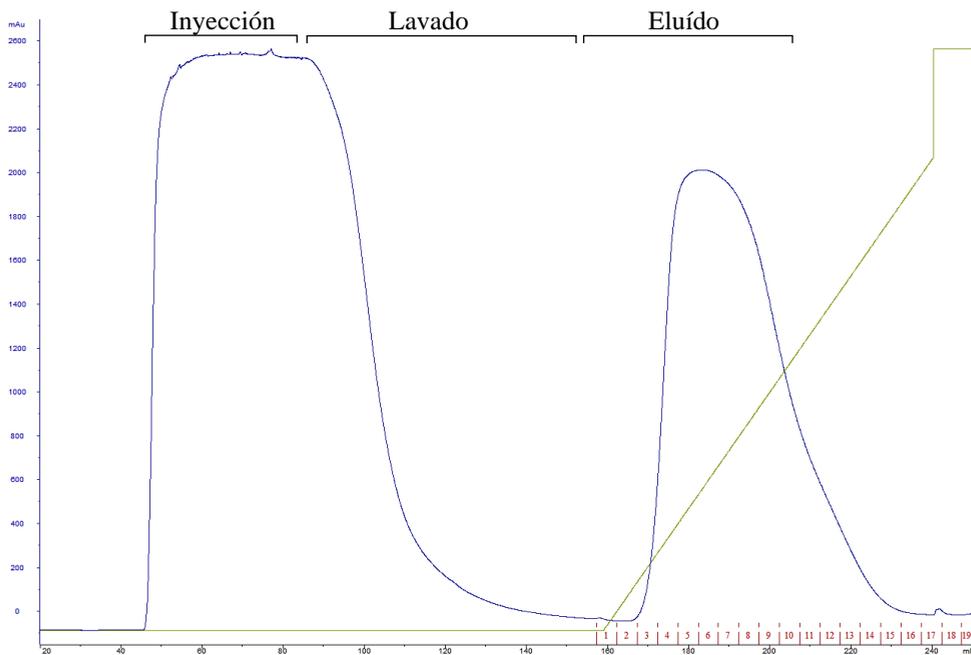


Figura 16. Cromatograma de la purificación de AM062771-506 donde se puede observar la inyección de la muestra, el lavado de la columna y el posterior pico de proteína eluida.

3.1.3.3. Electroforesis desnaturante en gel de acrilamida (SDS-PAGE)

Para confirmar la pureza de la muestra se realizó una electroforesis desnaturante en geles de acrilamida con dodecilsulfato de sodio (SDS) o SDS-PAGE (Brunelle y Green, 2014). Esta técnica nos permite separar las proteínas en función de su peso molecular haciéndolas pasar por un gel de acrilamida al someterlas a una diferencia de potencial, previamente cargando la proteína negativamente.

Para el SDS-PAGE se cargan 18 μ L de muestra junto con 6 μ L de tampón de carga laemmli (24% (v/v) Tris 1 M pH 6,8, 8% (w/v) SDS, 40% (v/v) glicerol,

10% (v/v) β -mercaptoetanol, 0,1% (w/v) azul de bromofenol) y se calientan a 95 °C durante 5 minutos. El SDS que contiene el tampón de carga es un detergente aniónico que se une a las proteínas de manera inespecífica confiriéndoles una relación peso/carga constante para que su migración solo dependa del peso y no de las cargas inherentes de las proteínas.

Posteriormente se cargan las muestras en un gel de acrilamida junto a un patrón de pesos moleculares como el Page Ruler Protein Ladder (Thermo Fisher). Se someten las muestras a un voltaje de 130 V durante aproximadamente 80 minutos y las proteínas se desplazan a una velocidad inversamente proporcional a su peso separando así las bandas más ligeras de las más pesadas.

El porcentaje de acrilamida en el gel dependerá del tamaño de la proteína de interés, ya que se resuelven las bandas de pesos inversamente relacionados al porcentaje de acrilamida presente. Ya que AM0627 tiene un peso de 55 kDa, se usaron geles de 12% de acrilamida.

Una vez migradas las proteínas se puede teñir el gel con una disolución de tinción como el Quick Coomassie (Neo-Biotech) para observar las bandas de proteína presentes en la muestra (**Figura 17**).



Figura 17. SDS-PAGE ejemplo del corte del sitio TEV donde se puede ver el marcador de pesos en el primer carril, AM0627⁷¹⁻⁵⁰⁶-12His-MBP junto con AM0627⁷¹⁻⁵⁰⁶ y 12His-MBP en el segundo carril, y AM0627⁷¹⁻⁵⁰⁶ una vez separada de la 12His-MBP en el tercer carril.

3.1.3.4. Intercambio de tampón mediante HiPrep 26/10 Desalting

Una vez se observa qué fracciones tienen la proteína de interés, hay que eliminar el imidazol del tampón utilizando una columna de HiPrep 26/10 Desalting (Cytiva).

La matriz de esta columna contiene canales de distintos tamaños por donde ha de pasar la muestra de tal manera que al pasar la proteína de interés esta pasa a una velocidad mayor al tener una ruta directa solo por los canales grandes mientras que moléculas pequeñas como aquellas que componen el tampón serán ralentizadas por los canales más pequeños y separadas de la proteína que saldrá en el tampón en el que estaba equilibrada la columna inicialmente (**Figura 18**).

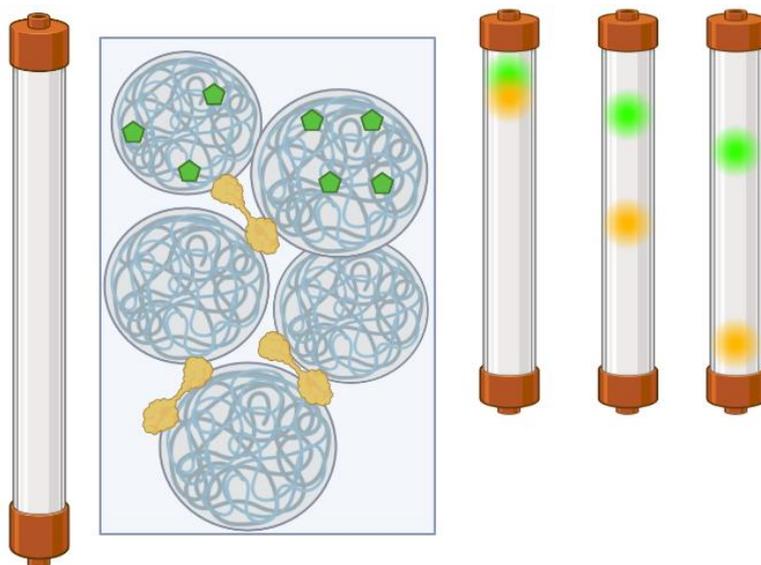


Figura 18. Esquema de la separación de proteína y tampón gracias a los diferentes canales de la matriz. Imagen generada en <https://www.biorender.com>.

La columna deberá ser equilibrada con tampón D (25 mM Tris pH 8, 150 mM NaCl). Una vez lista la columna se procede a inyectar un máximo de 15 mL de muestra ya que, si se trabaja con volúmenes mayores, los picos de proteína y tampón inicial podrían solaparse. La columna separa la proteína del tampón de tal manera que se tendrá un pico inicial de UV a 280 correspondiente a la proteína y un segundo pico de conductividad que corresponde al tampón inicial. En caso de tener más de 15 mL de muestra se lavará la columna con 70-80 mL de tampón D hasta que se observe una disminución y estabilización tanto de la conductividad como de la señal de UV a 280 nm. A partir de este punto se procederá con la siguiente inyección de 15 mL, repitiendo el proceso hasta terminar el volumen de muestra.

3.1.3.5. Proteólisis de la HisTag y eliminación

Una vez se obtuvo la proteína en tampón D, se procedió a eliminar la MBP-His con la proteasa TEV-His la cual contiene una HisTag. Esta proteasa corta una secuencia específica localizada de tal manera que se separa nuestra cola de histidinas junto con la MBP para no tener ya una proteína de fusión sino solo la proteína de interés AM0627.

Se cuantificó la muestra por Bradford (Biorad) para conocer la concentración en mg/mL de proteína a la cual se le añadió la proteasa TEV-His en un ratio de 1/30 (w/w) con respecto a la proteína de interés. Se tomaron muestras previas a añadir la proteasa de modo que se pudo comprobar el corte mediante un gel SDS-PAGE y se dejó la mezcla incubando a 18 °C o/n para que actuara la proteasa.

Una vez comprobado el corte se procedió a la purificación de la proteína ya sin HisTag, para lo cual se utilizó una columna HisTrap HP (Cytiva).

Esta vez se utilizó tampón D para equilibrar la columna tras lo cual se inyectó la muestra recogiendo todo aquello que no es capturado por la columna o *flow-through*, ya que la proteína al no contener la HisTag no interacciona con la matriz mientras que la MBP-His y la proteasa TEV-His se quedan adheridas a la columna, separándolas de la proteína de interés. Una vez se inyectó toda la muestra se pasó algo de tampón D para asegurar la recuperación de toda la proteína de interés tras lo cual pasamos un 10% de tampón B, permitiéndonos recuperar la proteína en caso de que interaccione débilmente con la matriz mediante histidinas superficiales. Finalmente se pasó un 100% de tampón B para despegar la MBP-His y la TEV-His de la columna (**Figura 19**).

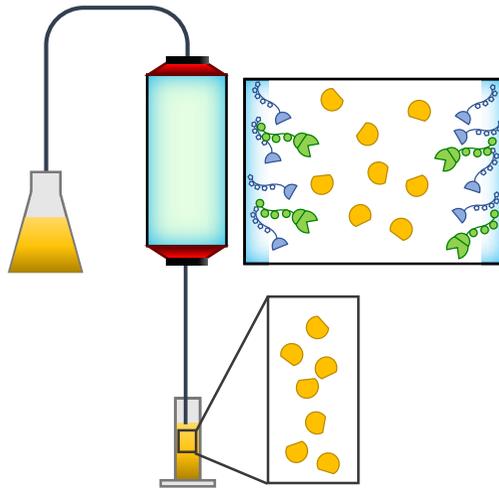


Figura 19. Esquema de la separación de la proteasa TEV (verde) y la MBP-His (Azul) de AM0627 (amarillo).

Se comprobó la pureza con un gel SDS-PAGE.

3.1.3.6. Concentración y cuantificación

Para concentrar la muestra y tener volúmenes de trabajo más pequeños se utilizaron unos concentradores amicon de 15 mL (Merck Millipore). Estos concentradores tienen una membrana con un tamaño de poro determinado, en este caso de 30 kDa, de tal manera que la proteína al ser de mayor tamaño no atraviesa la membrana.

Los amicones se equilibraron en H₂O MiliQ (MQ) y posteriormente en tampón D, se cargaron con hasta 15 mL de muestra y se centrifugaron aplicando 4000 rpm y 4 °C en una centrifuga Beckman Coulter Allegra X-30R. Este paso se

repitió hasta tener el volumen deseado y en caso de interés se puede cambiar el tampón de la muestra diluyendo la muestra en el tampón de interés y centrifugando, repitiendo este proceso por lo menos 3 veces.

Para conocer la concentración de las proteínas se pasó a la cuantificación, para lo cual se utilizó un nanodrop DeNovix Serie DS-11FX para medir la absorbancia de cada muestra a 280 nm. Conociendo el coeficiente de extinción teórico de las proteínas junto con su peso molecular, calculados con ProtParam (Gasteiger et al., n.d.) a partir de la secuencia de cada proteína (**Tabla 1**), se puede calcular su concentración molar y transformarla a mg/mL.

Una vez cuantificadas las muestras, si no van a usarse inmediatamente ni van dirigidas a la generación de cristales, se someten a una congelación rápida en nitrógeno líquido o *flash freeze* y se reservan a -80 °C para su posterior uso.

Tabla 1. Coeficientes de extinción y pesos moleculares para cada construcción de AM0627.

Construcción	Coeficiente de extinción ($M^{-1} \cdot cm^{-1}$)	Peso molecular (Da)
AM0627 ²¹⁻⁵⁰⁶ WT	78980	55107,62
AM0627 ²¹⁻⁵⁰⁶ Y288A	77490	55015,52
AM0627 ²¹⁻⁵⁰⁶ D318A	78980	55063,61
AM0627 ²¹⁻⁵⁰⁶ W321A	73480	54992,48
AM0627 ²¹⁻⁵⁰⁶ E326A	78980	55049,58
AM0627 ²¹⁻⁵⁰⁶ N347A	78980	55064,59
AM0627 ²¹⁻⁵⁰⁶ R362A	78980	55022,51
AM0627 ²¹⁻⁵⁰⁶ F390A	78980	55031,52
AM0627 ²¹⁻⁵⁰⁶ Y470A	77490	55015,52
AM0627 ²¹⁻⁵⁰⁶ V389A	78980	55079,56
AM0627 ⁷¹⁻⁵⁰⁶ WT	77365	49945,70
AM0627 ⁷¹⁻⁵⁰⁶ E326A	77365	49887,66

3.1.3.7. Eliminación de impurezas y agregados mediante cromatografía de exclusión molecular

En caso de que se quiera generar cristales, es importante que la muestra tenga la mayor pureza posible y no tenga agregados que podrían interferir en la formación de cristales. Para ello la muestra se pasa por una columna de exclusión molecular HiLoad 26/60 Superdex 75 pg (Cytiva).

El mecanismo de esta columna es muy similar a la HiPrep 26/10 Desalting, separando moléculas por tamaño, pero su resolución de separación es de proteínas entre 3 kDa y 70 kDa, de tal manera que somos capaces de separar proteínas en distintos picos según su tamaño. Esto permite eliminar las

impurezas de distinto tamaño y posibles agregados solubles ya que estos tienen un gran tamaño en comparación con la proteína.

El volumen máximo de inyección para no perder resolución en los picos son 5 mL por lo que si la muestra es de mayor volumen se procederá a concentrarla como se ha explicado en el **apartado 3.1.3.6**.

Una vez concentrada la muestra se inyectó en la HiLoad 26/60 Superdex 75 pg previamente equilibrada en tampón D y se dejó pasar el volumen muerto de la columna de 100 mL, a partir de entonces se recogieron fracciones de 5 mL. Dependiendo del peso de la proteína se tendrá un volumen de elución diferente donde se obtendrá el pico de absorbancia de la proteína de interés (**Figura 20**).

Una vez se obtuvo la proteína purificada y sin agregados se concentró utilizando amicones (**apartado 3.1.3.6**) y se comprobó mediante un gel SDS-PAGE. Finalmente se cambió el tampón en el propio amicon a un tampón sin NaCl (25 mM Tris pH 8) ya que no es recomendable tener NaCl presente para la generación de cristales.

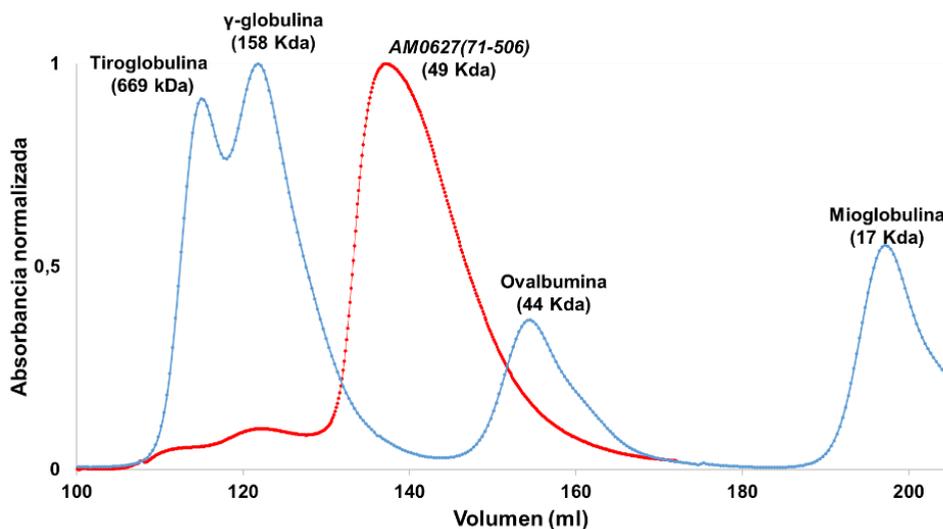


Figura 20. Filtración en gel de AM0627⁷¹⁻⁵⁰⁶ junto con un conjunto de proteínas con diferentes tamaños que se usan como marcadores de peso molecular.

3.1.4. Preparación y síntesis de los glicopéptidos empleados con AM0627

3.1.4.1. Síntesis de glicopéptidos en fase sólida (SPPS)

La secuencia peptídica de TEAQT**T**PPPA donde ** equivale a Gal β 1-3GalNAc, es común para todos los péptidos de este trabajo y está basada en un consenso de secuencias de corte de AM0627 determinada por espectrometría de masas (Shon et al., 2020).

A partir del glicopéptido TEAQT**T**PPPA (P1) se diseñaron el resto de péptidos de este trabajo (**Figura 21**) de la siguiente manera: eliminando los O-glicanos para dejar solo la secuencia peptídica (P9); añadiendo un grupo

GalNAc en la Thr5 y otro en la Thr6 (P2); añadiendo Gal β 1-3GalNAc en la Thr5 (P3); añadiendo Gal β 1-3GalNAc en la Thr6 (P4); añadiendo un grupo GalNAc en la Thr5 (P5); añadiendo un grupo GalNAc en la Thr6 (P6); añadiendo un grupo GalNAc en la Thr6 y un grupo galactosamina (GalN) en la Thr5 (P7); y añadiendo un grupo GalNAc en la Thr5 y un grupo GalN en la Thr6 (P8).

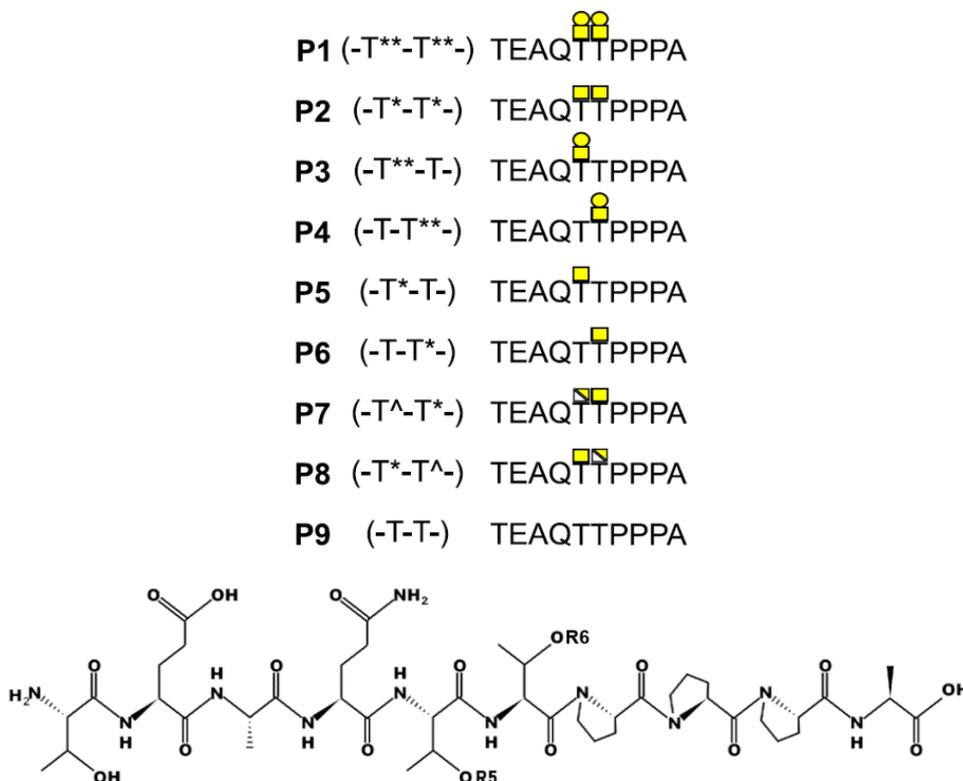


Figura 21. Representación de los glicopéptidos P1 a P9, así como su estructura química donde modificando los sustituyentes en R5 y R6 obtenemos los péptidos P1 a P9.

Los glicopéptidos con antígeno Tn utilizados en este trabajo se sintetizaron mediante síntesis en fase sólida asistida por microondas por etapas en un sintetizador Liberty Blue utilizando la estrategia Fmoc en resina Rink Amide MBHA (0,1 mmol).

Fmoc-Thr[GalNAc(Ac)₃- α -D]-OH (2,0 equivalentes) y/o Fmoc-Thr(GalN₃(Ac)₃- α -D)-OH (2,0 equivalentes) se sintetizaron como se describe en la literatura (Plattner et al., 2011) y se acoplaron manualmente usando HBTU [hexafluorofosfato de (2-(1H-benzotriazol-1-il)-1,1,3,3-tetrametiluronio)], mientras que todos los demás aminoácidos Fmoc (5,0 equivalentes) se acoplaron automáticamente usando oxyma pure/DIC (N,N'-diisopropilcarbodiimida). Los grupos O-acetilo del residuo GalNAc del antígeno Tn se eliminaron en una mezcla de NH₂NH₂/MeOH (7:3). En el caso de los péptidos P7 a P8, el grupo azido fue transformado en el correspondiente grupo amino por hidrogenación Pd/C. A continuación, los glicopéptidos se liberaron de la resina y todos los grupos protectores de las cadenas laterales sensibles a ácido se eliminaron simultáneamente utilizando TFA al 95%, TIS (triisopropilsilano) al 2,5% y H₂O al 2,5%, seguido de su precipitación con éter dietílico frío. Los productos crudos se purificaron por HPLC en una columna Phenomenex Luna C18(2) (10 μ m, 250·21,2 mm) y un detector de absorbancia dual, con un caudal de 10 mL/min.

La síntesis y purificación de los péptidos fue realizada por el grupo del Doctor Francisco Corzana.

Una vez purificados todos los péptidos, se disolvieron a 100 mM en tampón D tras lo que se comprobó el pH con tiras de pH y se ajustó entorno a pH 7-8 añadiendo 0,1-1 μL de NaOH 2 M.

3.1.4.2. Síntesis de glicopéptidos con antígenos T

Una vez se obtuvieron todos los glicopéptidos Tn preparados, se procedió a añadir una galactosa para generar el antígeno T o *Core 1*.

La enzima C1GalT1 lleva a cabo la transferencia de una galactosa de UDP-Galactosa al GalNAc invirtiendo la configuración en presencia de Mn^{2+} (**Figura 22**).

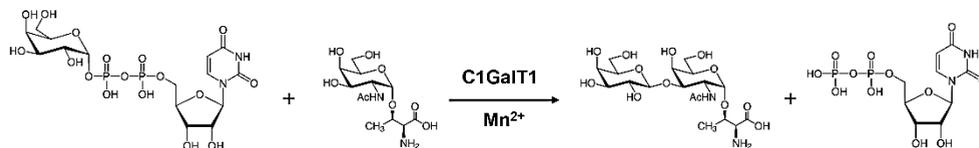


Figura 22. Reacción de C1GalT1.

Los péptidos P2, P5 y P6 fueron incubados o/n a 37 °C a una concentración de 23 mM junto con 51 μM de *DmC1GalT1*, 200 μM MnCl_2 y 97 mM UDP-Galactosa en tampón D a un volumen final de 180 μL .

Una vez se obtuvo los péptidos resultantes (P1, P3 y P4 respectivamente, **Figura 21**) se procedió a purificarlos como se explica en el apartado 3.1.4.1.

La purificación de los péptidos fue realizada por el grupo del Doctor Francisco Corzana.

3.1.5. Técnicas estructurales: Cristalografía de rayos X

3.1.5.1. Generación de cristales

El primer paso para resolver una estructura cristalográfica es la formación de cristales de un tamaño y calidad óptimos, lo que conlleva múltiples ensayos de cristalización ya que la más mínima diferencia entre dos proteínas puede dar lugar a condiciones de cristalización completamente diferentes.

Un experimento de cristalización requiere una solución de proteína concentrada que se mezcla con una disolución de agente precipitante, de manera que comience su precipitación de forma gradual y controlada. Esto permite la formación de núcleos cristalinos microscópicos que pueden crecer hasta un tamaño óptimo.

Para los cristales generados en este trabajo se utilizó la técnica de difusión de vapor en formato de gota sentada, una de las más utilizadas actualmente para la generación de cristales. En este método se mezcla la proteína de interés con la disolución de agente precipitante formando una gota de 0,6-1 μL y se aíslan junto con un reservorio de mayor volumen de la misma disolución de agente precipitante. Esto permite que se establezca un equilibrio entre la gota de mezcla proteína/agente precipitante y el reservorio, donde el desplazamiento de agua en forma de vapor se produce de la gota de proteína al reservorio, ya que ahí la concentración del agente precipitante es mayor. De esta manera la concentración de proteína y agente precipitante aumenta lenta y gradualmente, lo que representa la zona subsaturada en el diagrama de fase (**Figura 23**) hasta que comienza a formar núcleos cristalinos, lo que corresponde a la zona de nucleación (**Figura 23**). Con la formación de estos núcleos la concentración

de proteína soluble disminuye, llegando a la zona de crecimiento en el diagrama de fase (**Figura 23**) donde aumenta el tamaño del cristal. Si la concentración de proteína o de agente precipitante son demasiado altas, la proteína precipitara de manera no cristalina (**Figura 23**).

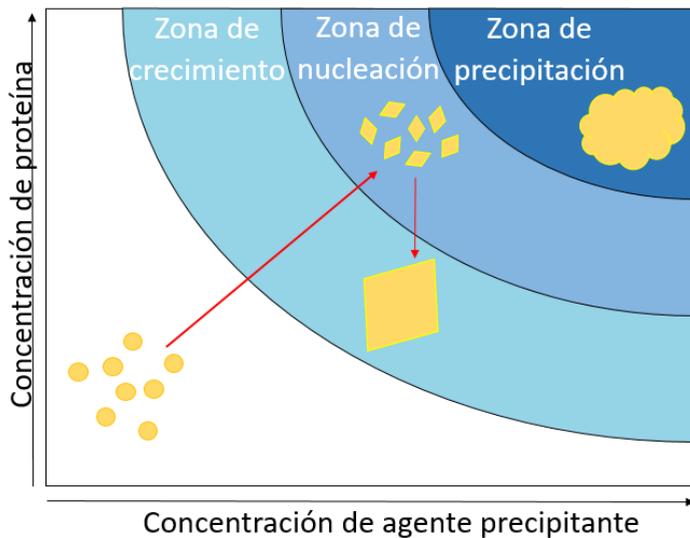


Figura 23. Diagrama de fases mostrando las diferentes zonas del equilibrio entre proteína y agente precipitante.

Todo esto lleva a la generación de cristales ordenados donde su estructura se describe según la geometría en la que se ordenan las partículas en la celdilla unidad. Esta es la unidad mínima de repetición que contiene toda la simetría del cristal y que se describe con la longitud de sus aristas (a , b , c) y los ángulos entre ellas (α , β , γ). La celdilla unidad se puede recrear si aplicamos los elementos de simetría al componente mínimo de la celdilla unidad que es la unidad asimétrica.

Una vez se obtuvieron los cristales fueron sumergidos en una disolución con un agente crioprotector, tal como glicerol o etilenglicol, que previene la formación de hielo cristalino. Este podría no solo dañar los cristales sino también dar su propio patrón de difracción enmascarando el patrón del propio cristal de proteína. Para congelarlos se utilizan crioloops (Hampton) sintéticos que permiten manipular el cristal, pero no interfieren en la difracción.

Para la cristalización de complejos proteína-ligando se pueden tomar dos posibles aproximaciones:

- La co-cristalización, en la que el ligando se pone directamente en la gota de proteína antes de los ensayos de tal manera que cristalice el complejo y no la proteína sola.
- El remojado o *soaking*, donde se sumergen los cristales en una disolución con una alta concentración de ligando. Esto permite que el ligando se desplace por los grandes canales de la estructura cristalina y se una a las proteínas que forman el cristal, pero tiene el riesgo de afectar a los cristales tanto como para disolverlos.

Por último, se puede tener problemas con la difracción del cristal y que la resolución no sea lo suficientemente buena, en tal caso se puede generar una nueva construcción eliminando zonas móviles, elegir otro ortólogo del que cristalizar su proteína o probar condiciones cercanas a la estudiada.

Los experimentos de co-cristalización realizados en esta tesis se llevaron a cabo en placas de poliestireno de 96 pocillos MRC 2 Well Crystallization Plate (Jena Bioscience) donde se mezclaron 0,5 μL de disolución de proteína (15

mg/mL AM0627⁷¹⁻⁵⁰⁶E326A en tampón 25 mM Tris pH 8, 0,5 mM ZnCl₂ y 5 mM P1) junto con 0,5 µL de disolución de agente precipitante frente a un reservorio de agente precipitante de 60 µL y fueron incubadas a 18 °C aisladas de vibraciones.

Los cribados que se emplearon fueron JCSG HTS +, Morpheus I, Morpheus II y ProPlex (Molecular Dimensions, Jena Bioscience). Finalmente, de las condiciones donde se obtuvieron cristales (**Figura 24**), los que tuvieron la mejor difracción se obtuvieron en las condiciones: 0,07 M de monosacáridos, 0,1 M del sistema tampón 1 pH 6,5 y 30% de mezcla precipitante 2 (Morpheus I, Molecular Dimensions).

Estos cristales se crioprotegieron remojándolos en una mezcla de la disolución de agente precipitante junto con 25% (v/v) de glicerol e inmediatamente se congelaron en nitrógeno líquido, reservándolos en un dewar hasta que fueron difractados.

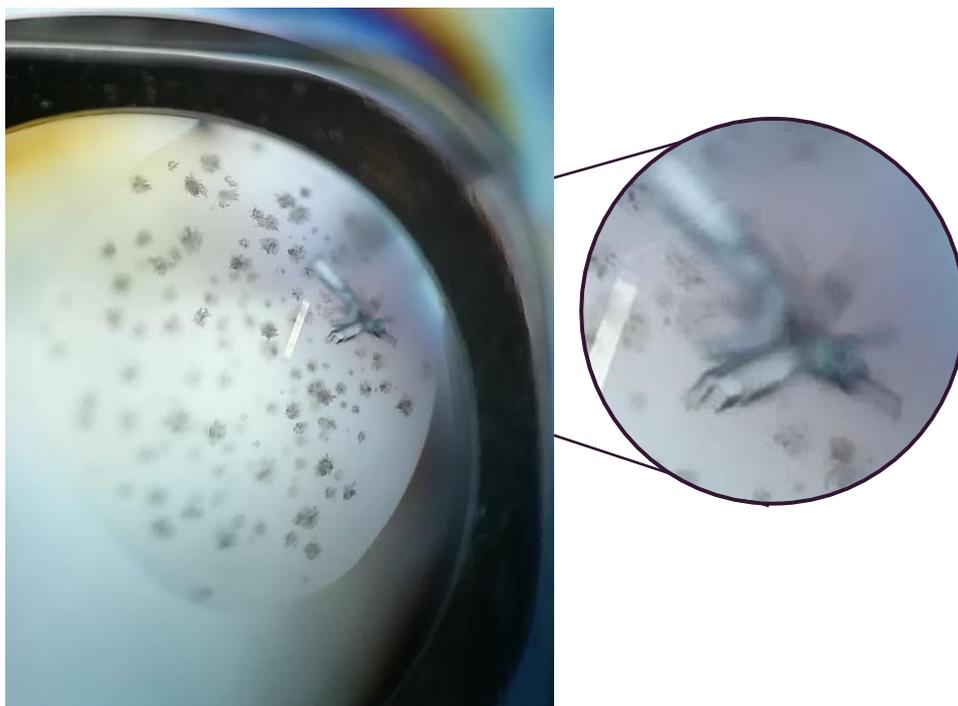


Figura 24. Cristales de AM0627 que se obtuvieron en las condiciones: 0,07 M de monosacáridos, 0,1 M de tampón 1 pH 6,5 y 30% de mezcla precipitante 2 (Morpheus I, Molecular Dimensions).

3.1.5.2. Difracción de rayos X y toma de datos

Una vez se obtuvieron los cristales congelados se mandaron a un sincrotrón, donde se les incidió con un haz de rayos X monocromático enfocado.

Cuando el haz incide sobre la estructura cristalina, los electrones contenidos en el cristal emiten radiación de rayos X en todas las direcciones. Los átomos en el cristal están colocados de manera periódica, lo que lleva a que en ciertas

direcciones se produzcan interferencias constructivas a ángulos concretos (**Figura 25**). Esto es lo que se conoce como el fenómeno de la difracción.

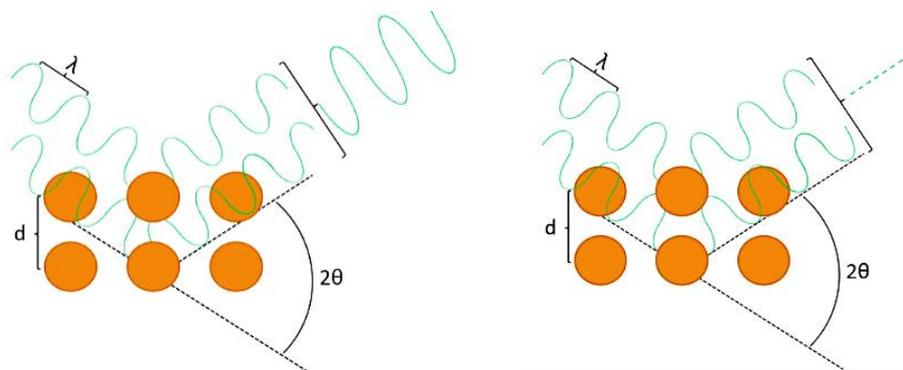


Figura 25. Izquierda. Representación del fenómeno de la difracción mostrando interferencias constructivas y **Derecha.** destructivas.

El máximo de intensidad de estos haces dispersados sigue la ley de Bragg (**Eq 1**) (Humphreys, 2013) produciendo un patrón de difracción que contiene la información sobre la disposición atómica dentro del cristal.

$$\text{Eq 1} \quad \textit{Bragg law: } n\lambda = 2d \sin\theta$$

Los cristales de proteína necesitan una luz muy intensa para conseguir una señal razonable del patrón de difracción por lo que hay que mandarlos a un sincrotrón, que da alta resolución y luminosidad dado a la intensidad del haz.

Los cristales se montan en una cabeza goniométrica durante el experimento para facilitar su rotación ya que esta es necesaria para observar todos los planos cristalográficos de la muestra.

Durante el experimento, se utiliza una corriente de nitrógeno líquido, ya que a bajas temperaturas los cristales macromoleculares resisten mejor los efectos de la radiación de rayos X y permiten obtener mejores datos de difracción. Además, facilita la ubicación de los átomos en la estructura cristalina, ya que la baja temperatura minimiza la vibración térmica atómica. Un experimento típico de rayos X de proteínas consta de miles de patrones de difracción que se registran en un detector de rayos X.

Los cristales se mandaron al sincrotrón ALBA en Barcelona, España (beamline 13, XALOC). donde se tomaron los datos usando una longitud de onda de 0,97 Å y 1,28 Å a una temperatura de 100 K.

3.1.5.3. Procesamiento y refinamiento de la estructura cristalográfica

Una vez se tuvieron los datos de difracción se tuvo que resolver la función de densidad electrónica (**Eq 2**) que representa la transformada de Fourier entre las posiciones atómicas (espacio real) y el patrón de difracción (espacio recíproco).

En el experimento se obtienen los factores de estructura ($|F_{\text{obs}}(\text{hkl})|$), pero se pierden la información relativa a las fases ($\phi_{\text{obs}}(\text{hkl})$)

$$\text{Eq 2} \quad \rho(\text{xyz}) = \frac{1}{V} \sum_{-\infty}^{+\infty} |F(\text{hkl})| \cdot e^{-2\pi i[\text{hx}+\text{ky}+\text{lz}-\phi(\text{hkl})]}$$

Esto se conoce como el problema de las fases que en el caso de cristales de proteína se puede resolver mediante tres estrategias: dispersión anómala de

longitud de onda múltiple o simple (MAD/SAD), remplazo isomorfo múltiple (MIR) y remplazo molecular (MR).

Si existe un modelo de una proteína homologa a la de interés se puede resolver el problema de las fases mediante MR, donde se utiliza la proteína modelo para obtener unas fases teóricas con las que empezar el refinamiento lo que nos da un primer mapa de densidad electrónica. Mediante refinamientos manuales y automáticos con herramientas computacionales, los aminoácidos de la proteína se encajan en la densidad electrónica lo mejor posible siguiendo dos parámetros, Rfactor y Rfree (Wlodawer et al., 2008), que están relacionados con la convergencia entre los datos experimentales y teóricos.

Cuando la proteína ha encajado en la densidad, no hay densidad electrónica sin explicar y los parámetros Rfactor y Rfree han convergido se obtiene el modelo final.

Los datos fueron procesados usando los paquetes de software XDS (Kabsch, 2010) y CCP4 (Winn et al., 2011).

Se aplicó SAD utilizando SHELXE (Usón y Sheldrick, 2018) al cristal recogido a una longitud de onda de 1,28 Å, y se encontraron dos sitios de zinc, uno para cada monómero, usando la señal anómala presente hasta una resolución de 1,9 Å. El coeficiente de correlación fue 43,0% para todos los datos y 29,2% para las reflexiones débiles.

Después se usó SHELXE para distinguir la quiralidad y para revelar los átomos de la proteína usando trazado de polialanina tanto para hélices como láminas.

De esta solución, con 3 ciclos de autotrazado y con 10 ciclos de modificación de la densidad, SHELXE distinguió la quiralidad correcta y fue capaz de encontrar 845 aminoácidos divididos en 10 cadenas (correspondientes a las dos moléculas en la unidad asimétrica). El coeficiente de correlación fue de 41,85%.

Después se resolvió la estructura cristalina del complejo AM0627⁷¹⁻⁵⁰⁶E326A-P1-Zn²⁺ recogido a 0,97 Å por remplazo molecular con Phaser (McCoy et al., 2007) usando el modelo de SHELXE.

Las fases iniciales se mejoraron con ciclos de construcción de modelo manuales con Coot (Emsley y Cowtan, 2004) y refinamiento automático con REFMAC5 (Winn et al., 2011) hasta obtener el modelo final.

El modelo de AM0627⁷¹⁻⁵⁰⁶E326A se validó (**Tabla 2**) con PROCHECK (Winn et al., 2011).

Las gráficas de Ramachandran para AM0627⁷¹⁻⁵⁰⁶E326A mostraron que 90,6%, 8,6%, 0,3%, y 0,5% de los aminoácidos estaban en regiones muy favorables, permitidas, generosamente permitidas y no permitidas respectivamente.

Tabla 2. Recopilación de datos y estadísticas del refinamiento de AM0627⁷¹⁻
506E326A en complejo con P1 y Zn²⁺.

	AM0627 ⁷¹⁻⁵⁰⁶ E326A en complejo con P1 y Zn ²⁺	AM0627 ⁷¹⁻⁵⁰⁶ E326A en complejo con P1 y Zn ²⁺
Toma de datos		
Grupo espacial	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁
Longitud de onda de emisión (Å)	1,2817	0,9792
Dimensiones de la celdilla unidad <ul style="list-style-type: none"> • a, b, c (Å) • α, β, γ (°) 	74,05/ 82,22/ 173,81 90/ 90/ 90	73,14/ 82,69/ 172,61 90/ 90/ 90
Numero de moléculas en la unidad asimétrica	2	2
Resolución (Å) Los valores entre paréntesis se refieren a la capa de mayor resolución	173,81-1,90 (2,00-1,90)	172,61-1,50 (1,58-1,50)
Rmerge	0,153 (2,140)	0,058 (1,092)
I / σ I	13,9 (2,5)	12,6 (1,6)
Compleitud (%)	100 (100)	100 (100)
Redundancia	24,8 (22)	7,3 (7,2)
Mn(I) CC(1/2)	0,999 (0,781)	0,999 (0,624)
Refinamiento		
Resolución (Å)	1,90	1,50
Número total de reflexiones observadas	2091294	1229955

Número total de reflexiones únicas	84443	167774
<i>R</i> work / <i>R</i> free		0,167/0,1821
Número de átomos		
Proteína		7048
P1		211
Zn ²⁺		2
Aguas		769
Glicerol		72
Factores B (Å²)		
Proteína		28,74
P1		52,30
Zn ²⁺		20,85
Aguas		39,87
Glicerol		55,63
Desviación R.M.S.		
Longitud de enlace (Å)		0,0083
Ángulos de enlace (°)		1,5532

3.1.6 Ensayos de proteólisis

3.1.6.1. Ensayos de proteólisis de los glicopéptidos *in vitro*

La actividad de corte glicopeptidasa se midió *in vitro* por MS MALDI-TOF de manera semicuantitativa. La reacción de la enzima silvestre o *wildtype* (WT) AM0627²¹⁻⁵⁰⁶ y todos sus mutantes se llevaron a cabo añadiendo 0,4 μM de enzima junto con 57 μM de P1 en tampón 50 mM de hidrógeno carbonato de amonio a pH 8 e incubando a 37 °C.

Los experimentos de corte con los distintos glicopéptidos P1 a P9 se llevaron a cabo añadiendo 0,4 μM de enzima AM0627²¹⁻⁵⁰⁶ WT junto con 500 μM de cada glicopéptido en tampón 50 mM de hidrógeno carbonato de amonio a pH 8 e incubando a 37 °C.

En cada caso se cogieron alícuotas de las mezclas de reacción a distintos intervalos de tiempos y la generación de producto se detectó utilizando MS MALDI-TOF.

Los ensayos de corte fueron realizados por el grupo del Doctor Henrik Clausen.

3.1.6.2. Ensayos de proteólisis con *reporters*

El diseño y la construcción de los *reporters* se realizó como se describe anteriormente en la literatura (Nason et al., 2021), así como la purificación de BT4244 (**apartado 4.1.8.**) (Konstantinidi et al., 2022a).

Se utilizó un *reporter* que contiene 18 repeticiones de 12 aminoácidos conteniendo bis O-glicositos (AEAAAT*T*PAPAK)_{n=18}, producido

homogéneamente en células HEK293 modificada para producir antígeno Tn (*knockout* (KO)^{C1GALT1}).

Se utilizó un *reporter* de los TR de MUC1 que contiene entre 5 y 6 repeticiones de TRs de 20 aminoácidos con secuencia GVT SAPDTRPAPGSTAPPAH que contiene cinco O-glicositos, repartidos en dos bis O-glicositos y un mono O-glicosito. Este se expresó utilizando líneas estables tanto para células HEK293 WT, así como HEK293 con la O-glicosilación modificada para producir antígeno Tn (KO^{C1GALT1}), *Core 1* (KO^{GCNT1, ST3GAL1/2, ST6GALNAC2/3/4}) o mono-sialil T (mSTa) (KO^{GCNT1, ST6GALNACT2/3/4}).

Todas las líneas celulares de HEK293 se crecieron a una densidad de $0,25 \cdot 10^6$ células/mL y fueron incubadas por 5 días en un agitador orbital con medio Freestyle 17 (Gibco) suplementado con 0,1%(w/v) Kolliphor P188 (Sigma-Aldrich) y 2%(v/v) Glutamax (Thermofisher)

Dado que los *reporters* fueron diseñados con una cola de 6 histidinas, el sobrenadante se purificó mediante cromatografía de afinidad Ni-NTA (Qiagen) utilizada igual que una HisTrap como se explica en el **apartado 3.1.3.2.** utilizando como tampón A' 25 mM fosfato de sodio, 0,5 M NaCl y 10 mM imidazol pH 7,4, y como tampón B' 25 mM fosfato de sodio, 0,5 M NaCl y 200 mM imidazol pH 7,4.

Una vez se purificaron los *reporters* se les cambió el tampón por H₂O MQ usando columnas Zeba Spin (Thermofisher) y se cuantificaron usando el Kit Pierce™ BCA Protein Assay Kit (ThermoFisher).

Los ensayos de corte con los diferentes *reporters* purificados se realizaron incubando diluciones seriadas de AM0627²¹⁻⁵⁰⁶, AM0627⁷¹⁻⁵⁰⁶ y BT4244 que

se mezclaron con 0,5 μg de cada *reporter* durante 2 o 3 horas incubando a 37 °C en tampón 50 mM de hidrógeno carbonato de amonio a pH 8.

Las reacciones se pararon desnaturalizando las enzimas a 95 °C durante 5 minutos. Las muestras se cargaron en geles NuPAGE Novex (Bis-Tris gradiente 4-12%) y se les aplicó un voltaje de 200 V durante 45 minutos. Después, se tiñeron usando la tinción de proteínas fluorescentes Krypton (ThermoFischer) y se observaron con un sistema ImageQuant LAS4000 (GE Healthcare).

Los ensayos de corte fueron realizados por el grupo del Doctor Henrik Clausen.

3.1.7. Técnicas computacionales

3.1.7.1. Simulaciones de dinámica molecular de AM0627⁷¹⁻⁵⁰⁶ y BT4224

Se usó la estructura de AM0627⁷¹⁻⁵⁰⁶E326A y la de BT4224 (PDB: 5KD8) como estructuras de partida para todas las simulaciones en este trabajo. Primero, se mutó la base catalítica eliminada en AM0627⁷¹⁻⁵⁰⁶E326A utilizando Chimera (Pettersen et al., 2004). La protonación de las histidinas se eligió basada en la red de puentes de hidrógeno y coordinación del metal comprobada manualmente con Chimera. Los residuos His325 y His329 que coordinan al Zn^{2+} se introdujeron como protonados en N δ y el resto de residuos de histidina se introdujeron como protonados en N ϵ .

Las simulaciones se llevaron a cabo en pH 7, por lo que los residuos de ácido aspártico y ácido glutámico estaban cargados negativamente mientras que todas las argininas y lisinas estaban cargadas positivamente. El sistema a

estudiar se introdujo en el centro de una caja de $98 \cdot 98 \cdot 98 \text{ \AA}^3$ con una distancia de al menos 10 \AA entre la superficie del solvente y el borde de la caja.

La caja se solvató luego con moléculas de agua TIP3P y contraiones que se añadieron para neutralizar el sistema. Se describió la proteína usando el campo de fuerza AMBER20 ff14SB (Maier et al., 2015) mientras que el campo de fuerza GLYCAM06j (Kirschner et al., 2008) se usó para describir las moléculas de carbohidratos de los sustratos P1, P2 y P9. El módulo LEaP de AMBER20 se usó para generar los archivos de topología y coordenadas para simulaciones de dinámica molecular (MD) clásicas que se llevaron a cabo usando la versión de CUDA del módulo PMEMD (Salomon-Ferrer et al., 2013) del paquete de simulación AMBER20.

Después, el sistema solvatado se sometió inicialmente a 5000 pasos de minimización del algoritmo de descenso de gradiente, seguido de 5000 pasos de minimización del algoritmo de gradiente conjugado con restricciones de posición en todos los átomos pesados del solvente, utilizando un potencial armónico de $50 \text{ kcal} \cdot \text{mol}^{-1} \cdot \text{\AA}^{-2}$.

El sistema minimizado se calentó después hasta 300 K usando el termostato de Berendsen, con una constante de tiempo de 1 ps para el acoplamiento y restricciones de posición de $50 \text{ kcal} \cdot \text{mol}^{-1} \cdot \text{\AA}^{-2}$ aplicadas en tres pasos de 500 ps del proceso de calentamiento. Las restricciones de posición se disminuyeron gradualmente a $5 \text{ kcal} \cdot \text{mol}^{-1} \cdot \text{\AA}^{-2}$ en cuatro pasos de 500 ps de equilibrado NPT, usando el termostato y baróstato de Berendsen para mantener el sistema a 300 K y 1 atm de manera constante.

Para las simulaciones finales, cada sistema se sometió a 200 o 400 ns de muestreo a temperatura y presión constante de 300 K y 1 atm, controlados por el termostato de Langevin, con una frecuencia de colisión de $2,0 \text{ ps}^{-1}$, y el baróstato de Berendsen con una constante de acoplamiento de 1,0 ps. El algoritmo SHAKE se aplicó para contener todos los enlaces que implicaran átomos de hidrógeno. Se aplicó un límite de 10 \AA a todas las interacciones no enlazantes, con las interacciones electrostáticas lejanas tratadas con el método de partícula-malla de Ewald (PME).

Se usaron pasos de 2 fs para todas las simulaciones clásicas, y las coordenadas de la simulación se guardaron cada 10 ps. Se llevaron a cabo tres simulaciones independientes.

Las simulaciones fueron realizadas por los grupos de la Doctora Carme Rovira y del Doctor Francisco Corzana.

3.1.7.2. Simulaciones QM/MM del complejo AM0627⁷¹⁻⁵⁰⁶-P1-Zn²⁺

Se usó un fotograma extraído de la trayectoria de la MD clásica para las posteriores simulaciones de mecánica cuántica (QM)/mecánica molecular (MM) del complejo AM0627⁷¹⁻⁵⁰⁶-P1-Zn²⁺, que combina simulaciones Born-Oppenheimer, basadas en la teoría del funcional de la densidad (DFT), con metodología de campo de fuerza. La región QM consistió en el ion de Zn²⁺ y los residuos con los que está coordinado (His325, His326, y Glu343) así como partes del glicopéptido sustrato (Thr5, Thr6, y Pro7), el agua catalítica y el Glu326, resultando en un total de 76 átomos. Los enlaces colocados entre la región QM y MM se truncaron por átomos de hidrógeno. La región QM se

contuvo en una supercélula aislada de 20,0·20,0·20,0 Å³. Todas las simulaciones de QM/MM se llevaron a cabo usando CP2K v7.1 interconectado con PLUMED v2.5 combinando el programa para QM QUICKSTEP y el driver MD FIST.

Se uso el método *multigrid* de espacio real para simular el acoplamiento electrostático entre las regiones QM y MM. La región QM se trató con DFT(BLYP), usando el set dual de gaussiana y ondas planas (GPW) mientras que el resto del sistema se simuló al nivel clásico utilizando los parámetros usados para las simulaciones de MD del **apartado 3.1.7.1**. Se usó el set de gaussiana de valencia triple Z polarizada (TZV2P) para expandir la función de onda, mientras que el set de ondas planas auxiliar con un límite de densidad de 350 Ry y pseudopotenciales GTH se utilizó para converger la densidad electrónica. Todas las simulaciones QM/MM se llevaron a cabo bajo el conjunto NVT usando una constante de acoplamiento de 10 fs y un tiempo de integración de 1,0 fs. Primero el sistema se equilibró sin ninguna restricción por 10 ps. Después se usó el método de las metadinámicas para explorar el perfil de energía libre para cada paso de la reacción.

La distancia entre el oxígeno del agua y el carbono del grupo carbonilo de la Thr5 (C*) se usó como coordenada de reacción o *collective variable* en inglés (CV) para el primer paso de reacción (CV1), mientras que la distancia entre el carbono del grupo carbonilo de la Thr5 (C*) y el nitrógeno amida de la Thr6 se usó como coordenada de reacción del segundo paso de la reacción (CV2). La transferencia de protón ocurrió espontáneamente durante las simulaciones de metadinámicas así que no se añadió ninguna CV para activarlas.

La altura de la campana de gauss se puso a 1,0 kcal/mol, y se cambió a 0,1 kcal/mol cuando estaba a punto de cruzar el estado de transición y el intervalo de tiempo de deposición entre dos gaussianas consecutivas se programó a 25s. Las anchuras de las campanas de gauss se retocaron de acuerdo a las oscilaciones de cada coordenada de reacción (0,2 Å para CV1 y CV2). Se observó un equilibrio en el primer estado de transición en el primer paso de reacción, pero no en el segundo paso.

En el segundo paso, la transferencia del protón de la Thr5 al grupo amida de la Thr6 fue espontánea, pero no volvió al estado intermedio. Por esta razón la energía libre del estado P es posible que este sobreestimada, mientras que el mecanismo de reacción no se encuentra afectado. Las trayectorias obtenidas en la simulación se analizaron con VMD y PyMOL (PyMOL 2.4.2), y los cálculos de distancia y análisis de clústeres se hicieron con CPPTRAJ de AMBER 20.

Las simulaciones fueron realizadas por el grupo de la Doctora Carme Rovira.

CAPÍTULO 4. RESULTADOS

4.1. AM0627

4.1.1. Características generales de la estructura del complejo AM0627⁷¹⁻⁵⁰⁶E326A-P1-Zn²⁺

Para entender el reconocimiento de los glicodominios y las preferencias en cuanto a sustratos de AM0627, se propuso obtener la estructura de rayos X con un bis T O-glicopéptido basado en la secuencia consenso de corte para AM0627, TEAQT**T**PPPA (P1). Las construcciones que se probaron de AM0627 fueron mutadas para cambiar el residuo Glu326 a alanina (E326A) para inactivar la enzima (Shon et al., 2020). La primera construcción probada, contenía los residuos Ala21 a Glu506. Se obtuvieron cristales con una difracción pobre, impidiendo la resolución de la estructura, así que se generó un nuevo constructo que contenía los residuos Pro71 hasta Glu506 (AM0627⁷¹⁻⁵⁰⁶E326A), ya que se planteó la hipótesis que los 50 primeros residuos, entre Ala21 y Lys70, no eran esenciales para la enzima ni estaban cerca del centro activo (**Figura 31**).

Para corroborar que la actividad de la construcción truncada AM0627⁷¹⁻⁵⁰⁶ mantenía niveles similares a la enzima WT, se probó su actividad utilizando un *reporter* que contiene 18 repeticiones de 12 aminoácidos conteniendo bis O-glicositos (AEAAAT*T*PAPAK)_{n=18}, producido homogéneamente en

células HEK293 con un KO de *CIGALT1* para garantizar que la O-glicosilación se detiene en un Tn O-glicano (Nason et al., 2021).

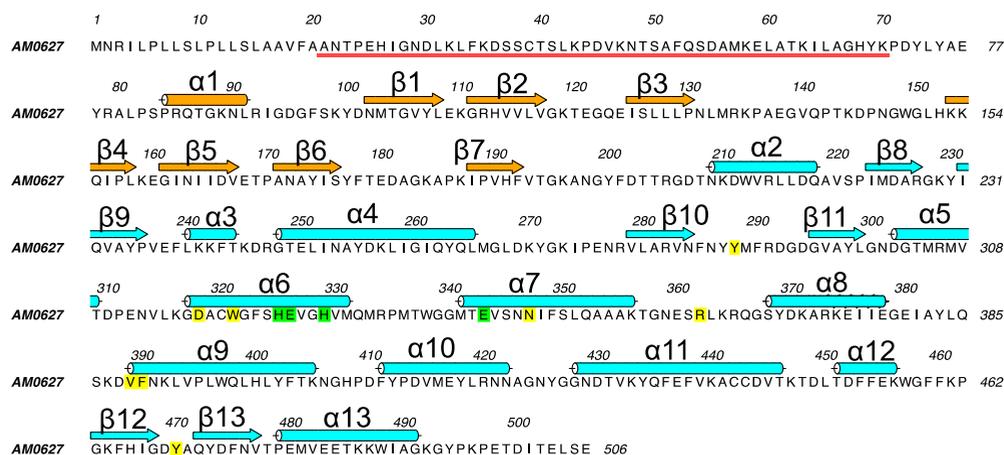


Figura 26. Secuencia y estructura secundaria (α -hélices y láminas β) de AM0627 obtenida de la estructura cristalográfica donde se puede observar en naranja el dominio tipo inmunoglobulina, en cian el dominio M60 catalítico y la parte eliminada en AM0627⁷¹⁻⁵⁰⁶ E326A subrayada en rojo. Los residuos His325, Glu326 e His329 que forman el motivo HEXXH, junto con Glu343, están resaltados en verde.

La proteína AM0627²¹⁻⁵⁰⁶ y la forma truncada AM0627⁷¹⁻⁵⁰⁶ mostraron ser cualitativamente similares a la hora de hidrolizar el *reporter* (**Figura 32**), demostrando así que la proteína truncada es comparable a la *wildtype* y por tanto que la construcción cristalizada es válida para entender la enzima *wildtype* al menos al nivel del dominio catalítico.

Una vez eliminados estos residuos, el complejo AM0627⁷¹⁻⁵⁰⁶ E326A-P1-Zn²⁺ cristalizó generando cristales ortorrómbicos del grupo espacial P2₁2₁2₁ con

mejor difracción que el complejo con la construcción larga AM0627²¹⁻⁵⁰⁶E326A.

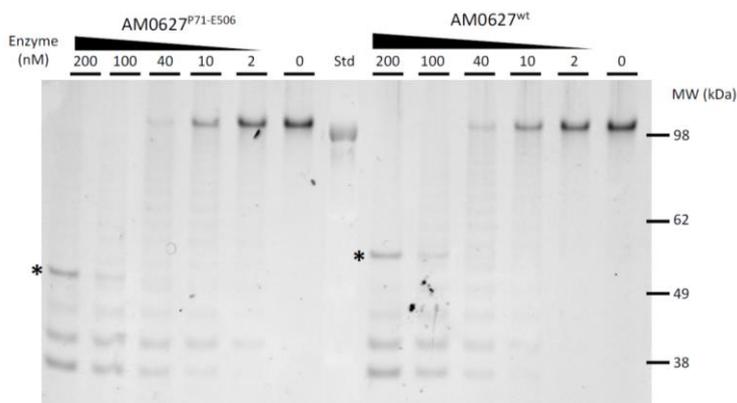


Figura 27. Análisis SDS-PAGE de la actividad de la enzima *wildtype* AM0627²¹⁻⁵⁰⁶ y el truncado AM0627⁷¹⁻⁵⁰⁶ con un *reporter* artificial con bis Tn O-glicanos. Se incubaron 0,5 µg de *reporter* de bis Tn O-glicano purificado durante 2 horas a 37 °C con la concentración de enzima variando de 0 a 200 nM. Los geles se visualizaron con tinción de proteína fluorescente Krypton. * indica las bandas correspondientes a la enzima AM0627²¹⁻⁵⁰⁶ y el truncado AM0627⁷¹⁻⁵⁰⁶. El carril central marcado como std contiene los marcadores de peso molecular. Además, el peso molecular de cada proteína en el patrón se indica en el lado derecho. Los experimentos se realizaron por duplicado.

La presencia de Zn²⁺ en los cristales permitió usar Zn-SAD para resolver la estructura a 1,9 Å y obtener un modelo usando SHELXE. Esto permitió resolver la estructura partiendo de otro cristal con el mismo grupo espacial a 1,5 Å por reemplazo molecular. Aunque se observaron dos moléculas de AM0627 en la unidad asimétrica que tenían ciertas interacciones, se puede descartar que formen un dímero en disolución ya que el cromatograma de la

columna de exclusión molecular muestra que se encuentra como un monómero en disolución (**Figura 20, apartado 3.1.3.7.**).

Se muestran en la estructura del complejo AM0627⁷¹⁻⁵⁰⁶E326A-P1-Zn²⁺ dos dominios diferenciados de la enzima AM0627, un dominio tipo inmunoglobulina encontrado previamente en otras estructuras cristalinas de mucinasas como BT4244, ZmpB y IMPa, y un dominio catalítico M60 típico de las metalopeptidasas dependientes de Zn²⁺ (**Figura 33**) compartido también con mucinasas como BT4244, ZmpB y AM0908 (Noach et al., 2017a; Rawlings et al., 2018).

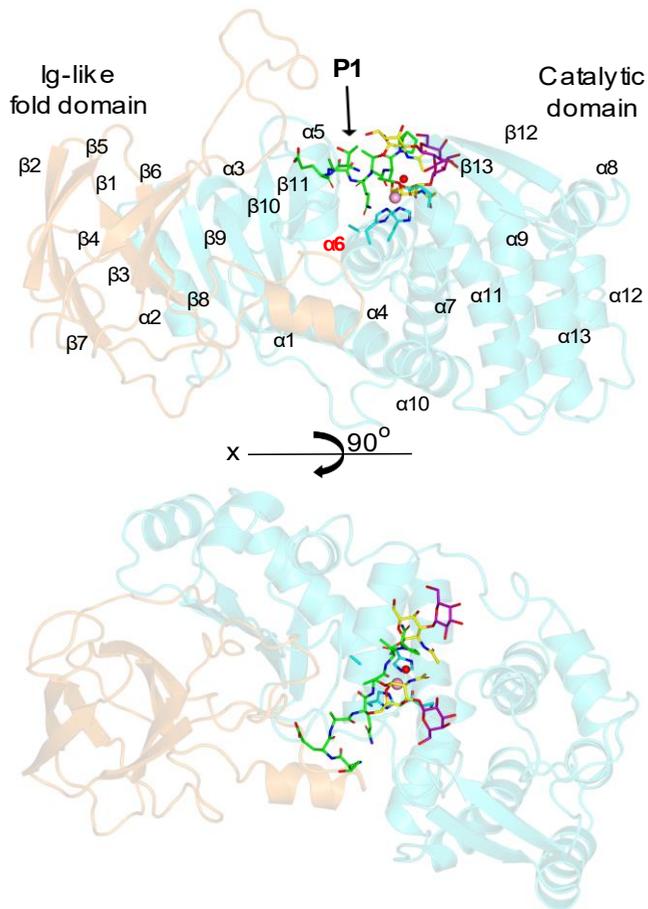


Figura 28. Estructura cristalina del complejo AM0627⁷¹⁻⁵⁰⁶E326A-P1-Zn²⁺ en formato *cartoon*. En el panel inferior se observa el mismo monómero con un giro de 90 grados a través del eje X. Los dominios se representan en naranja y cian para el dominio tipo inmunoglobulina y el dominio M60 catalítico, respectivamente. Las estructuras secundarias se indican en la estructura superior con $\alpha 6$ resaltado en rojo. Los aminoácidos de P1 se representan con átomos de carbono verdes, mientras que el GalNAc y Gal se muestran como átomos de carbono amarillo y magenta, respectivamente. El Zn²⁺ y una molécula de agua se muestran como una esfera rosa y roja respectivamente.

Comparando ambas cadenas A y B de ambos monómeros en la unidad asimétrica se observó que la raíz del error cuadrático medio (RMSD por sus siglas en inglés) entre ellas para 437 átomos de carbono alfa ($C\alpha$) fue de 0,44 Å, demostrando que no se presentaban diferencias significativas entre ambas cadenas. Por este motivo se centró el análisis de una de ellas y en particular de la cadena B de la estructura de AM0627⁷¹⁻⁵⁰⁶E326A ya que presentaba el mapa de densidad electrónica para el O-glicopéptido mejor definido.

El dominio catalítico M60 de AM0627 mostraba el motivo típico HEXXH, localizado en la hélice alfa conservada $\alpha 6$ (**Figura 33**). Los residuos que componían este motivo eran His325, Glu326 en esta estructura mutado a Ala326 y His329 (**Figura 34**). A parte, se encuentra un Glu343 conservado, que junto con His325 y His329 del motivo HEXXH y el grupo carbonilo de la Thr5 en P1 forman una coordinación con geometría pentagonal para el Zn^{2+} (**Figura 34 y Tabla 6**), característica que concuerda con el clan de las gluzincin al que pertenece AM0627.

En el sitio activo se puede encontrar una molécula de agua interaccionando mediante puentes de hidrógeno con el carbonilo de la Thr5, el grupo acetamido del GalNAc localizado en G1, según la nomenclatura utilizada para definir la red de glicanos (**Figura 34**), y el grupo acetamido de la Thr6 sugiriendo su posible implicación en la catálisis (**Figura 34**).

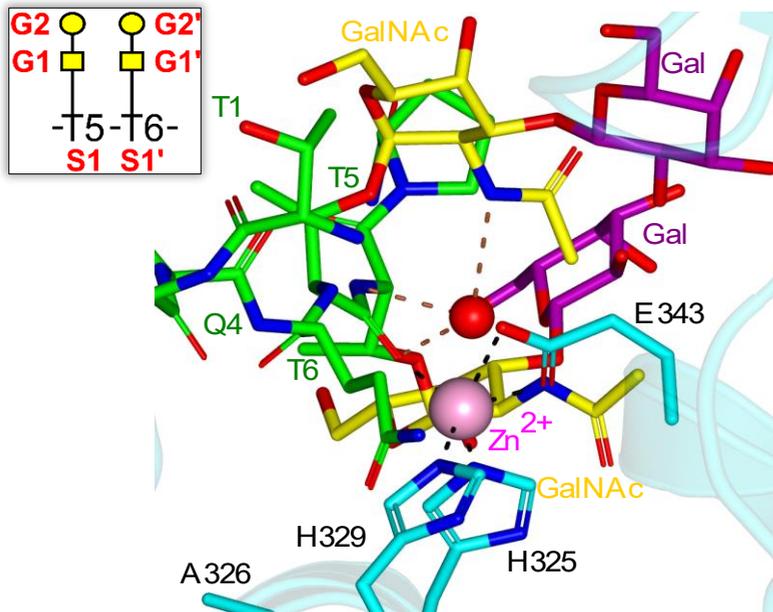


Figura 29. Vista en detalle del sitio activo donde se observa la geometría pentagonal formada por el Zn^{2+} y los residuos de coordinación (His325, His329, Glu343 y Thr5). También se observan las interacciones de la molécula de agua con la Thr5, Thr6 y el GalNAc. En líneas discontinuas negras se muestran las interacciones entre el metal y los aminoácidos y en línea de puntos marrón los puentes de hidrógeno entre la molécula de agua y P1. En el recuadro se muestra un esquema del subsitio y las moléculas de glicanos se muestran en amarillo y con diferentes símbolos según al SNFG (Neelamegham et al., 2019).

Tabla 3. Resumen de las interacciones entre el Zn^{2+} y la molécula de agua con los residuos de AM0627 y P1.

Zn^{2+}/H_2O	Aminoácido (AM0627/P1)	Tipo de interacción
Zn^{2+}	His325 ^{AM0627} (cadena lateral)	Esfera de coordinación del Zn^{2+}
Zn^{2+}	His329 ^{AM0627} (cadena lateral)	Esfera de coordinación del Zn^{2+}
Zn^{2+}	Glu343 ^{AM0627} (cadena lateral)	Esfera de coordinación del Zn^{2+}
Zn^{2+}	Thr5 ^{P1} (<i>backbone</i>)	Esfera de coordinación del Zn^{2+}
H ₂ O	Thr5 ^{P1} (grupo carbonilo)	Puente de hidrógeno
H ₂ O	GalNAc ^{G1} (grupo acetamido)	Puente de hidrógeno
H ₂ O	Thr6 ^{P1} (grupo acetamido)	Puente de hidrógeno

4.1.2. Detalles de la interacción entre AM0627 y el bis T O-glicopéptido P1

El sitio activo está formado por el sitio de unión a zinc ya mencionado y el sitio de unión al bis T-glicopéptido P1 (**Figura 35**).

Observando las interacciones entre P1 y la enzima se puede ver que la cadena lateral del Glu2 mantiene un puente de hidrógeno con la cadena lateral de Asp292, el grupo metilo de Ala3 establece interacciones de CH- π con las cadenas laterales de Trp149/Phe290, y la cadena lateral de Gln4 forma puentes de hidrógeno con el *backbone* de Arg291. Respecto a las dos treoninas O-glicosiladas, el *backbone* de Thr5 establece interacciones de puente de hidrógeno con la cadena lateral de Tyr470, Thr5 también mantiene interacciones de CH- π con la cadena lateral de Tyr287 a través del grupo metilo, y el *backbone* de la Thr6 forma interacciones de puente de hidrógeno con la cadena lateral de Tyr287 (**Figura 35**).

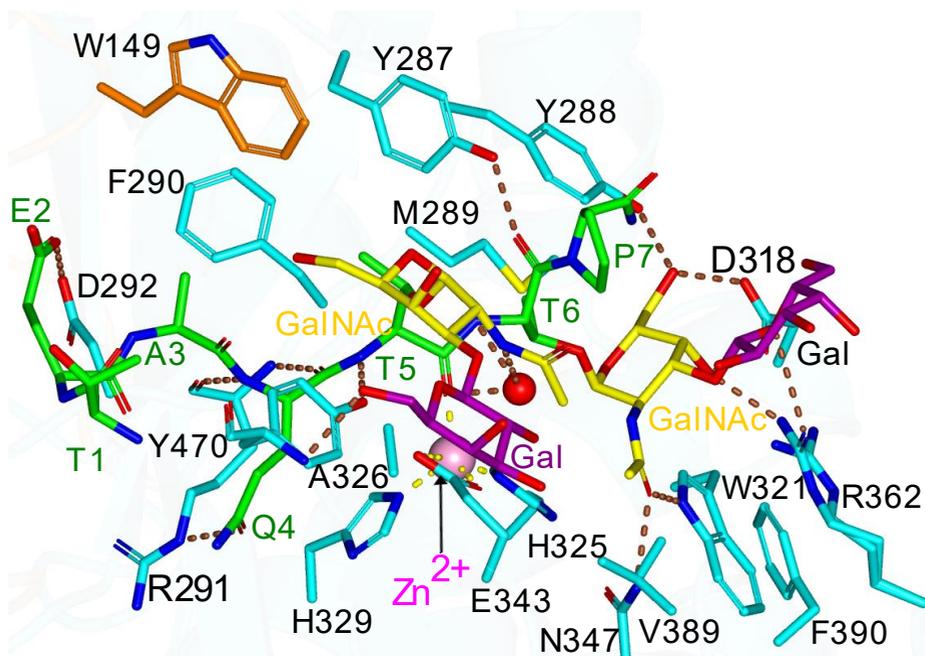


Figura 30. Detalle de las interacciones en el sitio activo del complejo AM062771-506E326A-P1-Zn²⁺. En línea de puntos amarilla las interacciones entre el metal y los aminoácidos y en línea de puntos marrón los puentes de hidrógeno entre la molécula de agua y P1, y entre AM062771-506E326A y P1.

Se observa en las interacciones que implican a los O-glicanos de P1, que el GalNAc y Gal de G1 y G2 respectivamente, establecen interacciones de CH- π con Tyr470 (**Figura 36**). Además, el grupo OH6 del Gal en G2 establece un puente de hidrógeno con el *backbone* de Tyr470 (**Figura 36**).

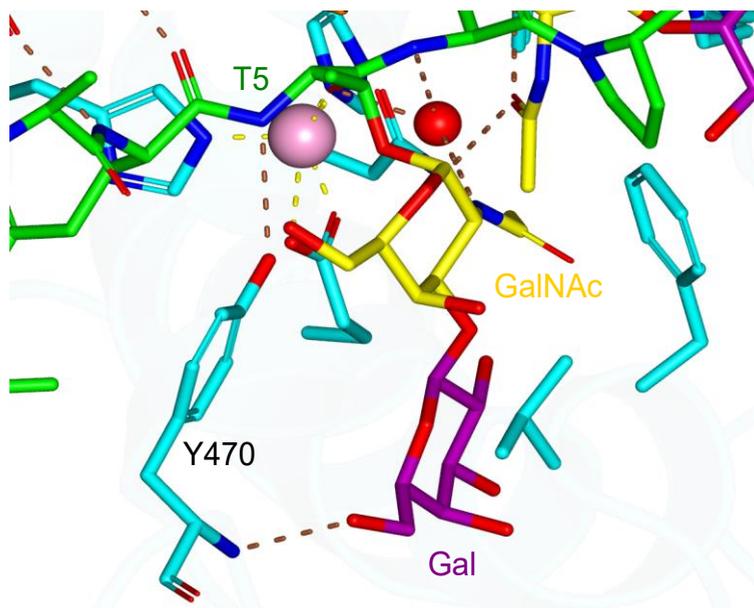


Figura 31. Detalle de las interacciones de Tyr470 con el glicopéptido. Se muestran las interacciones entre Tyr470 con el GalNAc y Gal en G1 y G2 respectivamente.

Cabe destacar que, de todos los azúcares, el GalNAc localizado en G1' es el que establece el mayor número de interacciones con AM0627 (**Figura 35 y Tabla 7**). El grupo carbonilo del acetamido y el grupo metilo del GalNAc establecen puentes de hidrógeno con las cadenas laterales de Trp321 y Asn347 respectivamente, así como interacciones de CH- π con Phe390 (**Figura 35**). También se ven puentes de hidrógeno entre los grupos OH3 y OH6 del GalNAc con las cadenas laterales de Asp318 y Arg362 por parte del OH3, y Tyr288 y Asp318 por parte del OH6 (**Figura 35**). Por último, el Gal en la subunidad G2' está poco reconocido y expuesto al solvente, estableciendo únicamente un

puede de hidrógeno con la cadena lateral de Arg362 a través de su oxígeno cíclico (**Figura 35**).

Todas estas interacciones revelan que tanto la cadena peptídica como la mayoría de los O-glicanos de P1 se muestran bien reconocidos.

Tabla 4. Resumen de las interacciones entre los residuos de AM0627 y el O-glicopéptido P1.

Aminoácido (AM0627)	Aminoácido/glicano (P1)	Tipo de interacción
Asp292 (cadena lateral)	Glu2 (cadena lateral)	Puente de hidrógeno
Trp149 (cadena lateral)	Ala3 (metilo)	CH- π
Phe290 (cadena lateral)	Ala3 (metilo)	CH- π
Arg291 (<i>backbone</i>)	Gln4 (cadena lateral)	Puente de hidrógeno
Tyr470 (cadena lateral)	Thr5 (<i>backbone</i>)	Puente de hidrógeno
Tyr287 (cadena lateral)	Thr5 (metilo)	CH- π
Tyr287 (cadena lateral)	Thr6 (<i>backbone</i>)	Puente de hidrógeno
Tyr470 (cadena lateral)	GalNAc ^{G1}	CH- π
Tyr470 (cadena lateral)	Gal ^{G2}	CH- π
Tyr470 (<i>backbone</i>)	Gal ^{G2} (OH6)	Puente de hidrógeno
Trp321 (cadena lateral)	GalNAc ^{G1'} (grupo carbonilo)	Puente de hidrógeno
Asn347 (cadena lateral)	GalNAc ^{G1'} (metilo)	Puente de hidrógeno
Phe390 (cadena lateral)	GalNAc ^{G1'}	CH- π
Asp318 (cadena lateral)	GalNAc ^{G1'} (OH3)	Puente de hidrógeno
Arg362 (cadena lateral)	GalNAc ^{G1'} (OH3)	Puente de hidrógeno
Tyr288 (cadena lateral)	GalNAc ^{G1'} (OH6)	Puente de hidrógeno
Asp318 (cadena lateral)	GalNAc ^{G1'} (OH6)	Puente de hidrógeno
Arg362 (cadena lateral)	Gal ^{G2'} (oxígeno endocíclico)	Puente de hidrógeno

4.1.3. Determinación de los residuos clave para la actividad de AM0627

Para elucidar la participación de los residuos de AM0627 implicados en las interacciones con los O-glicanos del sustrato P1, se prepararon diferentes mutantes a alanina de los aminoácidos Tyr288, Asp318, Trp321, Asn347, Arg362, Phe390 y Tyr470, y la actividad de las enzimas mutantes resultantes se caracterizaron *in vitro* junto con el mutante a alanina V389A como control positivo, ya que Val389 no mantiene interacciones con P1.

Los ensayos de corte se siguieron por MALDI-TOF y mostraron que tanto el mutante AM0627²¹⁻⁵⁰⁶V389A como la enzima *wildtype* cortaron un 90% del sustrato pasados 10 minutos, que corresponde con el primer punto de la reacción. Por el contrario, la mayoría de mutantes mostraron poca o ninguna actividad metalopeptidasa. En general, la mayoría de los mutantes a alanina mostraron menos del 20% de hidrólisis pasados 10 minutos, aunque cabe destacar que pasadas 2 horas solo R362A y F309A mostraron un incremento significativo de la hidrólisis llegando al 80 y 60% respectivamente (**Figura 37**).

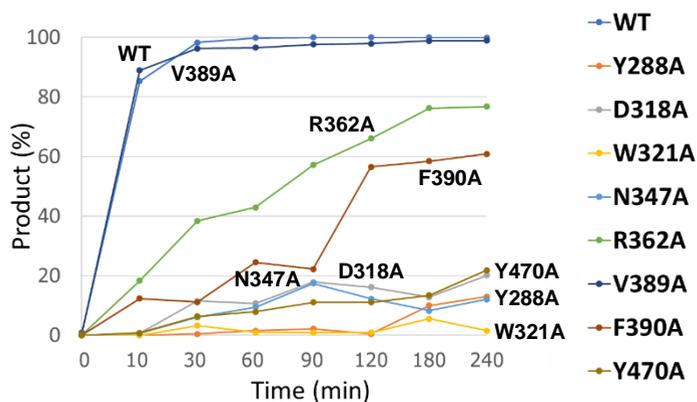


Figura 32. Evolución temporal de los ensayos de corte con $0,4 \mu\text{M}$ de *wildtype* AM0627²¹⁻⁵⁰⁶ y los diferentes mutantes a alanina junto con $57 \mu\text{M}$ de P1 incubados a 37°C . El corte se siguió mediante espectros de MALDI-TOF para detectar el pico de sustrato y de producto de donde se estimó porcentaje de sustrato remanente y el porcentaje de producto formado. Los espectros de MALDI-TOF se pueden ver en la **Figura 38**. Todos los experimentos se realizaron por duplicado. Los colores para la enzima *wildtype* y sus mutantes vienen indicados en la leyenda.

El análisis por MALDI-TOF de la reacción del corte usado para seguir la reacción no permite un análisis más cuantitativo ya que es difícil cuantificar el corte a bajas concentraciones (**Figura 38**).

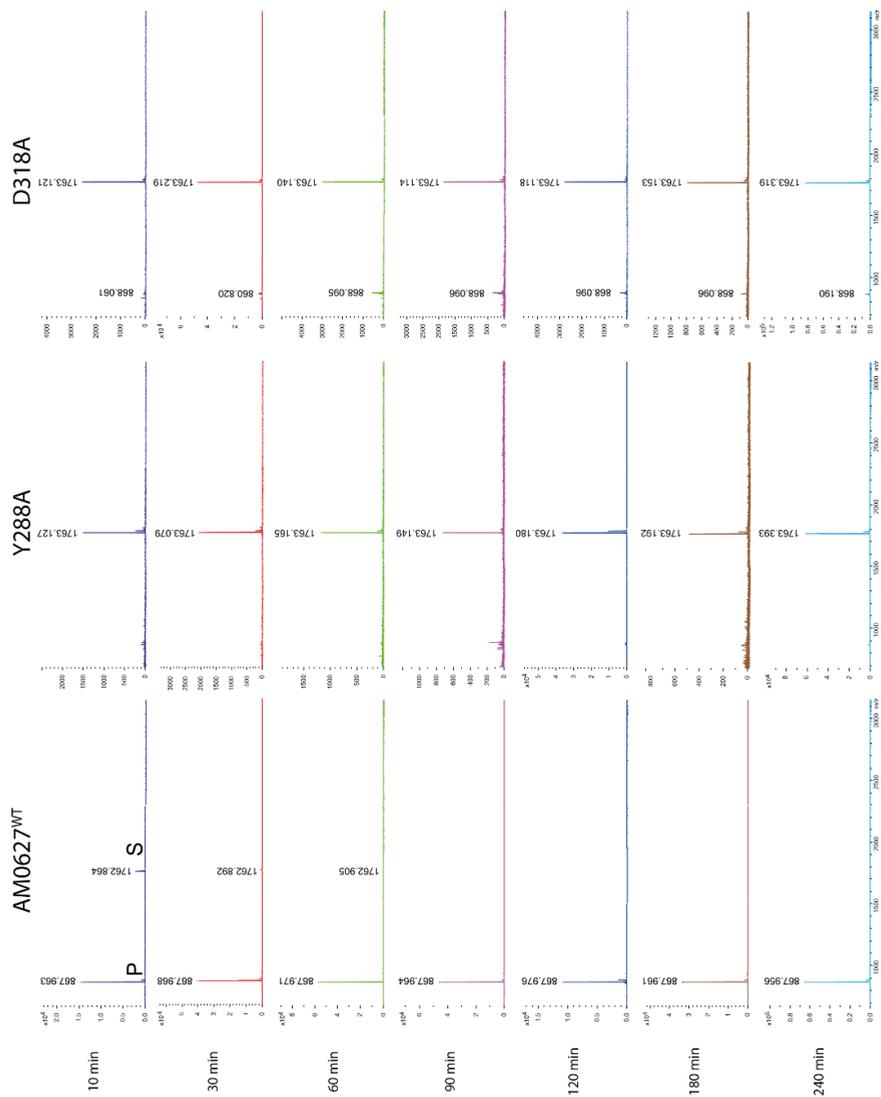


Figura 33. Espectros de MALDI-TOF de los ensayos de corte *in vitro* del AM0627²¹⁻⁵⁰⁶ y sus mutantes frente a P1. En todos los experimentos se usaron 600 ng de AM0627²¹⁻⁵⁰⁶ y sus mutantes. Se semicuantificó una intensidad máxima de sustrato (S) y producto (P) y se estimó la relación de sustrato restante.

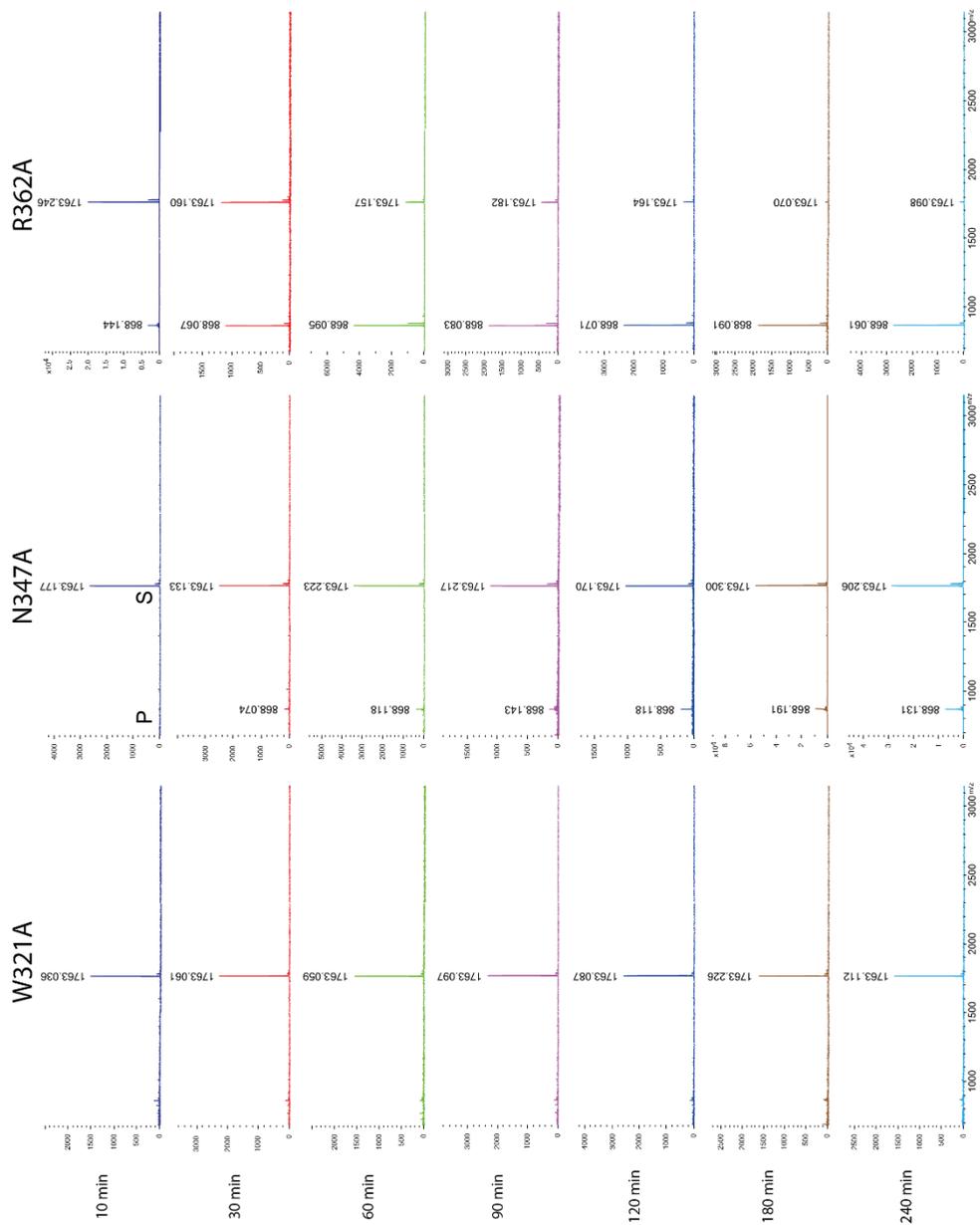


Figura 38. Continuación

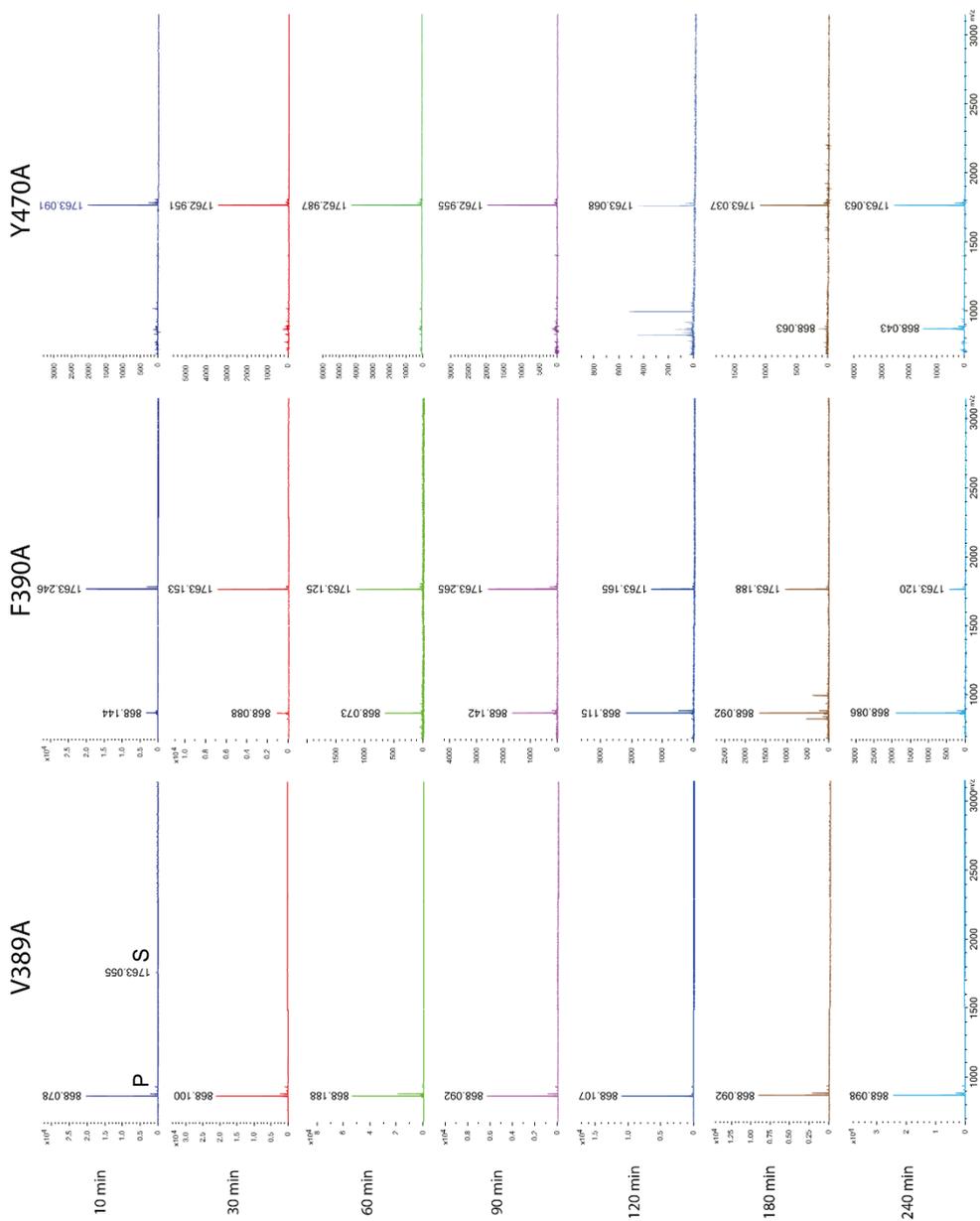


Figura 38. Continuación

Cabe destacar que los mutantes de R362A y F390A, aun teniendo ~30 veces menos actividad que la enzima *wildtype*, mostraron significativamente más actividad que el resto de mutantes a alanina, demostrando que no son esenciales para la actividad metalopeptidasa.

Tyr470, el único residuo que establece interacciones con el GalNAc y Gal en las subunidades G1 y G2 respectivamente (**Figura 36, apartado 4.1.2.**), así como en los residuos Tyr288, Asp318, Trp321, y Asn347 que interactúan con el GalNAc en la subunidad G1' (**Figura 35, apartado 4.1.2.**), todos ellos demuestran ser críticos para la actividad de AM0627.

4.1.4. Caracterización del epítipo mínimo de interacción en P1 para el reconocimiento por parte de AM0627

Una vez se determinó cómo AM0627 reconoce al O-glicopéptido y los aminoácidos que son esenciales para que tenga lugar la catálisis, se quiso demostrar cual es la estructura de O-glicanos mínima necesaria para el reconocimiento y la actividad de AM0627.

Con esta idea, se diseñó un conjunto de O-glicopéptidos (P1 a P9) que contenían GalNAc, GalN, y Gal β 1-3GalNAc (**Figura 21, apartado 3.1.4.1.**). Se usó GalN en vez de GalNAc en algunos de estos O-glicopéptidos para estudiar la participación del grupo acetilo en el reconocimiento de los sustratos por parte de AM0627.

En los ensayos de corte seguidos de nuevo por MALDI-TOF en los que se usó la enzima AM0627²¹⁻⁵⁰⁶ junto con los diferentes O-glicopéptidos P1 a P9, se observó que tanto P1 y P2, con bis T y bis Tn O-glicanos, respectivamente, fueron los mejores sustratos para AM0627, siendo P1 ligeramente más rápido que P2 en los puntos iniciales de la reacción (**Figura 39**).

Cabe destacar que la mayor actividad específica de AM0627 con el sustrato P1 se tuvo que estimar, ya que a altas concentraciones de sustrato (500 μM) y tiempos cortos de reacción, la reacción era prácticamente lineal llegando al 50% de corte de P1 a los 10 minutos y 100% a los 120 minutos. La actividad específica para P1 fue de $1,9 \mu\text{mol enzima} \cdot \text{min}^{-1} \cdot \text{mg}^{-1}$ (1,9 U/mg). Para P2 se observó una actividad 1,6 veces más baja que la de P1, mostrando una actividad de 1,2 U/mg, siendo por tanto P2 el segundo mejor sustrato. El mono Tn O-glicopéptido P5 (-T*-T-) mostró menos del 10% de corte a los 5 minutos, siendo un sustrato más débil seguido por el bis Tn O-glicopéptido modificado con una GalN en la Thr6 de P8 (-T*-T^). Mientras P8 tuvo menos del 10% de corte a los 15 minutos, el mono T O-glicopéptido P3 (-T**-T-) tuvo un corte prácticamente indetectable con menos del 5% pasados 15 minutos de reacción (**Figura 39**).

En el caso de los O-glicopéptidos mono T P4 (-T-T**-), mono Tn P6 (-T-T*) y el bis Tn O-glicopéptido modificado en la Thr5 con una GalN (P7, -T^*T*-), ninguno mostró capacidad de hidrólisis (**Figura 39**). Como era de esperar, el péptido sin O-glicanos P9 (-T-T-) fue completamente inactivo (**Figura 39**).

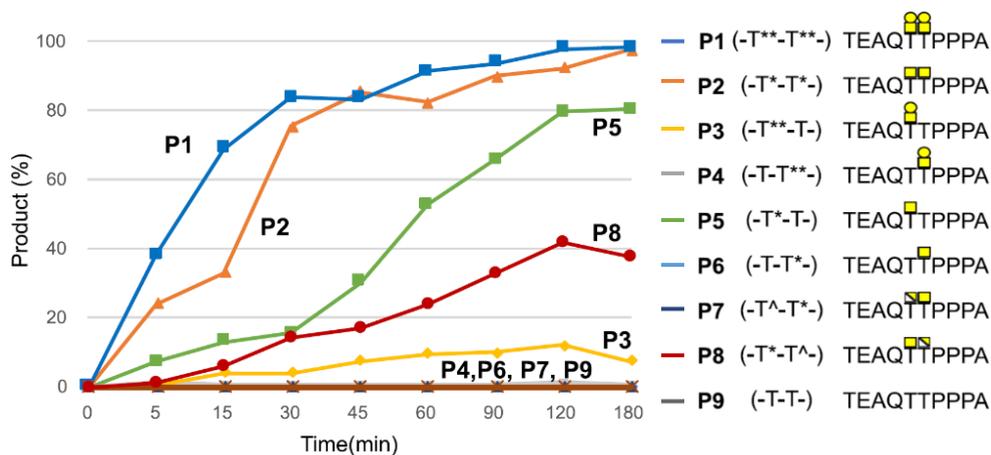


Figura 34. Evolución temporal de los ensayos de corte con $0,4 \mu\text{M}$ de AM0627²¹⁻⁵⁰⁶ con los diferentes glicopéptidos a $500 \mu\text{M}$ incubados a $37 \text{ }^\circ\text{C}$. Los espectros de MALDI-TOF se pueden ver en la **Figura 40**. Todos los experimentos se realizaron por duplicado. Los colores para cada péptido aparecen indicados en la leyenda, así como un esquema de su estructura.

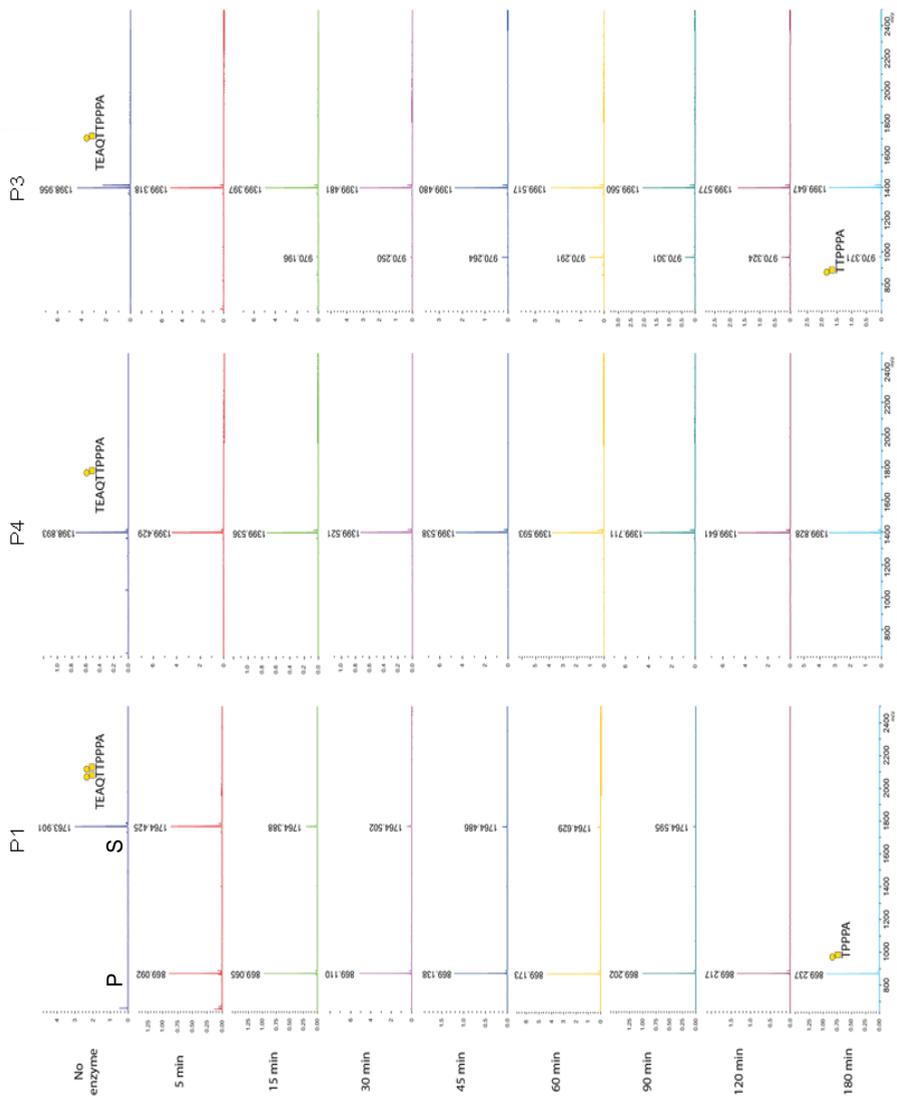


Figura 35. Espectros de MALDI-TOF de los ensayos de corte *in vitro* de AM0627²¹⁻⁵⁰⁶ frente a los glicopéptidos P1 a P9. En todos los experimentos se usaron 600 ng de AM0627²¹⁻⁵⁰⁶. Se semicuantificó una intensidad máxima de sustrato (S) y producto (P) y se estimó la relación de sustrato restante.

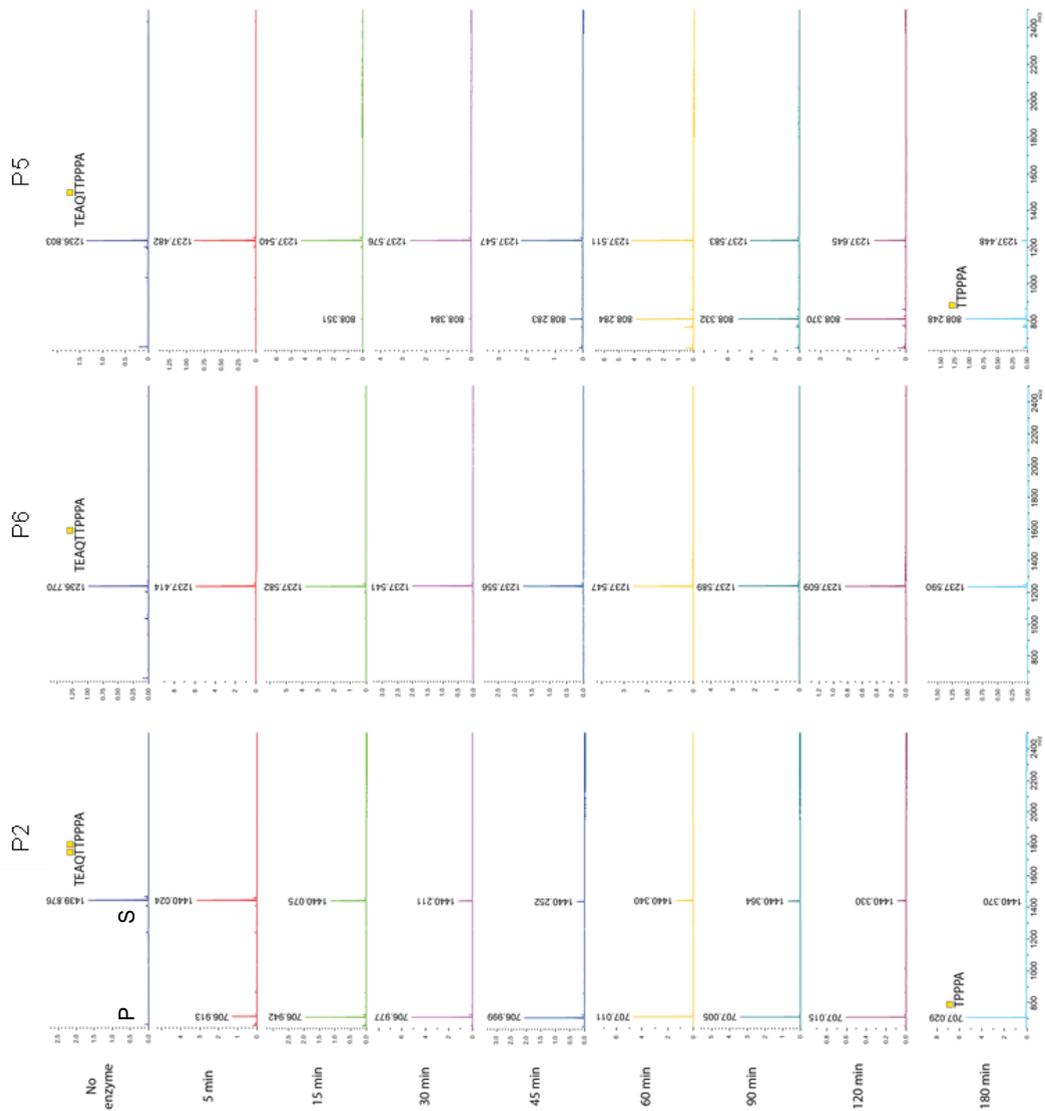


Figura 40. Continuación

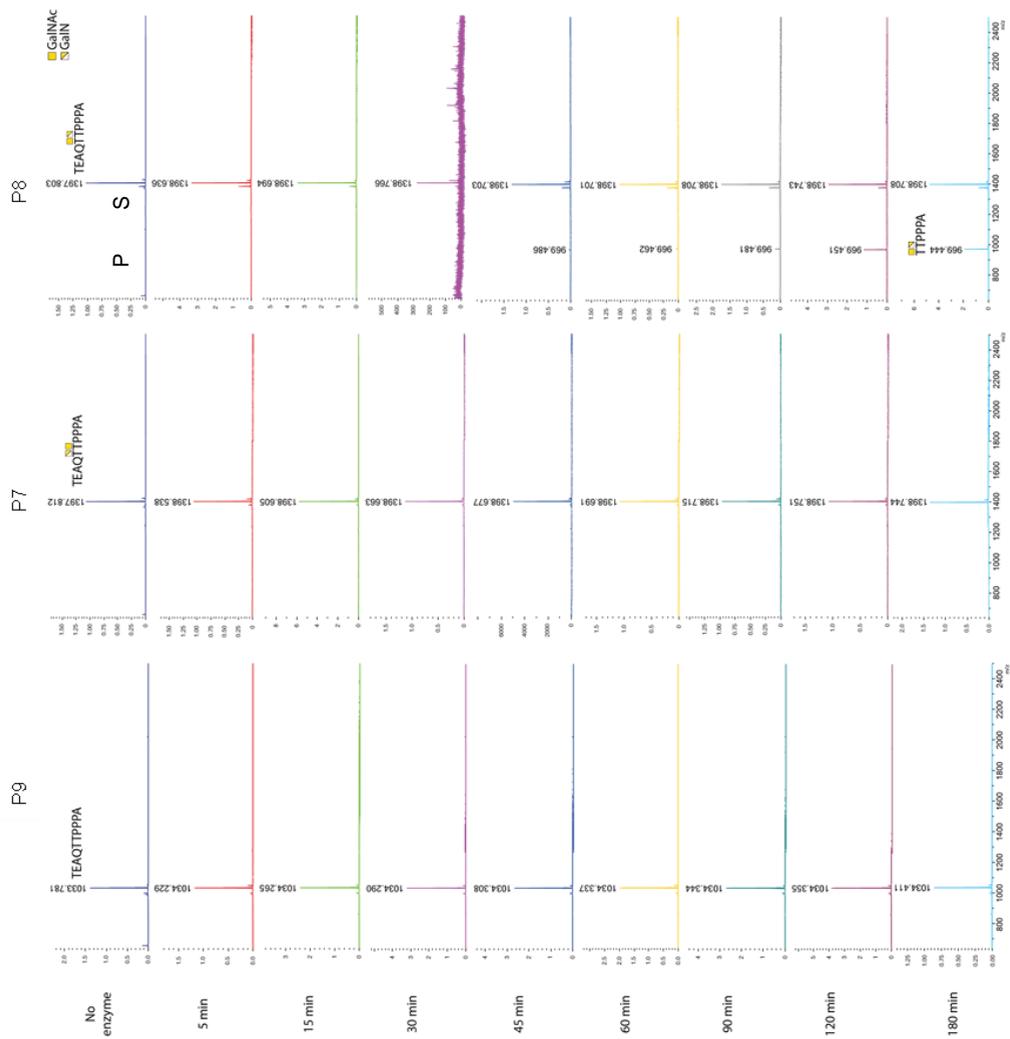


Figura 40. Continuación

4.1.5. Estudio de la estabilidad de los complejos de AM0627 con P1, P2 y P9

Para terminar de comprender estos resultados, se realizaron simulaciones de dinámica molecular de AM0627 junto con los sustratos P1, P2 y P9. La estructura del complejo AM0627⁷¹⁻⁵⁰⁶E326A-P1-Zn²⁺ se usó como modelo para las simulaciones, pero reestableciendo el Glu326 en la estructura.

Para analizar el movimiento entre los sustratos y AM0627⁷¹⁻⁵⁰⁶ se siguieron las distancias moleculares de coordinación para el zinc con los residuos de histidina (His324-Nε⋯Zn²⁺ y His329-Nε⋯Zn²⁺) y las distancias de interacción entre Arg291 y la cadena lateral del Glu4 del glicopéptido (NR291⋯OQ4 y NR291⋯NQ4) ya que estas solo se mantienen cuando el complejo adopta una conformación reactiva (**Figura 41**).

Lo primero que cabe destacar es que el péptido P9 era muy dinámico e inestable como cabía esperar, siendo incapaz de mantener completamente las interacciones en ningún punto de la simulación y la mayoría del tiempo tomaba conformaciones en las que el complejo está prácticamente desensamblado (**Figura 41 Abajo**).

Sin embargo, se observa en las simulaciones de P1 y P2, estos se mostraron mucho más estables por largos periodos de tiempo durante la simulación (**Figura 41 Arriba**), manteniendo las interacciones relevantes (**Figura 42**).

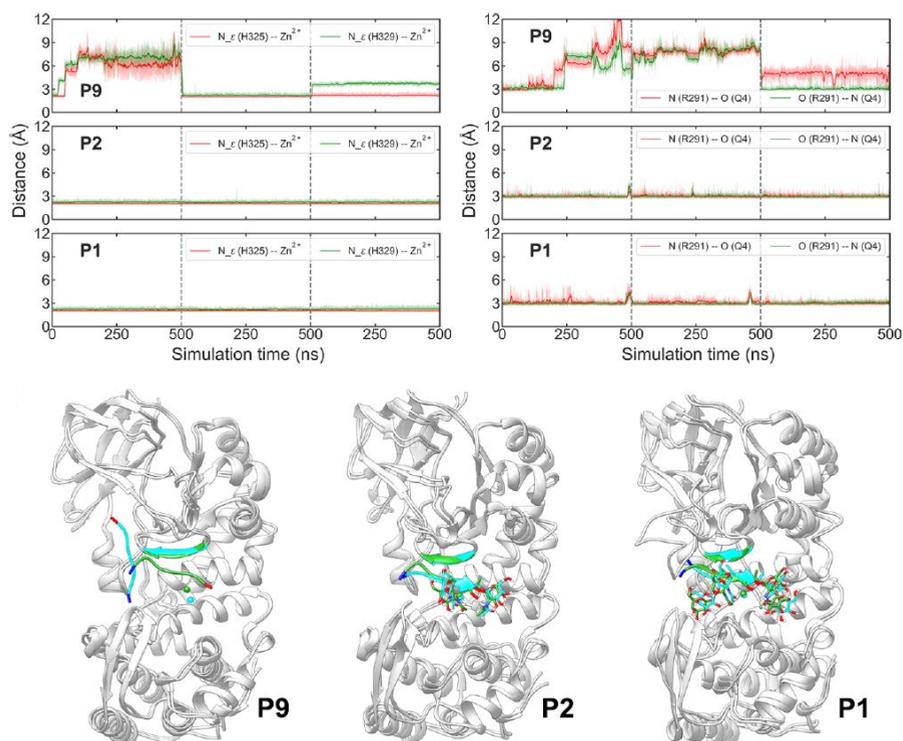


Figura 36. Arriba. Evolución de las distancias ($\text{His324-N}\epsilon\cdots\text{Zn}^{2+}$, $\text{His329-N}\epsilon\cdots\text{Zn}^{2+}$ y $\text{NR291}\cdots\text{OQ4}$ y $\text{NR291}\cdots\text{NQ4}$) en las simulaciones de dinámica molecular de AM0627^{71-506} en complejo con P1, P2 y P9, así como el Zn^{2+} . Se muestra cómo P9 es más inestable y por tanto tiene mayor variación en las distancias que los O-glicopéptidos P1 y P2. Las simulaciones se realizaron por triplicado (3 experimentos independientes). **Abajo.** Estructura general de AM0627^{71-506} en complejo con P1, P2 y P9, así como el Zn^{2+} mostrando la dinámica de los O-glicopéptidos durante las simulaciones donde se observa cómo P9 se aleja del sitio activo mientras que P2 y P1 se mantienen estables en el sitio activo. Las posiciones iniciales de los glicopéptidos y la lámina β formada por Met289-Asp292 y el Zn^{2+} se muestran en verde, mientras que las posiciones finales (después de una de las réplicas de 500 ns) se muestran en cian. Los extremos N y C terminales de los glicopéptidos están coloreados en azul y rojo, respectivamente.

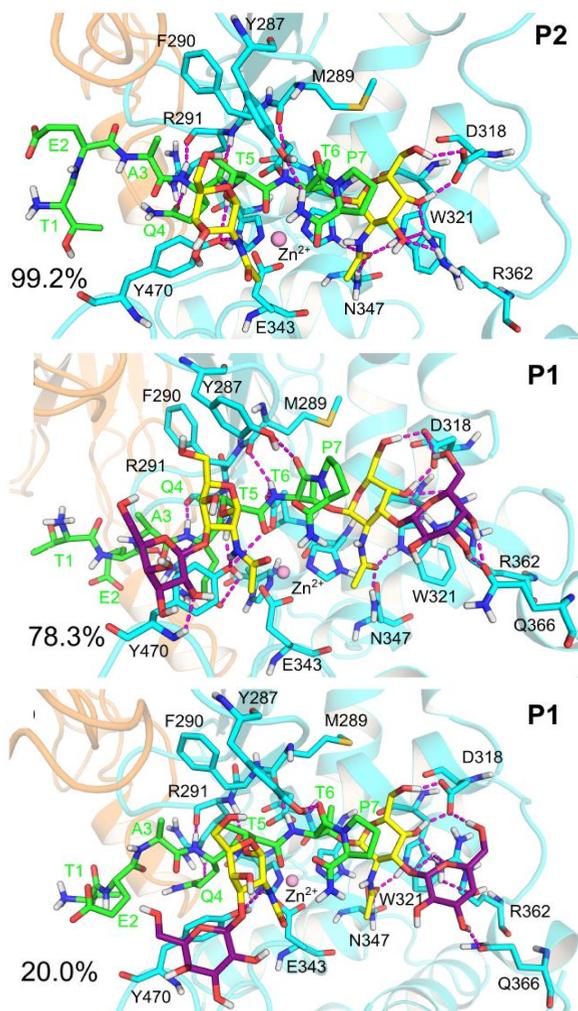


Figura 37. Conformaciones representativas del clúster más poblado del complejo AM0627⁷¹⁻⁵⁰⁶-P2-Zn²⁺ y de los dos grupos más poblados del complejo AM0627⁷¹⁻⁵⁰⁶-P1-Zn²⁺. El agrupamiento se realizó sobre las tres réplicas en función del RMSD de los átomos del *backbone* del péptido con un límite de 0,5 Å después de un alineamiento de los átomos del *backbone* de AM0627⁷¹⁻⁵⁰⁶ usando CPPTRAJ de Amber 20. También aparece representado el porcentaje de representación para cada clúster.

4.1.6. Estudio del mecanismo de reacción de AM0627 a nivel computacional

Para conocer en más detalle el mecanismo de catálisis de AM0627, se llevaron a cabo simulaciones de metadinámica de QM/MM del complejo AM0627⁷¹⁻⁵⁰⁶-P1-Zn²⁺.

Se comprobó cuál de los dos mecanismos de catálisis posibles para zinc metalopeptidasas seguía AM0627, dependiendo de si el residuo nucleofílico, Glu326, ataca directamente al átomo de carbono del carbonilo del enlace peptídico que se va a hidrolizar o lo hace a través de una molécula de agua de manera indirecta (**Figura 43**).

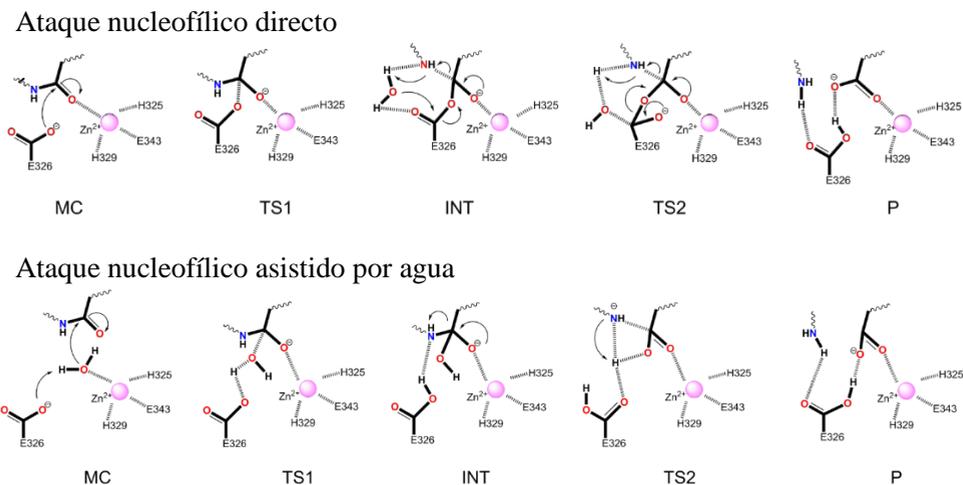


Figura 38. Esquema de los dos mecanismos propuestos para la hidrólisis de enlaces peptídicos catalizada por AM0627. **Arriba.** El ataque nucleofílico directo donde el Glu326 ataca directamente al enlace peptídico, y **Abajo.** el ataque nucleofílico asistido por agua, donde el Glu326 actúa como base desprotonando la molécula de agua que efectuará el ataque nucleofílico sobre el enlace peptídico.

4.1.6.1. Mecanismo guiado por el ataque nucleofílico directo de Glu326

La conformación del sitio activo para el complejo AM0627⁷¹⁻⁵⁰⁶-P1-Zn²⁺ obtenida de las simulaciones de QM/MM muestra que el Zn²⁺ mantiene la coordinación típica tetraédrica (Laitaoja et al., 2013) y un átomo de oxígeno del grupo carboxilato de Glu326 se mantiene a unos 4 Å del átomo C* (**Figura 44**) la cual es una conformación reactiva en la que Glu326 realizaría el ataque nucleofílico directo al carbonilo del péptido.

Siguiendo la distancia Glu326-O...C* como variable colectiva (CV1) (**Figura 44 a**), las simulaciones de metadinámica QM/MM para la catálisis mediante el ataque directo de Glu326 mostraron que esto conllevaba una alta barrera de energía con la formación de un complejo inestable (**Figura 44 b**).

Esto sugiere que el ataque nucleófilo directo de Glu326 seguramente no sea el mecanismo de acción, sino que estará mediado a través de una molécula de agua como se muestra a continuación.

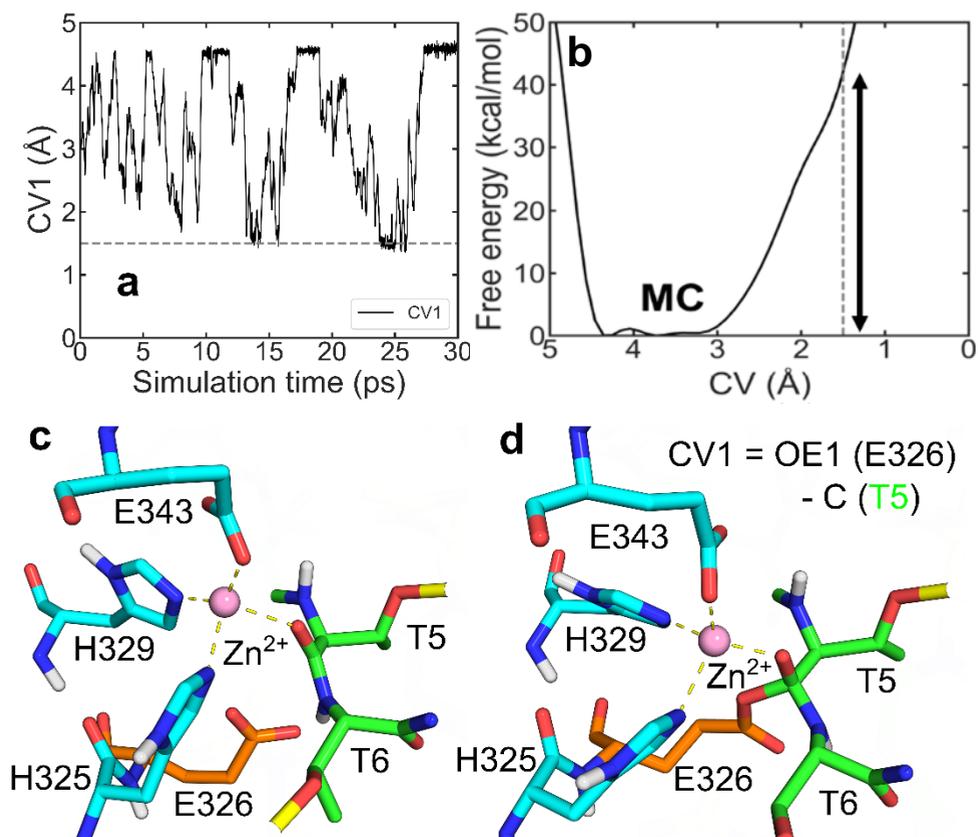


Figura 39. Perfil de energía libre y estados del mecanismo de hidrólisis de AM0627 obtenido de simulaciones de QM/MM para el ataque nucleofílico directo. **a.** Variaciones de la CV1 a lo largo de la dinámica. **b.** Perfil de energía libre donde se observa un valor de energía libre alto que lleva a un estado inestable del sistema. **c y d.** *Snapshots* de los distintos estados presentes durante la simulación: siendo **c**, el complejo de Michaelis (MC), y **d**, el estado intermedio inestable en el que Glu326 está unido a Thr5. Las líneas de puntos en amarillo representan los enlaces de coordinación del Zn^{2+} . Para los aminoácidos se mantiene el esquema de colores anterior exceptuando el aminoácido Glu326 (naranja) que actúa de base durante la catálisis.

4.1.6.2. Mecanismo guiado por el ataque nucleofílico indirecto mediado por una molécula de H₂O

Las simulaciones de QM/MM se repitieron, pero teniendo en cuenta una molécula de agua que se encontró en el sitio activo de la estructura cristalina. Ésta se mantuvo en el sitio activo y estable durante la simulación de QM/MM y se encontraba coordinada con el ion de Zn²⁺ y formando un puente de hidrógeno con el Glu326. El átomo de oxígeno de la molécula de agua se encuentra a 3 Å del C*, estando bien orientada para el ataque nucleofílico hacia el átomo C*.

Para comprobar este posible mecanismo para la catálisis, se llevaron a cabo de nuevo simulaciones de QM/MM del ataque nucleófilo y se siguieron usando la distancia entre C* y el oxígeno del agua como CV.

Durante la simulación, un protón se transfiere desde el agua al Glu326, que actúa como base, mostrando una energía de activación de 12,1 kcal/mol (**Figura 45 a**), el cual es un valor en consonancia con los valores publicados previamente para zinc metalopeptidasas (Wu et al., 2010; Xu y Guo, 2009), sugiriendo que la reacción es posible. Esta simulación llevó a la formación de un intermedio estable (INT) (**Figura 45 d**) en el que el C* se encuentra tetracoordinado y el enlace C*-N se estira con respecto al MC de 1,39 Å a 1,52 Å en el estado intermedio, pero sin llegar a ser hidrolizado.

Para el segundo paso de la reacción enzimática, se llevaron a cabo otro conjunto de simulaciones de metadinámica QM/MM, pero partiendo del intermedio tetraédrico (**Figura 45 d**). En estas simulaciones se observa la transferencia del protón del C*-OH al átomo de N lo que conlleva al corte

completo del enlace peptídico (**Figura 45 f**) con una energía de activación de 5,6 kcal/mol.

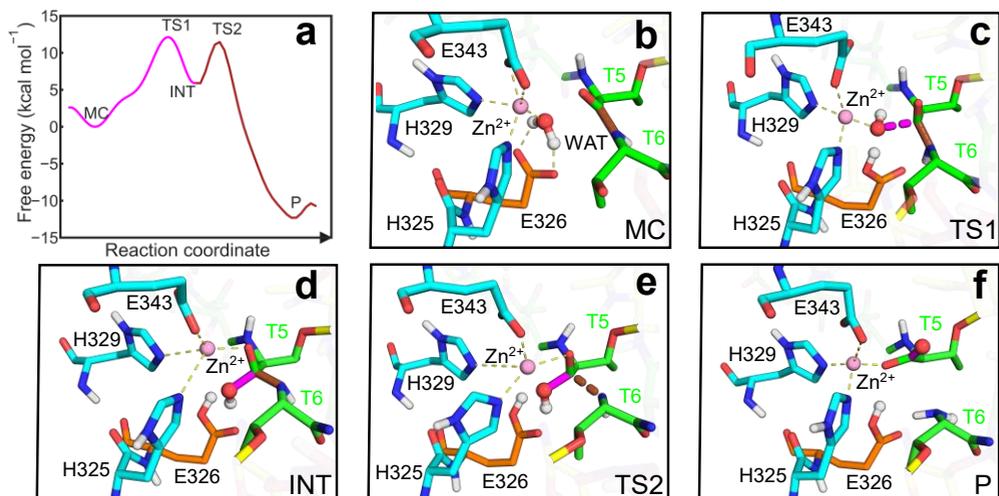


Figura 40. Perfil de energía libre y estados del mecanismo de hidrólisis de AM0627 obtenido de simulaciones de QM/MM para el ataque nucleofílico asistido por una molécula de agua. **a.** Perfil de energía libre donde se observa un valor de 12,1 kcal/mol para el primer estado de transición (TS1) y una energía de 5,6 kcal/mol entre el estado intermedio (INT) y el segundo estado de transición (TS2). **b-f.** *Snapshots* de los distintos estados presentes durante la simulación, el MC, el TS1, el INT, el TS2 y el producto de la reacción (P). Las líneas de puntos en marrón representan enlaces formándose o deshaciéndose en los estados de transición, las líneas de puntos en amarillo representan los enlaces de coordinación del Zn²⁺. Para los aminoácidos se mantiene el esquema de colores anterior exceptuando el aminoácido Glu326 (naranja) que actúa de base durante la catálisis.

4.1.7. Comparación estructural de AM0627 frente a otras mucinasas y O-glicoproteasas

4.1.7.1. AM0627 y las mucinasas BT4244, ZmpB, ZmpC e IMPa

AM0627 se comparó a nivel estructural con otras mucinasas y O-glicoproteasas previamente publicadas para entender el por qué AM0627 parecía ser la única mucinasa con capacidad de hidrólisis dependiente de bis O-glicanos. Por ello, se analizaron los sitios activos de las mucinasas previamente publicadas BT4244 (PDB: 5KD8) (Noach et al., 2017b), ZmpB /ZmpC (PDBs: 5KDU/6XT1) (Noach et al., 2017b; Pluinage et al., n.d.) e IMPa (PDB: 5KDX) (Noach et al., 2017b) en complejo con sus sustratos Tn, mSTb y T O-glicanos respectivamente (**Figura 46**), y se compararon con el sitio activo de AM0627.

Inspeccionando los distintos sitios activos de las mucinasas se puede ver que los residuos de cada enzima que rodean a la molécula de GalNAc en G1' (Trp321, Asn347, Arg362 y Tyr288 para AM0627; Trp570, Asn595, Arg611 y Tyr538 para BT4244; Trp752, Asn775, Arg790 y Phe727 para ZmpB; Trp692, Gln720, Arg742 y Trp685 para IMPa; y Trp747, Asn770, Arg785 y Phe722 para ZmpC) e incluso sus interacciones con este azúcar están principalmente conservadas (**Figura 46**) siendo las únicas excepciones Phe727 de ZmpB y Phe722 de ZmpC, que establecen interacciones de CH- π con el Sia de su sustrato mSTb, y Trp685 de IMPa que no interacciona con ningún azúcar (**Figura 46**).

En cuanto al nivel de conservación de los residuos en la subunidad S, solo BT4244 comparte el residuo de tirosina que ha demostrado ser clave en AM0627, Tyr470 para AM0627 y Tyr723 para BT4244 (**Figura 46**).

Para profundizar en la posible función de Tyr723 de BT4244, se superpuso la estructura cristalina de AM0627 sobre las de BT4244, ZmpB, IMPa, y ZmpC. Se encontró que para AM0627/BT4244 el RMSD era el menor, de 2,02 Å y el número de residuos alineados el mayor, de 394, seguida de AM0627/ZmpB (2,25 Å y 351 residuos) junto con AM0627/ZmpC (2,26 Å y 342 residuos) y por último AM0627/IMPa (2,99 Å y 342 residuos).

Inspeccionando en detalle el centro activo del complejo entre BT4244, su sustrato Tn O-glicano y P1 superpuesto del complejo AM0627⁷¹⁻⁵⁰⁶-P1-Zn²⁺, se observa cómo la Tyr723 se encuentra localizada de tal manera que seguramente reconozca los O-glicanos localizados en G1 y G2 (**Figura 46**).

Esto indica que AM0627 y BT4244 se deben comportar de manera muy similar en cuanto al reconocimiento de bis O-glicanos y que BT4244 seguramente corte glicopéptidos que contengan bis O-glicanos, como se ha demostrado recientemente sobre los motivos Tn de MUC1, GS*T*A y VT*S*A (Coelho et al., 2022; Konstantinidi et al., 2022b).

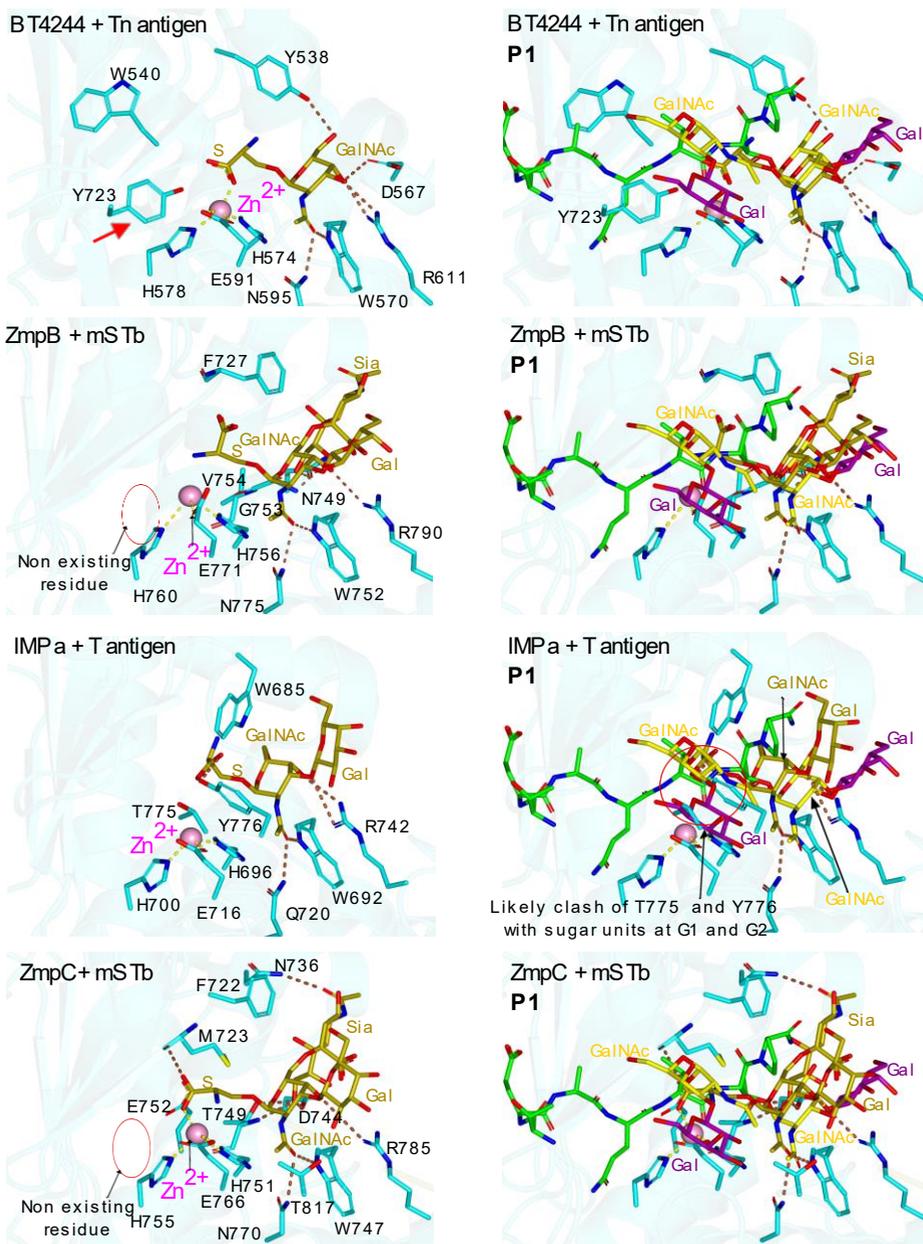


Figura 41. Análisis de los sitios activos de BT4244, ZmpB, IMPa y ZmpC. **Izquierda.** Vista en detalle de los sitios activos de los complejos BT4244-Tn O-glicano, ZmpB-mSTb, IMPa-

T O-glicanos y ZmpC-mSTb. **Derecha.** Vista en detalle de los sitios activos de los complejos BT4244-Tn O-glicano, ZmpB-mSTb, IMPa-T O-glicano y ZmpC-mSTb superponiendo P1 de la estructura del complejo AM0627⁷¹⁻⁵⁰⁶E326A-P1-Zn²⁺. Se mantiene el esquema de colores para los aminoácidos procedentes de las enzimas, P1, el Zn²⁺ y los enlaces de coordinación del Zn²⁺ y los puentes de hidrógeno entre las enzimas y los sustratos. Los ligandos en estas estructuras se muestran como átomos de carbono en verde oliva. La flecha roja indica la Tyr723 de BT4244 la cual está conservada con la Tyr470 en AM0627.

En cuanto al resto de mucinasas, mientras que ZmpB y ZmpC no contienen el residuo aromático cerca de los O-glicanos G1 y G2, IMPa contiene dos treoninas, Thr775 y Tyr776, que generan impedimento estérico con el GalNAc y Gal en las subunidades G1 y G2 respectivamente, siendo por lo tanto incapaz de acomodar estos O-glicanos en el sitio activo (**Figura 46**).

Podemos decir, por tanto, que la ausencia del residuo aromático colocado para interactuar con los glicanos en las subunidades G1 y G2 es la razón por la que ZmpB, ZmpC, y IMPa no son capaces de cortar O-glicopéptidos que contengan bis-O-glicanos.

4.1.7.2. AM0627 y la O-glicoproteasa OgpA

La O-glicoproteasa inactiva de *Akkermansia muciniphila* OgpA-H205A-E206A (PDB: 6Z2P) (Trastoy et al., 2020) en complejo con su sustrato, un T O-glicopéptido (**Figura 47**) se comparó con el sitio activo de AM0627 con la intención de entender la diferencia a nivel de estructura entre mucinasas y O-glicoproteasas.

Inspeccionando el sitio activo de OgpA, esta muestra un modo de reconocimiento a su ligando diferente al encontrado por las mucinasas mencionadas previamente. OgpA mostró un RMSD de 4,50 Å y solo 111 residuos alineados con AM0627. Además, los O-glicanos del T O-glicopéptido con respecto a los O-glicanos que están localizados en la subunidad G' de P1, están orientados de distinta manera (**Figura 47**) y los residuos que interaccionan con los O-glicanos (Val164, Phe166, Tyr236, y Lys198) son diferentes (**Figura 47**).

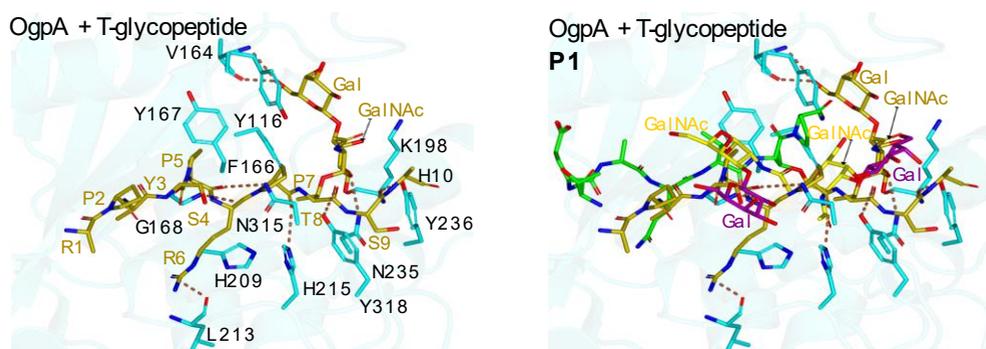


Figura 42. Análisis del sitio activo de OgpA. **Izquierda.** Vista en detalle del sitio activo del complejo OgpA-T O-glicopéptido. **Derecha.** Vista en detalle del sitio activo del complejo OgpA-T O-glicopéptido superponiendo P1 de la estructura del complejo AM0627⁷¹⁻⁵⁰⁶-P1-Zn²⁺. Se mantiene el esquema de colores para los aminoácidos procedentes de las enzimas, P1, el Zn²⁺ y los enlaces de coordinación del Zn²⁺ y los puentes de hidrógeno entre la enzima y los sustratos. Los ligandos en estas estructuras se muestran como átomos de carbono en verde oliva.

Curiosamente, aunque OgpA muestra un residuo aromático Tyr116 en la cercanía del GalNAc en la subunidad G' de P1, éste prefiere interactuar

con el Gal en la subunidad G2' del T O-glicopéptido. Por lo tanto, se encuentran enormes diferencias entre OgpA y el resto de mucinasas en cuanto al sitio activo y al reconocimiento de O-glicanos.

Además, se puede observar varios aminoácidos como Asn315 que generan impedimento estérico impidiendo la localización de los O-glicanos en las subunidades G (**Figura 47**), sugiriendo porqué OgpA no es capaz de cortar enlaces peptídicos en clústeres de bis O-glicanos.

Por último, cabe destacar que, en cuanto al reconocimiento de la secuencia peptídica, OgpA reconoce mayoritariamente el *backbone* de los aminoácidos del sustrato T O-glicopéptido excepto por una interacción de puente de hidrógeno entre la cadena lateral de Arg6 y el *backbone* de Leu213 (**Figura 47**), indicando que OgpA es menos específica respecto a la secuencia de aminoácidos que AM0627, aun siendo esta última promiscua en cuanto a las secuencias de O-glicopéptidos.

4.1.8. Ensayos de actividad de BT4244 por O-glicopéptidos con bis Tn y bis T O-glicanos

Para entender mejor la actividad de AM0627 y BT4244 frente a bis O-glicanos y su tolerancia por O-glicanos complejos, se comparó la actividad de ambas enzimas frente a diferentes formas de O-glicanos usando un *reporter* de los TR de MUC1 formado por entre 5 y 6 repeticiones de TRs de 20 aminoácidos de longitud con secuencia GVT SAPDTRPAPGSTAPPAH y que contiene cinco O-glicositos, repartidos en dos bis O-glicositos y un mono O-glicosito.

Se probó una glicofoma estilo *wildtype* del *reporter* de MUC1 (WT-MUC1), el cual consistía en una mezcla de estructuras tipo *Core 1* y *Core 2* que pueden estar mono o bi-sialiladas, y tres O-glicofomas más homogéneas que contenían mSTa (mSTa-MUC1), T (T-MUC1) y Tn (Tn-MUC1).

Se observó que ni AM0627 ni BT4244 cortaban de manera eficiente los *reporters* de MUC1 sialilados WT-MUC1 y mSTa-MUC1 (**Figura 48**), aun cuando anteriormente se ha sugerido que AM0627 es capaz de cortar O-glicanos bis mSTa (Shon et al., 2020).

Por el contrario, ambas enzimas fueron capaces de hidrolizar los *reporters* de T y Tn MUC1 eficientemente a un ratio de 1:100 enzima/sustrato (**Figura 48**), lo que consolida los estudios de corte usando los O-glicopéptidos P1 a P9 (**Figura 39, apartado 4.1.4.**). Se observó preferencia por parte de AM0627 por la glicofoma T frente a la Tn, mientras que BT4244 mostró preferencia por la glicofoma Tn (**Figura 48**).

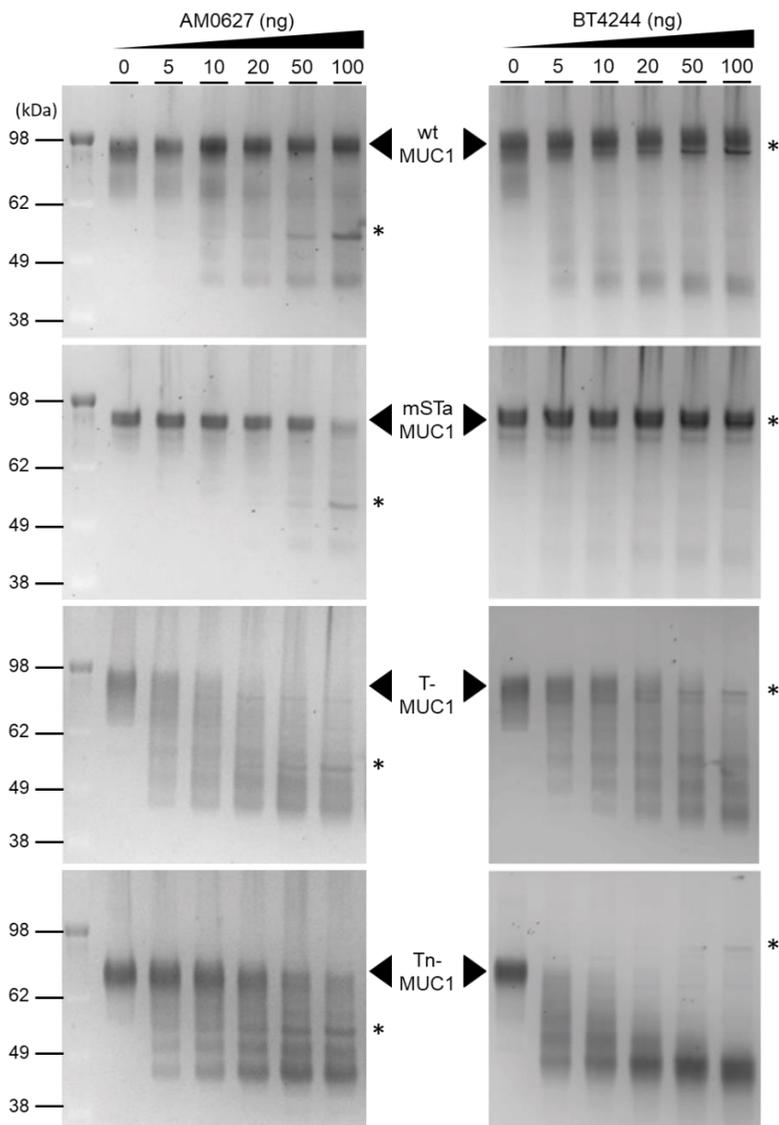


Figura 43. Análisis SDS-PAGE de la actividad de AM0627 y BT4244 frente a *reporters* de MUC1. SDS-PAGE de la digestión de *reporters* de los TRs de MUC1 por parte de AM0627²¹⁻⁵⁰⁶ y BT4244. Estos *reporters* fueron producidos en células HEK293 modificadas para tener O-glicofomas WT, mSTa, T y Tn. Se incubaron 0,5 μ g de cada *reporter* con AM0627²¹⁻⁵⁰⁶ y BT4244 durante 3 horas a 37 °C, con la concentración de enzima variando de 0 a 200 nM. Los

geles se visualizaron con tinción de proteína fluorescente Krypton. * indica las bandas correspondientes a la enzima AM0627²¹⁻⁵⁰⁶ y BT4244. El carril de la izquierda en los geles de AM0627 contiene los marcadores de peso molecular. Además, el peso molecular de cada proteína en el patrón se indica en el lado izquierdo. Los experimentos se realizaron por duplicado.

Para comprender mejor este resultado se realizaron simulaciones de dinámica molecular de los complejos de BT4244 con dos variantes de un O-glicopéptido presente en MUC1 que contienen un bis Tn O-glicano (AHGVT*S*A) y un bis T O-glicano (AHGVT**S**A). En la simulación, se pudo observar que los puentes de hidrógeno formados entre la enzima y el GalNAc en la subunidad G1' del bis Tn O-glicano, son más estables que los que se forman entre BT4244 y el GalNAc en la subunidad G1' del bis T O-glicano (**Figura 49**). Esto explicaría porqué BT4244 prefiere actuar sobre glicopéptidos que contienen clústeres bis Tn sobre bis T.

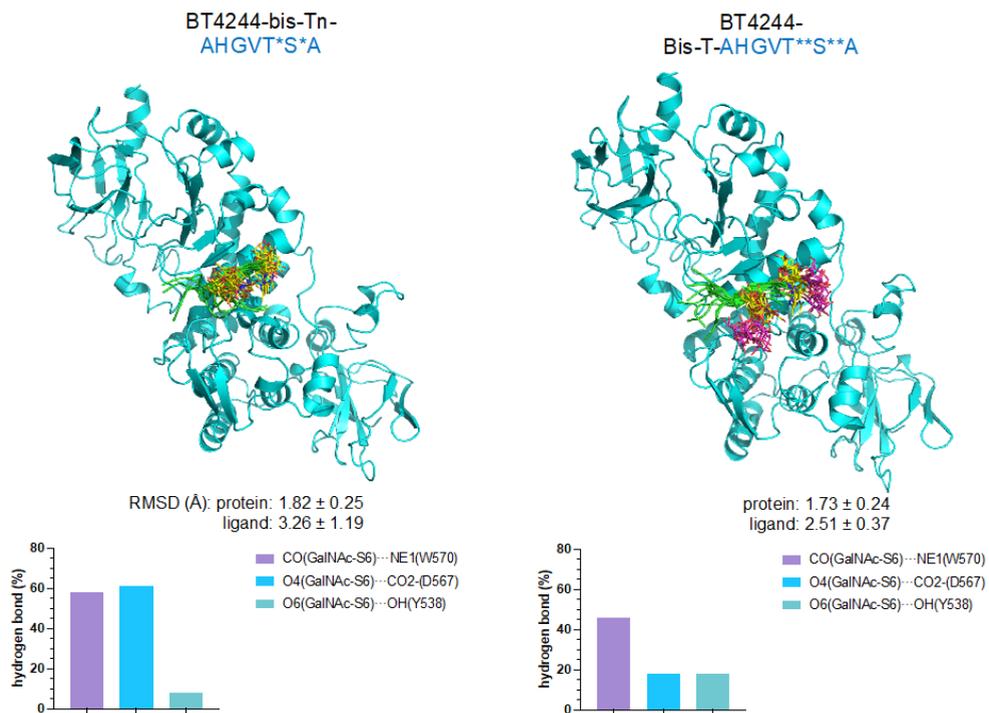


Figura 44. Conjunto de estructuras obtenidas de 0,5 μ s de las simulaciones de MD junto con los valores de RMSD para la proteína (átomos del *backbone*) y el ligando. También se muestra el porcentaje de puentes de hidrógeno encontrados en las simulaciones entre BT4244 y el GalNAc en la subunidad G1' de los ligandos bis Tn (AHGVT*S*A) y bis T (AHGVT**S**A).

4.1.9. Estudio de la capacidad de AM0627 de acomodar residuos de ácido siálico en el sitio activo

Para explicar mejor por qué ambas mucinas no fueron capaces de actuar sobre los *reporters* que contienen ácidos siálicos (WT-MUC1 y mSTa-

MUC1), se inspecciono el sitio activo de la estructura de AM0627 en complejo con P1 (**Figura 50**).

Se pudo observar como el Sia unido al OH3 del Gal en la subunidad G2 provocaría impedimento estérico y colisiones con varios residuos, entre ellos Lys387, Asp388 y Val389 (**Figura 50**). Además, el Gal en la subunidad G2' se encuentra cerca de una superficie de carga negativa de AM0627 que repelería al Sia, desfavoreciendo la unión (**Figura 50**).

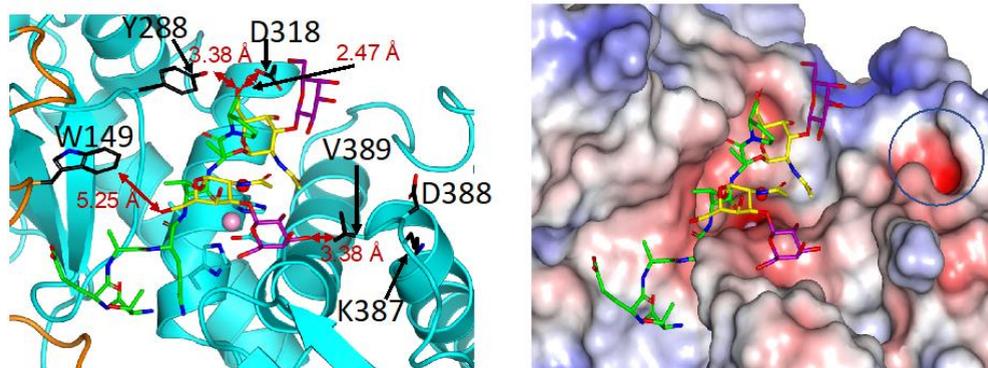


Figura 45. Izquierda. Detalle del sitio activo del complejo AM0627⁷¹⁻⁵⁰⁶E326A-P1-Zn²⁺ donde se muestran las distancias entre los residuos de AM0627⁷¹⁻⁵⁰⁶E326A y los OH6 de los GalNAc en G1 y en G1', y el OH3 del Gal en G2. **Derecha.** Representación de la superficie electrostática de AM0627⁷¹⁻⁵⁰⁶E326A (entre -5 kTe⁻¹ (rojo) y +5 kTe⁻¹ (azul)). El círculo azul indica una superficie con carga negativa que repelería un Sia unido a Gal formando la O-glicoforma mSTa en la subunidad S1'. Se mantiene el esquema de colores utilizado hasta ahora.

Finalmente, también se abordó por qué la glicoforma STn bloquea la escisión del péptido. Cuando se mira en detalle el sitio activo de AM0627 (**Figura 50**), se puede observar que si hubiera residuos de Sia unidos en α -2-6 a los residuos

de GalNAc probablemente chocarían con los aminoácidos Trp149, Tyr288 y Asp318. Hay que tener en cuenta que el OH6 de GalNAc en la subunidad G1 está a 5,25 Å de Trp149, y el OH6 de GalNAc en la subunidad G1' está a 3,38 Å de Tyr288 y 2,47 Å de Asp318 (**Figura 50**).

CAPÍTULO 5. DISCUSIÓN Y CONCLUSIONES

5.1. AM0627

Las mucinas tienen grandes clústeres de O-glicanos como se ha comentado a lo largo de la tesis y sin embargo las mucinasas estudiadas hasta la fecha reconocen únicamente mono O-glicositos. Durante el desarrollo de esta tesis, se ha conseguido demostrar que AM0627 y BT4244 muestran una actividad capaz de hidrolizar bis O-glicopéptidos.

La estructura de la forma inactiva de AM0627 en complejo con P1, es capaz de mostrar como esta mucinasa reconoce bis O-glicanos para llevar a cabo su hidrólisis, reconociendo tanto la cadena peptídica del sustrato P1, como la mayoría de los O-glicanos. Esto sugiere que AM0627 tiene preferencias por secuencias de aminoácidos concretas. El análisis de la estructura cristalográfica de AM0627 demuestra la importancia de los residuos Tyr288, Asp318, Trp321, y Asn347 que interaccionan con el GalNAc en la subunidad G1', así como Tyr470 que interacciona con los azúcares en las posiciones G1 y G2. Se ha podido demostrar que, al mutar estos residuos, la enzima pierde totalmente su actividad, reforzando la idea de que estos residuos resultan la fuerza motora para el reconocimiento de GalNAc en la posición G1' y para que AM0627 se una a sustratos bis T O-glicosilados a través de los glicanos de la subunidad S1. Los análisis cinéticos de AM0627 frente a los distintos sustratos O-glicopeptídicos sugieren que AM0627 tiene una mayor preferencia por O-glicopéptidos con bis T O-glicanos frente a bis Tn O-glicanos y solo es

capaz de actuar sobre mono O-glicanos cuando este se encuentra colocado en la subunidad G1, así como que la ausencia de O-glicanos como en P9 hace imposible la catálisis. Lo que, es más, al realizar MDs de AM0627 junto a P1, P2 y P9, se ha podido mostrar como esta ausencia de O-glicanos hace que el reconocimiento del péptido sea débil y le cueste mantenerse en el sitio activo mientras que los O-glicopéptidos están mejor reconocidos.

Comparando esta estructura cristalina de AM0627²¹⁻⁵⁰⁶E236A con la de otras O-glicoproteasas, se observa una gran diferencia entre este tipo de enzimas y las mucinasas a nivel de reconocimiento de sustrato, donde ninguna de las interacciones entre las enzimas y sus sustratos coinciden aun teniendo un rol similar. A su vez, dentro de las mucinasas, se observa cómo la ausencia del residuo de tirosina que interacciona con los O-glicanos en las posiciones G1 y G2 respectivamente, impide a ZmpB, ZmpC, y IMPa tener la capacidad de hidrolizar este tipo de estructuras. Curiosamente se ha observado que BT4244 sí conserva esta tirosina y se ha probado que es capaz de escindir bis O-glicanos sencillos, como se ha demostrado en otros estudios muy recientemente (Coelho et al., 2022; Konstantinidi et al., 2022b). Estos resultados corroboran la importancia de este residuo en el reconocimiento de este tipo de O-glicanos.

Casi al mismo tiempo que apareció nuestro trabajo sobre AM0627, se publicó por otro grupo la estructura cristalina de AM0627 junto con Zn²⁺, pero en este estudio no se determinaron las bases moleculares para el reconocimiento y preferencia de bis O-glicanos (Shon et al., 2022), ya que el sustrato bis T O-glicano, que fue añadido utilizando técnicas de *docking*, no parecía estar en el

sitio correcto. En este estudio se establece que las interacciones prioritarias con el primer O-glicano se producen con los residuos Trp149 y Phe290 (Shon et al., 2022), lo que claramente no coincide con la estructura cristalina obtenida de AM0627. Igualmente, la función esencial de la Tyr470 de reconocer los O-glicanos en la subunidad S1, no se pudo observar debido a la localización errónea del sustrato (Shon et al., 2022).

El análisis estructural llevado a cabo, respalda el hallazgo de que las mucinasas AM0627 y BT4244 no son capaces de hidrolizar sustratos con O-glicanos cubiertos por Sia debido a posibles factores estéricos y colisiones entre estos glicanos y los aminoácidos presentes en el sitio activo. El grupo de Carolina Bertozzi (Shon et al., 2022) sugirió previamente que AM0627 era capaz de cortar estas estructuras sialiladas provenientes de un sustrato heterogéneo de MUC2 nativa. Sin embargo, los ensayos de hidrólisis con *reporters* de MUC1 de estructura definida (Nason et al., 2021), permiten confirmar que AM0627 y BT4244 son incapaces de hidrolizar estructuras sialiladas de manera efectiva, mostrando ambas preferencias de sustrato similares incluso cuando AM0627 y BT4244 muestran una ligera preferencia por bis T y bis Tn O-glicanos, respectivamente.

Gracias a las simulaciones de QM/MM se ha podido definir el mecanismo de AM0627 como un mecanismo de reacción en dos pasos en el que una molécula de agua es la que lleva a cabo el ataque nucleofílico sobre el grupo carbonilo del enlace peptídico hidrolizable una vez el residuo de Glu326 actúa como base arrebatándole un protón, lo que es consistente con la familia de zincin (Cerdà-Costa y Gomis-Rüth, 2014).

El hecho de que el sustrato P7 (-T⁺-T^{*}-) no muestre actividad cuando P8 (-T^{*}-T⁺-) sí la muestra, sugiere que este grupo acetilo del GalNAc en G1 resulta esencial para retener el agua que lleva a cabo el ataque nucleofílico y por tanto crítico para la catálisis. A su vez se puede hipotetizar que la ausencia del acetilo podría llevar a la formación de un residuo de GalN protonado que podría impedir la captura de la molécula de agua.

Los hallazgos encontrados en esta parte de la tesis resultan de vital importancia ya que las mucinas no contienen únicamente sitios aislados de O-glicosilación sino grandes clústeres de O-glicanos que han de ser degradados, como en el caso de MUC2 y MUC5AC con clústeres de hasta seis O-glicositos (Schjoldager et al., 2020).

Por todo lo demostrado durante el desarrollo de este capítulo de la tesis, se pueden extraer las siguientes conclusiones respecto a AM0627:

- Se ha conseguido la estructura cristalográfica del complejo AM0627⁷¹⁻⁵⁰⁶E236A-P1-Zn²⁺, donde se han conseguido identificar los residuos clave en el reconocimiento de los O-glicanos, y se ha visto que el GalNAc en G1` es el más reconocido.
- Se han conseguido expresar y purificar mutantes a alanina de dichos residuos y se ha conseguido caracterizar su capacidad de hidrólisis, demostrando que los residuos que reconocen los O-glicanos son indispensables para la catálisis.
- Se ha demostrado la importancia de la presencia de O-glicanos para que AM0627 sea capaz de llevar a cabo la hidrólisis, así como sus

preferencias en cuanto a sustrato, siendo como mínimo necesario un GalNAc en la posición G1 para que tenga lugar la catálisis.

- Se ha demostrado la importancia del residuo de Tyr470 en AM0627 en el reconocimiento de bis T O-glicanos. Este mismo residuo de tirosina se ha encontrado en BT4244 y se ha demostrado su capacidad de escindir bis O-glicanos, corroborando la función de este aminoácido.
- Se ha demostrado que el mecanismo de reacción se produce en dos etapas por medio de un ataque nucleofílico mediado por agua donde el Glu236 actúa como base desprotonando al agua.
- La comparación entre OgpA y AM0627, muestra las diferencias estructurales entre estas familias de enzimas y su capacidad de reconocer sustratos tipo mucina.
- La estructura del complejo AM0627⁷¹⁻⁵⁰⁶E236A-P1-Zn²⁺ ha mostrado como en caso de que la enzima intente reconocer O-glicanos sialilados se produciría impedimento estérico con los aminoácidos del entorno, corroborando porqué esta enzima no es capaz de hidrolizar O-glicanos complejos conteniendo ácido siálico.

CAPÍTULO 6. REFERENCIAS

- Belzer, C., & De Vos, W. M. (2012). Microbes inside from diversity to function: The case of Akkermansia. *ISME Journal*, 6(8), 1449–1458. <https://doi.org/10.1038/ismej.2012.6>
- Birchough, G. M. H., Nystrom, E. E. L., Johansson, M. E. V., & Hansson, G. C. (2016). A sentinel goblet cell guards the colonic crypt by triggering Nlrp6-dependent Muc2 secretion. *Science*, 352(6293), 1535–1542. <https://doi.org/10.1126/science.aaf7419>
- Cerdà-Costa, N., & Gomis-Rüth, F. X. (2014). Architecture and function of metallopeptidase catalytic domains. In *Protein Science* (Vol. 23, Issue 2, pp. 123–144). Cold Spring Harbor Laboratory Press. <https://doi.org/10.1002/pro.2400>
- Coelho, H., Rivas, M. D. Las, Grosso, A. S., Diniz, A., Soares, C. O., Francisco, R. A., Dias, J. S., Compañon, I., Sun, L., Narimatsu, Y., Vakhrushev, S. Y., Clausen, H., Cabrita, E. J., Jiménez-Barbero, J., Corzana, F., Hurtado-Guerrero, R., & Marcelo, F. (2022). Atomic and Specificity Details of Mucin 1 O-Glycosylation Process by Multiple Polypeptide GalNAc-Transferase Isoforms Unveiled by NMR and Molecular Modeling. *JACS Au*, 2(3), 631–645. <https://doi.org/10.1021/jacsau.1c00529>
- Collado, M. C., Derrien, M., Isolauri, E., De Vos, W. M., & Salminen, S. (2007). Intestinal integrity and Akkermansia muciniphila, a mucin-

- degrading member of the intestinal microbiota present in infants, adults, and the elderly. *Applied and Environmental Microbiology*, 73(23), 7767–7770. <https://doi.org/10.1128/AEM.01477-07>
- 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>
- Derrien, M., Collado, M. C., Ben-Amor, K., Salminen, S., & De Vos, W. M. (2008). The mucin degrader *Akkermansia muciniphila* is an abundant resident of the human intestinal tract. *Applied and Environmental Microbiology*, 74(5), 1646–1648. <https://doi.org/10.1128/AEM.01226-07>
- Derrien, M., Vaughan, E. E., Plugge, C. M., & de Vos, W. M. (2004). *Akkermansia muciniphila* gen. nov., sp. nov., a human intestinal mucin-degrading bacterium. *International Journal of Systematic and Evolutionary Microbiology*, 54(5), 1469–1476. <https://doi.org/10.1099/ijs.0.02873-0>
- Desai, M. S., Seekatz, A. M., Koropatkin, N. M., Kamada, N., Hickey, C. A., Wolter, M., Pudlo, N. A., Kitamoto, S., Terrapon, N., Muller, A., Young, V. B., Henrissat, B., Wilmes, P., Stappenbeck, T. S., Núñez, G., & Martens, E. C. (2016). A Dietary Fiber-Deprived Gut Microbiota Degrades the Colonic Mucus Barrier and Enhances Pathogen Susceptibility. *Cell*, 167(5), 1339–1353.e21. <https://doi.org/10.1016/j.cell.2016.10.043>

- 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>
- Ensing, B., Laio, A., Parrinello, M., & Klein, M. L. (2005). A recipe for the computation of the free energy barrier and the lowest free energy path of concerted reactions. *Journal of Physical Chemistry B*, 109(14), 6676–6687. <https://doi.org/10.1021/jp045571i>
- Fernández-Blanco, J. A., Fakih, D., Arike, L., Rodríguez-Piñeiro, A. M., Martínez-Abad, B., Skansebo, E., Jackson, S., Root, J., Singh, D., McCrae, C., Evans, C. M., Åstrand, A., Ermund, A., & Hansson, G. C. (2018). Attached stratified mucus separates bacteria from the epithelial cells in COPD lungs. *JCI Insight*, 3(17). <https://doi.org/10.1172/jci.insight.120994>
- Frenkel, E. S., & Ribbeck, K. (2015). Salivary mucins protect surfaces from colonization by cariogenic bacteria. *Applied and Environmental Microbiology*, 81(1), 332–338. <https://doi.org/10.1128/AEM.02573-14>
- Frenkel, E. S., & Ribbeck, K. (2017). Salivary mucins promote the coexistence of competing oral bacterial species. *ISME Journal*, 11(5), 1286–1290. <https://doi.org/10.1038/ismej.2016.200>
- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., & Bairoch, A. (n.d.). *Protein Analysis Tools on the ExPASy Server* 571 571 From: *The Proteomics Protocols Handbook Protein Identification and Analysis Tools on the ExPASy Server*. <http://www.expasy.org/tools/>.

- Hansson, G. C. (2019). Mucus and mucins in diseases of the intestinal and respiratory tracts. *Journal of Internal Medicine*, 285(5), 479–490. <https://doi.org/10.1111/joim.12910>
- Hansson, G. C. (2020). Mucins and the Microbiome. In *Annual Review of Biochemistry* (Vol. 89, pp. 769–793). Annual Reviews Inc. <https://doi.org/10.1146/annurev-biochem-011520-105053>
- Holm, L. (2022). Dali server: structural unification of protein families. *Nucleic Acids Research*, 50(W1), W210–W215. <https://doi.org/10.1093/nar/gkac387>
- Humphreys, C. J. (2013). The significance of Bragg's law in electron diffraction and microscopy, and Bragg's second law. *Acta Crystallographica Section A: Foundations of Crystallography*, 69(1), 45–50. <https://doi.org/10.1107/S0108767312047587>
- Johansson, M. E. V., Ambort, D., Pelaseyed, T., Schütte, A., Gustafsson, J. K., Ermund, A., Subramani, D. B., Holmén-Larsson, J. M., Thomsson, K. A., Bergström, J. H., Van Der Post, S., Rodriguez-Piñeiro, A. M., Sjövall, H., Bäckström, M., & Hansson, G. C. (2011). Composition and functional role of the mucus layers in the intestine. In *Cellular and Molecular Life Sciences* (Vol. 68, Issue 22, pp. 3635–3641). <https://doi.org/10.1007/s00018-011-0822-3>
- Johansson, M. E. V., Holmén Larsson, J. M., & Hansson, G. C. (2011). The two mucus layers of colon are organized by the MUC2 mucin, whereas the outer layer is a legislator of host-microbial interactions. *Proceedings*

of the National Academy of Sciences of the United States of America,
108(SUPPL. 1), 4659–4665. <https://doi.org/10.1073/pnas.1006451107>

Johansson, M. E. V., Phillipson, M., Petersson, J., Velcich, A., Holm, L.,
Hansson, G. C., & Greenberg, E. P. (2008). *The inner of the two Muc2
mucin-dependent mucus layers in colon is devoid of bacteria*.
www.pnas.org/cgi/doi/10.1073/pnas.0803124105

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O.,
Tunyasuvunakool, K., Bates, R., Žídek, 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>

Kabsch, W. (2010). *XDS*. *Acta Crystallographica Section D Biological
Crystallography*, 66(2), 125–132.
<https://doi.org/10.1107/S09074444909047337>

Kavanaugh, N. L., Zhang, A. Q., Nobile, C. J., Johnson, A. D., & Ribbeck, K.
(2014). Mucins suppress virulence traits of *Candida albicans*. *MBio*, 5(6).
<https://doi.org/10.1128/mBio.01911-14>

Kirschner, K. N., Yongye, A. B., Tschampel, S. M., González-Outeiriño, J.,
Daniels, C. R., Foley, B. L., & Woods, R. J. (2008). GLYCAM06: A
generalizable biomolecular force field. carbohydrates. *Journal of
Computational Chemistry*, 29(4), 622–655.
<https://doi.org/10.1002/jcc.20820>

- Konstantinidi, A., Nason, R., Caval, T., Sun, L., Sørensen, D. M., Furukawa, S., Ye, Z., Vincentelli, R., Narimatsu, Y., Vakhrushev, S. Y., & Clausen, H. (2022a). Exploring the glycosylation of mucins by use of O-glycodomain reporters recombinantly expressed in glycoengineered HEK293 cells. *Journal of Biological Chemistry*, 298(4). <https://doi.org/10.1016/j.jbc.2022.101784>
- Konstantinidi, A., Nason, R., Caval, T., Sun, L., Sørensen, D. M., Furukawa, S., Ye, Z., Vincentelli, R., Narimatsu, Y., Vakhrushev, S. Y., & Clausen, H. (2022b). Exploring the glycosylation of mucins by use of O-glycodomain reporters recombinantly expressed in glycoengineered HEK293 cells. *Journal of Biological Chemistry*, 298(4). <https://doi.org/10.1016/j.jbc.2022.101784>
- Kufe, D. W. (2009). Mucins in cancer: Function, prognosis and therapy. In *Nature Reviews Cancer* (Vol. 9, Issue 12, pp. 874–885). <https://doi.org/10.1038/nrc2761>
- Laio, A., & Parrinello, M. (n.d.). *Escaping free-energy minima*. www.pnas.org/cgi/doi/10.1073/pnas.202427399
- Laitaoja, M., Valjakka, J., & Jänis, J. (2013). Zinc coordination spheres in protein structures. *Inorganic Chemistry*, 52(19), 10983–10991. <https://doi.org/10.1021/ic401072d>
- Laville, E., Perrier, J., Bejar, N., Maresca, M., Esque, J., Tauzin, A. S., Bouhajja, E., Leclerc, M., Drula, E., Henrissat, B., Berdah, S., Di Pasquale, E., Robe, P., & Potocki-Veronese, G. (2019). Investigating host

microbiota relationships through functional metagenomics. *Frontiers in Microbiology*, 10(JUN). <https://doi.org/10.3389/fmicb.2019.01286>

Linden, S. K., Sutton, P., Karlsson, N. G., Korolik, V., & McGuckin, M. A. (2008). Mucins in the mucosal barrier to infection. In *Mucosal Immunology* (Vol. 1, Issue 3, pp. 183–197). <https://doi.org/10.1038/mi.2008.5>

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>

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., & Read, R. J. (2007). Phaser crystallographic software. *Journal of Applied Crystallography*, 40(4), 658–674. <https://doi.org/10.1107/S0021889807021206>

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>

Nath, S., & Mukherjee, P. (2014). MUC1: A multifaceted oncoprotein with a key role in cancer progression. In *Trends in Molecular Medicine* (Vol. 20,

Issue 6, pp. 332–342). Elsevier Ltd.
<https://doi.org/10.1016/j.molmed.2014.02.007>

Neelamegham, S., Aoki-Kinoshita, K., Bolton, E., Frank, M., Lisacek, F., Lütteke, 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>

Noach, I., Ficko-Blean, E., Pluvinage, B., Stuart, C., Jenkins, M. L., Brochu, D., Buenbrazo, N., Wakarchuk, W., Burke, J. E., Gilbert, M., & Boraston, A. B. (2017a). Recognition of protein-linked glycans as a determinant of peptidase activity. *Proceedings of the National Academy of Sciences of the United States of America*, 114(5), E679–E688.
<https://doi.org/10.1073/pnas.1615141114>

Noach, I., Ficko-Blean, E., Pluvinage, B., Stuart, C., Jenkins, M. L., Brochu, D., Buenbrazo, N., Wakarchuk, W., Burke, J. E., Gilbert, M., & Boraston, A. B. (2017b). Recognition of protein-linked glycans as a determinant of peptidase activity. *Proceedings of the National Academy of Sciences of the United States of America*, 114(5), E679–E688.
<https://doi.org/10.1073/pnas.1615141114>

Paone, P., & Cani, P. D. (2020). Mucus barrier, mucins and gut microbiota: The expected slimy partners? In *Gut* (Vol. 69, Issue 12, pp. 2232–2243). BMJ Publishing Group. <https://doi.org/10.1136/gutjnl-2020-322260>

Pedram, K., Laqtom, N. N., Shon, D. J., Spiezio, A. Di, Riley, N. M., Saftig, P., Abu-Remaileh, M., & Bertozzi, C. R. (2022). *Lysosomal cathepsin D*

mediates endogenous mucin glycodomain catabolism in mammals.
<https://doi.org/10.1073/pnas>

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13), 1605–1612. <https://doi.org/10.1002/jcc.20084>

Plattner, C., Höfener, M., & Sewald, N. (2011). One-pot azidochlorination of glycals. *Organic Letters*, 13(4), 545–547. <https://doi.org/10.1021/ol102750h>

Pluvinage, B., Ficko-Blean, E., Noach, I., Stuart, C., Thompson, N., McClure, H., Buenbrazo, N., Wakarchuk, W., & Boraston, A. B. (n.d.). *Architecturally complex O-glycopeptidases are customized for mucin recognition and hydrolysis.* <https://doi.org/10.1073/pnas.2019220118/-/DCSupplemental>

Raba, G., & Luis, A. S. (2023). Mucin utilization by gut microbiota: recent advances on characterization of key enzymes. *Essays in Biochemistry*, 67(3), 345–353. <https://doi.org/10.1042/EBC20220121>

Rawlings, N. D., Barrett, A. J., & Finn, R. (2016). Twenty years of the MEROPS database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Research*, 44(D1), D343–D350. <https://doi.org/10.1093/nar/gkv1118>

Rawlings, N. D., Barrett, A. J., Thomas, P. D., Huang, X., Bateman, A., & Finn, R. D. (2018). The MEROPS database of proteolytic enzymes, their

substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Research*, 46(D1), D624–D632. <https://doi.org/10.1093/nar/gkx1134>

Salomon-Ferrer, R., Götz, A. W., Poole, D., Le Grand, S., & Walker, R. C. (2013). Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh ewald. *Journal of Chemical Theory and Computation*, 9(9), 3878–3888. <https://doi.org/10.1021/ct400314y>

Schjoldager, K. T., Narimatsu, Y., Joshi, H. J., & Clausen, H. (2020). Global view of human protein glycosylation pathways and functions. In *Nature Reviews Molecular Cell Biology* (Vol. 21, Issue 12, pp. 729–749). Nature Research. <https://doi.org/10.1038/s41580-020-00294-x>

Shon, D. J., Fernandez, D., Riley, N. M., Ferracane, M. J., & Bertozzi, C. R. (2022). Structure-guided mutagenesis of a mucin-selective metalloprotease from *Akkermansia muciniphila* alters substrate preferences. *Journal of Biological Chemistry*, 298(5). <https://doi.org/10.1016/j.jbc.2022.101917>

Shon, D. J., Kuo, A., Ferracane, M. J., & Malaker, S. A. (2021). Classification, structural biology, and applications of mucin domain-targeting proteases. In *Biochemical Journal* (Vol. 478, Issue 8, pp. 1585–1603). Portland Press Ltd. <https://doi.org/10.1042/BCJ20200607>

Shon, D. J., Malaker, S. A., Pedram, K., Yang, E., Krishnan, V., Dorigo, O., & Bertozzi, C. R. (2020). *An enzymatic toolkit for selective proteolysis*,

detection, and visualization of mucin-domain glycoproteins.
<https://doi.org/10.1073/pnas.2012196117/-/DCSupplemental>

Stenfors Arnesen, L. P., Fagerlund, A., & Granum, P. E. (2008). From soil to gut: *Bacillus cereus* and its food poisoning toxins. In *FEMS Microbiology Reviews* (Vol. 32, Issue 4, pp. 579–606). <https://doi.org/10.1111/j.1574-6976.2008.00112.x>

SUPRUNIUK, K., & RADZIEJEWSKA, I. (2021). MUC1 is an oncoprotein with a significant role in apoptosis (Review). In *International Journal of Oncology* (Vol. 59, Issue 3). Spandidos Publications. <https://doi.org/10.3892/ijo.2021.5248>

Svensson, F., Lang, T., Johansson, M. E. V., & Hansson, G. C. (2018). The central exons of the human MUC2 and MUC6 mucins are highly repetitive and variable in sequence between individuals. *Scientific Reports*, 8(1). <https://doi.org/10.1038/s41598-018-35499-w>

Taleb, V., Liao, Q., Narimatsu, Y., García-García, A., Compañón, I., Borges, R. J., González-Ramírez, A. M., Corzana, F., Clausen, H., Rovira, C., & Hurtado-Guerrero, R. (2022). Structural and mechanistic insights into the cleavage of clustered O-glycan patches-containing glycoproteins by mucinases of the human gut. *Nature Communications*, 13(1). <https://doi.org/10.1038/s41467-022-32021-9>

Trastoy, B., Naegeli, A., Anso, I., Sjögren, J., & Guerin, M. E. (2020). Structural basis of mammalian mucin processing by the human gut O-glycopeptidase OgpA from *Akkermansia muciniphila*. *Nature Communications*, 11(1). <https://doi.org/10.1038/s41467-020-18696-y>

- Usón, I., & Sheldrick, G. M. (2018). An introduction to experimental phasing of macromolecules illustrated by SHELX; New autotracing features. *Acta Crystallographica Section D: Structural Biology*, *74*, 106–116. <https://doi.org/10.1107/S2059798317015121>
- van Kempen, M., Kim, S. S., Tumescheit, C., Mirdita, M., Lee, J., Gilchrist, C. L. M., Söding, J., & Steinegger, M. (2023). Fast and accurate protein structure search with Foldseek. *Nature Biotechnology*. <https://doi.org/10.1038/s41587-023-01773-0>
- Van Putten, J. P. M., & Strijbis, K. (2017). Transmembrane Mucins: Signaling Receptors at the Intersection of Inflammation and Cancer. In *Journal of Innate Immunity* (Vol. 9, Issue 3, pp. 281–299). S. Karger AG. <https://doi.org/10.1159/000453594>
- Wagner, C. E., Wheeler, K. M., & Ribbeck, K. (2018). Mucins and Their Role in Shaping the Functions of Mucus Barriers. *Annu. Rev. Cell Dev. Biol*, *34*, 189–215. <https://doi.org/10.1146/annurev-cellbio-100617>
- Walther, G., Wagner, L., & Kurzai, O. (2020). Outbreaks of Mucorales and the Species Involved. In *Mycopathologia* (Vol. 185, Issue 5, pp. 765–781). Springer Science and Business Media B.V. <https://doi.org/10.1007/s11046-019-00403-1>
- 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>

Wlodawer, A., Minor, W., Dauter, Z., & Jaskolski, M. (2008). Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. In *FEBS Journal* (Vol. 275, Issue 1, pp. 1–21). <https://doi.org/10.1111/j.1742-4658.2007.06178.x>

Wu, S., Zhang, C., Xu, D., & Guo, H. (2010). Catalysis of carboxypeptidase A: Promoted-water versus nucleophilic pathways. *Journal of Physical Chemistry B*, 114(28), 9259–9267. <https://doi.org/10.1021/jp101448j>

Xu, D., & Guo, H. (2009). Quantum mechanical/molecular mechanical and density functional theory studies of a prototypical zinc peptidase (carboxypeptidase A) suggest a general acid-general base mechanism. *Journal of the American Chemical Society*, 131(28), 9780–9788. <https://doi.org/10.1021/ja9027988>

CAPÍTULO 7. ABREVIATURAS

2xTY. 2x triptona de extracto de levadura

3Tn. Tri Tn O-glicopéptido

4Tn. Tetra Tn O-glicopéptido

ADN. Ácido desoxirribonucleico

Antígeno ST. Sia α 2-3Gal β 1-3GalNAc α Ser/Thr

Antígeno STn. Sia α 2-6GalNAc α Ser/Thr

Antígeno T o *Core* 1. Gal β 1-3GalNAc α Ser/Thr

Antígeno Tn. GalNAc α Ser/Thr

APC. Aloficocianina

BSA. Albumina de suero bovino

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

CBM. Dominio de unión a carbohidratos

CHO. Células de ovario de hámster (*Chinese hamster ovary*)

Core 2. GlcNAc β 1-6(Gal β 1-3) GalNAc α Ser/Thr

Core 3. GlcNAc β 1-3GalNAc α Ser/Thr

Core 4. GlcNAc β 1-6(GlcNAc β 1-3) GalNAc α Ser/Thr

CV. Coordinada de reacción o *collective variable*

CysD. Dominios ricos en cisteína

C α . Carbono alfa

DFT. Teoría del funcional de la densidad

DMSO. Dimetilsulfóxido

DTT. Ditioneitol

E. coli. Escherichia coli

EDTA. Ácido etilendiaminotetraacético

FN3. Dominio tipo fibronectina III

GalNAc. N-Acetil galactosamina

GlcNAc. N-Acetil glucosamina

HEK293. Células embrionarias de riñón humano 293 (*Human embryonic kidney*)

IMAC. Cromatografía de afinidad por metales inmovilizados

IPTG. Isopropil- β -D-1-tiogalactopiranosido

LB. Caldo de Lisogenia (*Lysogeny broth*)

MAD/SAD. Dispersión anómala de longitud de onda múltiple o simple

MBP. Proteína de unión a maltosa

MC. Complejo de Michaelis

McMuc. Mucinasas de *Mucor circinelloides* 1006PhL

MD. Dinámica molecular

MFI. Media de intensidad de fluorescencia

MIR. Reemplazo isomorfo múltiple

MR. Remplazamiento molecular

MUC. Mucina

P1. TEAQT**T**PPPA

PBS. Tampón fosfato salino

PMSF. Fluoruro de fenilmetilsulfonilo

PTS. Secuencias con una alta proporción de prolinas, treoninas y serinas

RMSD. Raíz del error cuadrático medio

SDS. Dodecilsulfato de sodio

SDS-PAGE. Electroforesis desnaturizante en geles de acrilamida con SDS

Sia. Ácido siálico

SPPS. Síntesis de péptidos en fase sólida

TACA. Antígeno de carbohidratos asociado a cáncer

TEV. Virus del mosaico del tabaco. *Tobacco Etch Virus*

TR. Secuencias repetidas en tándem

TRIS. Tris(hidroxietil)aminometano

TS1. Primer estado de transición

TS2. Segundo estado de transición

UDP. Uridina difosfato

VWC. Dominios tipo C del factor von Willebrand

VWD. Dominios tipo D del factor von Willebrand

WT. *Wild type* o silvestre

CAPÍTULO 8. APENDICES

8.1. MEDIOS DE CULTIVO Y RECETAS

LB/Agar

- 1% (w/v) triptona
- 0,5% (w/v) extracto de levadura
- 1% (w/v) NaCl
- 1,5% (w/v) agar

LB

- 1% (w/v) triptona
- 0,5% (w/v) extracto de levadura
- 1% (w/v) NaCl

2XTY

- 1,6% (w/v) triptona
- 1% (w/v) extracto de levadura
- 0,5% (w/v) NaCl

Tampón de carga de proteínas laemmli

- 24% (v/v) 1 M Tris pH 6,8
- 8% (w/v) SDS

- 40% (v/v) glicerol
- 10% (v/v) β -mercaptoetanol
- 0,1% (w/v) azul de bromofenol

Ampicilina 100 mg/ml (10 ml)

- 1 g ampicilina
- 10 ml H₂O MQ

Cocktail de inhibidores de proteasas

- 20 mM PMSF
- 1 mM Leupeptina
- 200 mM Benzamidina

8.2 PROPIEDADES DE LAS PROTEÍNAS RECOMBINANTES

Se recogen a continuación las propiedades de las proteínas recombinantes *wildtype* con las que se ha trabajado durante esta tesis. Para los mutantes que derivan de algunas de estas proteínas, se indica únicamente sus pesos moleculares y coeficientes de extinción molar en las **Tablas 1** y **3** en los **apartados 3.1.3.6.** y **3.2.3.** respectivamente ya que sus propiedades son similares a las de la proteína *wildtype*.

8.2.1 AM0627²¹⁻⁵⁰⁶

ANTPEHIGNDLKLFKDSSCTSLKPDVKNTSAFQSDAMKELATKILAG
HYKPDYLYAEYRALPSRQTGKNLRIGDGFSSKYDNMTGVYLEKGRH
VVLVGKTEGQEISLLLPLNLMRKPAEGVQPTKDPNGWGLHKKQIPLKE
GINIIDVETPANAYISYFTEDAGKAPKIPVHFVTGKANGYFDTTRGDT
NKDWVRLLDQAVSPIMDARGKYIQVAYPVEFLKKFTKDRGTELINA
YDKLIGIQYQLMGLDKYGKIPENRVLARVNFNYMFRDGDGVAYLG
NDGTMRMVTDPENVLKGDACWGFSHEVGHVMQMRPMTWGGMTE
VSNNIFSLQAAAKTGNESRLKRQGSYDKARKEIIEGEIAYLQSKDVFN
KLVPLWQLHLYFTKNGHPDFYDPVMEYLRNAGNYGGNDTVKYQF
EFVKACCDVTKTDLTDFFEKWGFFKPGKFHIGDYAQYDFNVTPEMV
EETKKWIAGKGYPKPETDITELSE

Número de aminoácidos: 486

Peso molecular: 55107,62 Da

pI teórico: 7,28

Número de residuos cargados negativamente (Asp + Glu): 63

Número de residuos cargados positivamente (Arg + Lys): 63

Coefficiente de extinción molar: 78730 M⁻¹·cm⁻¹ (a 280 nm en agua, asumiendo que todas las cisteínas están reducidas)

8.2.2. AM0627⁷¹⁻⁵⁰⁶

PDYLYAEYRALPSRQTGKNLRIGDGFSKYDNMTGVYLEKGRHVVL
VGKTEGQEISLLLPNLMRKAEGVQPTKDPNGWGLHKKQIPLKEGINI
IDVETPANAYISYFTEDAGKAPKIPVHFVTGKANGYFDTTTRGDTNKD
WVRLLDQAVSPIMDARGKYIQVAYPVEFLKKFTKDRGTELINAYDK
LIGIQYQLMGLDKYGKIPENRVLARVNFNYYMFRDGDGVAYLGNDG
TMRMVTDPENVLKGDACWGFSHEVGHVMQMRPMTWGGMTEVSN
NIFSLQAAAKTGNESRLKRQGSYDKARKEIIEGEIAYLQSKDVFNKL
PLWQLHLYFTKNGHPDFYPDVMEYLRNNAAGNYGGNDTVKYQFEFV
KACCDVTKTDLTDFFEKWGFFKPGKFHIGDYAQYDFNVTPEMVEET
KKWIAGKGYPKPETDITELSE

Número de aminoácidos: 436

Peso molecular: 49660,44 Da

pI teórico: 6,85

Número de residuos cargados negativamente (Asp + Glu): 57

Número de residuos cargados positivamente (Arg + Lys): 56

Coefficiente de extinción molar: 77240 M⁻¹·cm⁻¹ (a 280 nm en agua, asumiendo que todas las cisteínas están reducidas)

8.3. LOS AMINOÁCIDOS

En la **Figura 74** se muestran todos los aminoácidos clasificados según sus propiedades como pueden ser aromaticidad, comportamiento ácido-base o presencia de azufre.

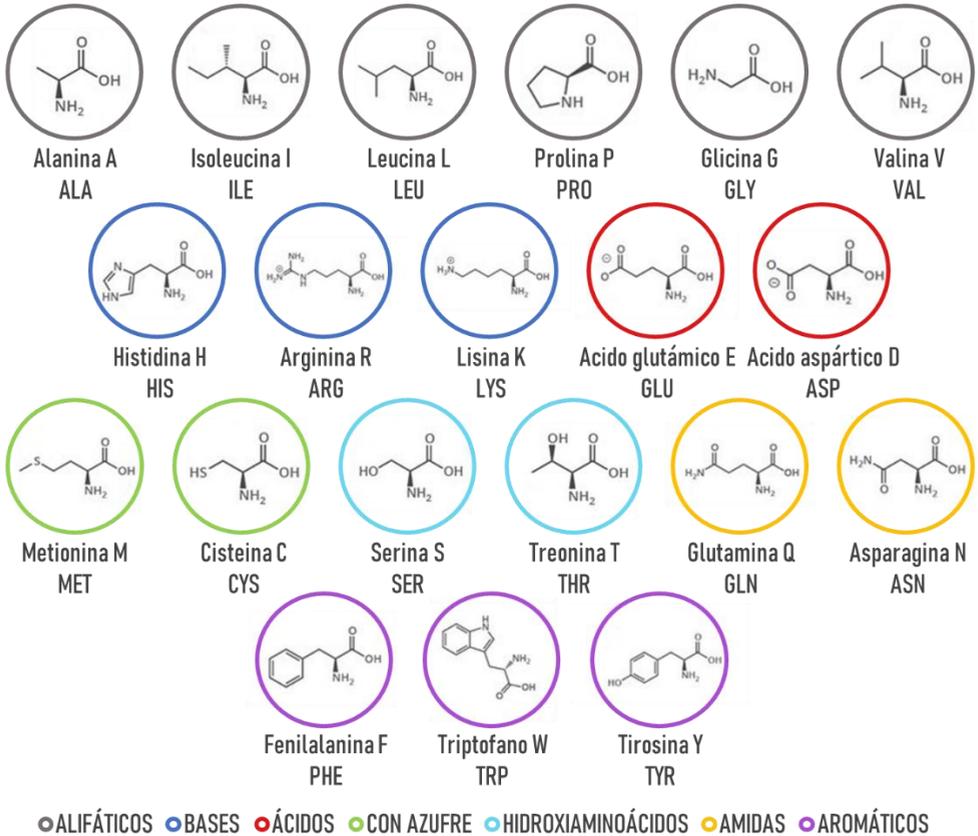


Figura 46. Clasificación de los aminoácidos según si son alifáticos (gris), bases (azul), ácidos (rojo), contienen azufre (verde), hidroxiaminoácidos (cian), amidas (amarillo) y aromáticos (morado).