

# GLICOPROTEÍNAS BIOMARCADORAS DE CÁNCER: UN RETO ANALÍTICO ACTUAL

Estela Giménez, Montserrat Mancera-Arteu, Laura Pont, José Barbosa, Fernando Benavente y Victoria Sanz-Nebot

Departamento de Ingeniería Química y Química Analítica, Instituto de Investigación en Nutrición y Seguridad Alimentaria (INSA·UB), Universidad de Barcelona, Martí i Franquès 1-11. 08028 Barcelona

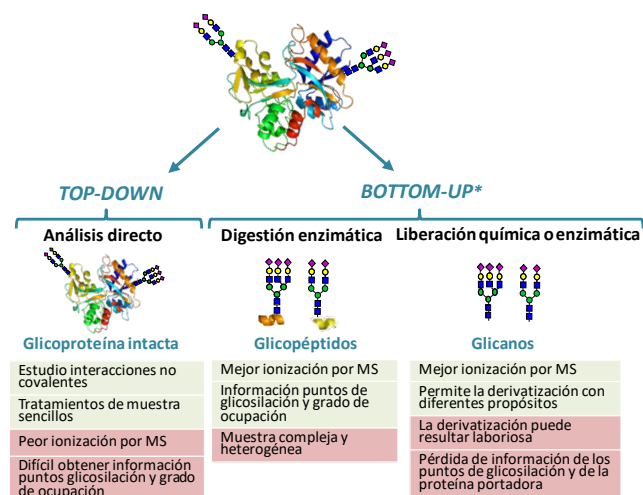
E-mail: estelagimenez@ub.edu, www.ub.edu/bioanalisi

## 1. Introducción

Las alteraciones en la glicosilación de las proteínas pueden ser la causa de enfermedades graves, como es el caso de los desórdenes congénitos de la glicosilación (CDGs) o pueden aparecer en respuesta a la presencia de un proceso patológico, como ocurre con el cáncer y los procesos inflamatorios. Por ello, las glicoproteínas se utilizan ampliamente en el diagnóstico de diversas patologías [1, 2]. Por ejemplo, el antígeno carbohidratado 19-9 (CA 19-9), el antígeno carcinoembrionario humano (CEA) o el antígeno prostático específico (PSA) se emplean en la actualidad como biomarcadores para el diagnóstico del cáncer de páncreas, colon o próstata, respectivamente. Sin embargo, sólo se mide el aumento de sus niveles por encima de unos valores establecidos, con lo que en muchos casos estos biomarcadores no son totalmente específicos ya que otras enfermedades pueden presentar niveles anormalmente elevados de estas glicoproteínas. Por este motivo, el reto analítico actual consiste en mejorar la selectividad y la sensibilidad de estos biomarcadores, llevando a cabo una caracterización exhaustiva de los puntos de glicosilación y de la composición y estructura de los carbohidratos, comúnmente llamados glicanos.

Las técnicas microseparativas como la electroforesis capilar y la cromatografía de líquidos capilar, ambas acopladas a la espectrometría de masas (CE-MS y CapLC-MS, respectivamente), son herramientas imprescindibles para abordar el análisis de glicoproteínas. Estas técnicas ofrecen un gran poder de resolución y proporcionan información estructural para identificar de manera inequívoca y caracterizar las diferentes proteoformas, que en este caso se denominan glicofomas. El estudio de la glicosilación de una proteína puede llevarse a cabo a tres niveles mediante dos estrategias de análisis. El análisis directo de la glicoproteína intacta (*top-down analysis*) es más rápido ya que no precisa de extensos tratamientos de muestra. Sin embargo, continúa siendo un gran reto analítico ya que las glicoproteínas son moléculas de masa molecular elevada, difíciles de ionizar y de detectar con la instrumentación actual [3]. La otra estrategia analítica se basa en el uso de enzimas o reactivos químicos para obtener especies glicosiladas de menor masa molecular, como son los glicopéptidos y los glicanos (*bottom-up*

*analysis*). En este caso, el tratamiento de muestra es más complejo, pero se mejora en sensibilidad y se obtiene información detallada sobre la estructura de los glicanos y de los puntos de glicosilación. En la Figura 1 se muestra un esquema de las tres aproximaciones con sus ventajas e inconvenientes principales.



**Figura 1:** Estrategias de análisis de glicoproteínas. (\*) El análisis bottom-up puede referirse como middle-down cuando la enzima empleada permite obtener fragmentos peptídicos de masa molecular elevada ( $3000 < M_r < 10000$ ) [4].

Por otro lado, el análisis de glicoproteínas en muestras biológicas requiere del desarrollo de tratamientos de muestra eficaces que permitan purificar y preconcentrar la glicoproteína de interés, aumentando la selectividad y sensibilidad del proceso analítico, dado que muchas de ellas se encuentran a baja concentración en los fluidos biológicos y en presencia de abundantes proteínas no glicosiladas que dificultan la detección por MS. Para ello, en glicoproteómica se utiliza la cromatografía de afinidad con anticuerpos, lectinas, grupos boronato y algunos óxidos metálicos como el óxido de titanio [5]. No obstante, el uso de anticuerpos suele ser más adecuado para el análisis de una glicoproteína en concreto (*targeted analysis*) puesto que ofrecen una elevada selectividad hacia su antígeno.

Finalmente, como en la mayoría de ciencias ómicas, la gran cantidad de datos generados mediante las estrategias analíticas empleadas dificulta llegar a conclusiones fiables a partir de la exploración de los

resultados sin herramientas matemáticas de análisis de datos. En el estudio de glicoproteínas biomarcadoras, el conjunto de datos obtenido suele ser muy grande y complejo a causa de la diversidad inherente de los glicanos. En este sentido, los métodos de análisis multivariante, como el análisis de componentes principales, PCA, y el análisis discriminante por mínimos cuadrados parciales, PLS-DA, se han empleado con éxito para investigar los glicanos y glicopéptidos que están significativamente alterados en presencia de ciertas patologías [6–8]. No obstante, la mejora de los métodos de análisis de datos es otro de los retos que existen actualmente en el campo de la glicoproteómica para identificar nuevos biomarcadores de manera sencilla y fiable [9].

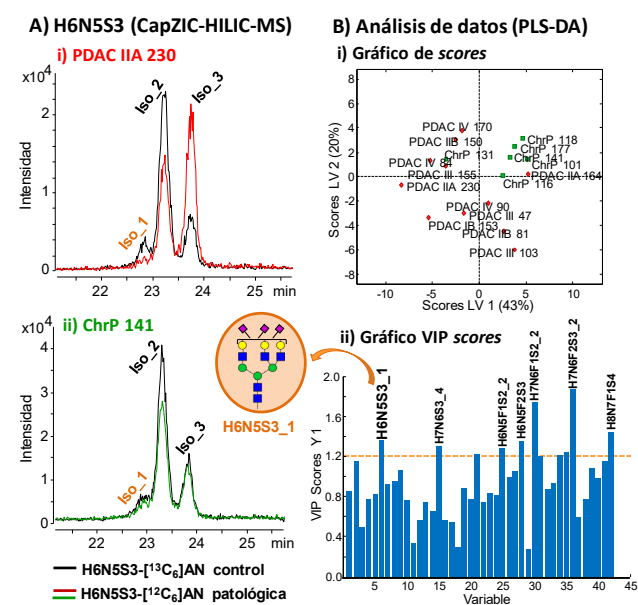
En este artículo se describen dos metodologías analíticas *bottom-up* desarrolladas por nuestro grupo de investigación para el estudio de glicoproteínas biomarcadoras de cáncer. La primera está centrada en una estrategia de análisis de glicanos mediante CapLC-MS para la caracterización de la alfa glicoproteína ácida humana (hAGP), como biomarcadora potencial de cáncer de páncreas [7]. La segunda presenta una metodología de análisis de glicopéptidos de la glicoproteína CEA por CE-MS/MS para estudiar su papel en la diferenciación de diferentes tipos de cáncer y estadios de la enfermedad [10].

## 2. Análisis de glicanos para el diagnóstico del cáncer de páncreas

Aunque el análisis de los glicanos implica la pérdida de información sobre los puntos de glicosilación, esta estrategia tiene mayor potencial para la caracterización a nivel de isómeros, lo que puede resultar importante porque en algunas patologías las alteraciones en la glicosilación se producen a este nivel [2]. Los isómeros más comunes que presentan los glicanos se deben, principalmente, a diferencias en el tipo de enlace de los ácidos siálicos (SiA), que pueden ser  $\alpha$ 2-3 o  $\alpha$ 2-6. En el caso de los N-glicanos, éstos se liberan de la cadena polipeptídica mediante la acción de una glucosidasa, PNGasa F, que rompe el enlace con la asparagina al que están unidos. Como estos analitos son moléculas muy hidrofílicas, no se ionizan bien por electrospray, por lo que es necesario derivatizarlos para mejorar su respuesta analítica. Asimismo, la derivatización permite modular su separación mediante CE o CapLC y cuantificarlos correctamente al poder introducir compuestos marcados isotópicamente [11, 12]. En este sentido, nuestro grupo de investigación ha desarrollado una metodología de marcaje isotópico de los glicanos con anilina [ $^{12}\text{C}_6$ ]/[ $^{13}\text{C}_6$ ] (GRIL) y de análisis de los glicanos derivatizados por cromatografía de líquidos capilar de interacción hidrofílica zwitteriónica acoplada a la espectrometría de masas (CapZIC-HILIC-MS) [13]. Esta metodología analítica permite separar los diferentes isómeros de un mismo glicano, cuantificándolos de manera fiable y rigurosa, y se ha aplicado con éxito al estudio de la hAGP en muestras de

suero de pacientes con cáncer de páncreas (PDAC) y pancreatitis crónica (ChrP) [7].

En primer lugar, la hAGP se purifica a partir de las muestras de suero por cromatografía de inmunofinidad y se cuantifica la cantidad de proteína mediante espectrofotometría UV-Vis. Después de digerir las muestras con PNGasa F, los glicanos se purifican por extracción en fase sólida con cartuchos de carbono grafito y se derivatizan con anilina (AN) aplicando la estrategia GRIL para cuantificarlos de manera relativa. Con este propósito, se preparan mezclas equimolares de los glicanos de la hAGP de cada muestra patológica (derivatizados con [ $^{12}\text{C}_6$ ]AN) y de los glicanos de la hAGP de una mezcla de muestras control de individuos sanos (derivatizados con [ $^{13}\text{C}_6$ ]AN). Finalmente, se analizan por CapZIC-HILIC-MS. Dada la gran cantidad de muestras y número de isómeros identificados, se aplican métodos de análisis multivariante de datos, para interpretar los resultados a partir de las áreas relativas de todos los isómeros detectados (área del glicano en la muestra patológica dividida entre el área del glicano en el *pool* de muestras control). En primer lugar, se realiza PCA para explorar los datos, detectar valores anómalos y estudiar la existencia de tendencias o grupos entre las diferentes muestras. A continuación, con el fin de mejorar la separación entre grupos e identificar qué isómeros pueden diferenciar ambas patologías (PDAC y ChrP), se lleva a cabo un PLS-DA. En la Figura 2 se muestran algunos de los resultados obtenidos en este estudio.



**Figura 2:** A) EICs del glicano H6N5S3 de la hAGP por CapZIC-HILIC-MS en dos muestras patológicas. B) Gráfico de scores (i) y de VIP scores (ii) del modelo de PLS-DA aplicado a las relaciones de áreas obtenidas para cada isómero de glicano de la hAGP en todas las muestras de PDAC y ChrP analizadas.

En el gráfico de *scores* (Figura 2B-i), se observa como los dos tipos de muestras patológicas (18 en total), a excepción de una muestra de ChrP y una de PDAC, son

agrupadas correctamente por el modelo de PLS-DA. Además, el gráfico de VIP scores (Figura 2B-ii) permite cuantificar la influencia de cada isómero de glicano en la separación de los dos grupos. Siete isómeros de un total de 42, cuya caracterización previa demostró que contenían mayoritariamente SiA  $\alpha$ 2-6 [14], son los que presentan mayor influencia en la diferenciación entre PDAC y ChrP (valores de VIP scores > 1,2). A modo de ejemplo, en la Figura 2A se muestran los cromatogramas de iones extraídos (EICs) del glicano H6N5S3 en las muestras patológicas PDAC IIA 230 (EIC en rojo) y ChrP 141 (EIC en verde), y en el pool de controles (EICs en negro). Como se puede observar, aunque los 3 isómeros del glicano H6N5S3 presentan alteraciones en las muestras patológicas con respecto al control, el modelo de PLS-DA sólo apunta al isómero 1 (H6N5S3\_1, iso\_1) como potencial biomarcador ya que se encuentra significativamente alterado en todas las muestras, disminuyendo su expresión en muestras de PDAC respecto a las de ChrP y a los controles sanos (Figura 2A).

Así, esta estrategia combinada permite, por un lado, comparar de manera rigurosa los perfiles de los glicanos en muestras control y patológicas y, por otro, reducir la complejidad de la interpretación de los datos obtenidos. Tras la identificación correcta de los isómeros de los glicanos de la hAGP, se proponen aquellos que podrían ser biomarcadores de PDAC, diferenciando este tipo de cáncer de la ChrP, incluso en estadios iniciales de la enfermedad.

### 3. Análisis de glicopéptidos para la diferenciación de cáncer de colon y metástasis de hígado

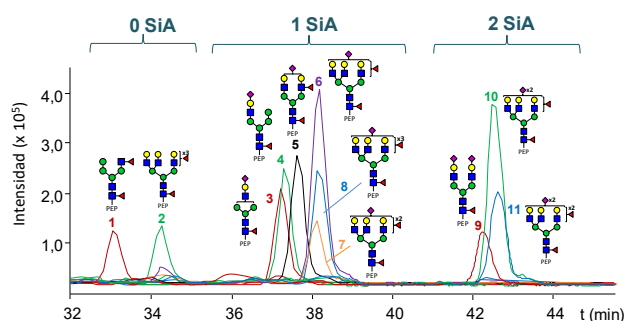
El análisis de los glicopéptidos obtenidos por digestión enzimática de las glicoproteínas presenta algunas ventajas con respecto al análisis de los glicanos. Además de proporcionar información sobre la composición y estructura de los glicanos, se obtiene información sobre los puntos de glicosilación y de la proteína portadora, al conservar parte de la secuencia polipeptídica [15]. Como contrapartida, tras la digestión se obtiene una mezcla compleja y heterogénea de péptidos y glicopéptidos, siendo imprescindible llevar a cabo una separación previa a la detección por MS, puesto que los péptidos presentan una mejor predisposición a la ionización e inhiben la de los glicopéptidos, disminuyendo la sensibilidad de los análisis. La CE presenta especial interés en el análisis de glicopéptidos debido a que su mecanismo de separación basado en las diferencias en la relación carga/radio hidrodinámico de los iones de los analitos favorece la separación de las glicofórmulas, que difieren en su carga y en el contenido y composición de los glicanos. En nuestro grupo de investigación, se han establecido métodos de CE-MS utilizando una interfase con líquido auxiliar (*sheath flow*) para detectar glicopéptidos de glicoproteínas de interés farmacológico y biomédico [16, 17]. Como alternativa a este tipo de interfase, que diluye los compuestos de interés a la salida del capilar de separación, en el estudio que se

presenta a continuación, se emplea una interfase sin líquido auxiliar (*sheathless*) para la mejora de la sensibilidad. La metodología analítica establecida se ha aplicado a la identificación y caracterización de los N-glicopéptidos de CEA por CE-MS/MS [10].

En primer lugar, las muestras comerciales de CEA procedentes de pacientes con cáncer de colon y con metástasis de hígado se digieren con tripsina, previa reducción y alquilación de la proteína. A continuación, los digestos se analizan por CE-MS/MS adquiriendo en modo “data dependent”, donde se fragmentan todos los iones con intensidades por encima de un umbral. Para encontrar la región del electroferograma donde migran los N-glicopéptidos, se extrae el electroferograma del ion de m/z 366,14 (ión oxonio), que es característico de todos los N-glicopéptidos. Finalmente, para identificar de manera inequívoca y caracterizar estructuralmente las glicofórmulas de los N-glicopéptidos detectados, se realiza una inspección visual de los espectros de MS/MS para asignar los diferentes fragmentos específicos de cada glicofórmula.

En los dos tipos de muestras analizadas, se logra identificar un total de 489 glicofórmulas correspondientes a 8 puntos de N-glicosilación, de los 28 puntos descritos en la bibliografía. En la Figura 3A se muestran, a modo de ejemplo, los electroferogramas de iones extraídos (EIEs) para las 11 glicofórmulas identificadas del glicopéptido N650 en una muestra de CEA procedente de metástasis de hígado.

A) Separación de las glicofórmulas del glicopéptido N650 en una muestra de metástasis de hígado por CE-MS/MS



B) Gráfico de scores del PLS-DA

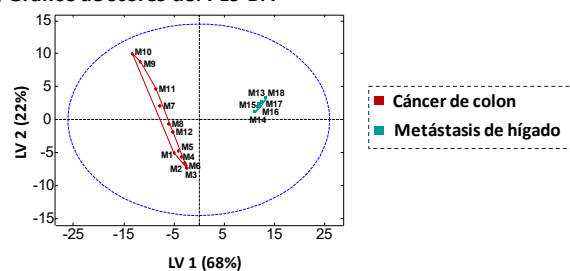


Figura 3: A) EIEs (full scan) de las glicofórmulas detectadas del glicopéptido N650 de CEA en una muestra de metástasis de hígado. B) Gráfico de scores del modelo de PLS-DA obtenido a partir de las áreas normalizadas de las glicofórmulas identificadas (de 8 N-glicopéptidos en total).

El método desarrollado permite separar las diferentes glicofomas en función del número de SiA dado que, al pH de trabajo, éstos se encuentran cargados negativamente. Además, el elevado poder de resolución de esta técnica permite separar glicofomas con el mismo contenido de SiA y que difieren en el número de ramificaciones. Con el objetivo de identificar qué glicopéptidos permiten diferenciar entre muestras patológicas, se estudian los mapas glicopéptídicos con herramientas de análisis multivariante de datos (PCA y PLS-DA), llevando a cabo una normalización de las áreas de pico de cada glicofoma respecto a un péptido de referencia del digesto con área y tiempo de migración similar a los N-glicopéptidos detectados. Tal y como se muestra en el gráfico de *scores* de la Figura 3B, el modelo de PLS-DA, construido con las áreas normalizadas de las glicofomas identificadas, permite discriminar entre muestras de CEA de cáncer de colon y de metástasis de hígado. Los valores de VIP *scores* sugieren que las glicofomas con mayor fucosilación y sialilación, que se encuentran más expresadas en las muestras de metástasis, son las más relevantes para esta diferenciación. Este aumento podría ser debido a la gravedad y la invasión del tumor en cuestión y no al tipo de cáncer, pero podría ser de gran aplicabilidad para la detección y el tratamiento precoz de las metástasis [18].

Es importante destacar, que las estrategias analíticas descritas, además de a otras glicoproteínas biomarcadoras, se podrían aplicar a biofármacos glicoproteicos (por ejemplo, hormonas o anticuerpos monoclonales), que comienzan a abundar en el mercado.

### Agradecimientos

Esta investigación del grupo Bioanálisis se ha financiado a través del Ministerio de Economía y Competitividad (RTI2018-097411-B-I00 y programa FPU de becas predoctorales), la cátedra UB Rector Francisco Buscarons Úbeda (Química e Ingeniería Química Forense) y el programa ADR de becas predoctorales de la Universidad de Barcelona.

### Referencias

[1] M.J. Kailemia, D. Park, B.L. Carlito, Glycans and Glycoproteins as Specific Biomarkers for Cancer, *Anal. Bioanal. Chem.* 409 (2017) 395–410.  
 [2] M. Hu, Y. Lan, A. Lu, X. Ma, L. Zhang, Glycan-based biomarkers for diagnosis of cancers and other diseases: Past, present, and future, *Glycan-based biomarkers for diagnosis of cancers and other diseases: Past, present, and future*, 1st ed., Elsevier Inc., 2019.  
 [3] A. Banazadeh, L. Veillon, K.M. Wooding, M. Zabetmoghaddam, Y. Mechref, Recent advances in mass spectrometric analysis of glycoproteins, *Recent advances in mass spectrometric analysis of glycoproteins*, 2017.  
 [4] A. Cristobal, F. Marino, H. Post, H.W.P. Van Den Toorn, S. Mohammed, A.J.R. Heck, Toward an Optimized Workflow for Middle-Down Proteomics, *Anal. Chem.* 89 (2017) 3318–3325.

[5] B.Y. Huang, C.K. Yang, C.P. Liu, C.Y. Liu, Stationary phases for the enrichment of glycoproteins and glycopeptides, *Electrophoresis* 35 (2014) 2091–2107.  
 [6] A. Barroso, E. Giménez, F. Benavente, J. Barbosa, V. Sanz-Nebot, Classification of congenital disorders of glycosylation based on analysis of transferrin glycopeptides by capillary liquid chromatography-mass spectrometry, *Talanta* 160 (2016) 614–623.  
 [7] M. Mancera-Arteu, E. Giménez, M. Balmaña, S. Barrabés, M. Albiol-Quer, E. Fort, R. Peracaula, V. Sanz-Nebot, Multivariate data analysis for the detection of human alpha-acid glycoprotein aberrant glycosylation in pancreatic ductal adenocarcinoma, *J. Proteomics* 195 (2019).  
 [8] M. Mancera-Arteu, E. Giménez, J. Sancho, V. Sanz-Nebot, Alterations in the Glycan Profile of Mouse Transferrin: New Insights in Collagen-Induced Arthritis, *J. Proteome Res.* 19 (2020) 1750–1759.  
 [9] A. Suppers, A.J. van Gool, H.J.C.T. Wessels, Integrated chemometrics and statistics to drive successful proteomics biomarker discovery, *Proteomes* 6 (2018).  
 [10] L. Pont, V. Kuzyk, F. Benavente, V. Sanz-Nebot, O.A. Mayboroda, M. Wuhler, G.S.M. Lageveen-Kammeijer, Site-Specific N-Linked Glycosylation Analysis of Human Carcinoembryonic Antigen by Sheathless Capillary Electrophoresis-Tandem Mass Spectrometry, *J. Proteome Res.* 20 (2021) 1666–1675.  
 [11] S. Zhou, L. Veillon, X. Dong, Y. Huang, Y. Mechref, Direct comparison of derivatization strategies for LC-MS/MS analysis of: N-glycans, *Analyst* 142 (2017) 4446–4455.  
 [12] D.G. Delafield, L. Li, Recent advances in analytical approaches for glycan and glycopeptide quantitation, *Mol. Cell. Proteomics* 20 (2021) 0–21.  
 [13] E. Giménez, V. Sanz-Nebot, A. Rizzi, Relative quantitation of glycosylation variants by stable isotope labeling of enzymatically released N-glycans using [12C]/[13C] aniline and ZIC-HILIC-ESI-TOF-MS, *Anal. Bioanal. Chem.* 405 (2013) 7307–7319.  
 [14] M. Mancera-Arteu, E. Giménez, J. Barbosa, V. Sanz-Nebot, Identification and characterization of isomeric N-glycans of human alpha-acid-glycoprotein by stable isotope labelling and ZIC-HILIC-MS in combination with exoglycosidase digestion, *Anal. Chim. Acta* 940 (2016) 92–103.  
 [15] J. Nilsson, Liquid chromatography-tandem mass spectrometry-based fragmentation analysis of glycopeptides, *Glycoconj. J.* 33 (2016) 261–272.  
 [16] E. Giménez, R. Ramos-Hernan, F. Benavente, J. Barbosa, V. Sanz-Nebot, Analysis of recombinant human erythropoietin glycopeptides by capillary electrophoresis electrospray-time of flight-mass spectrometry, *Anal. Chim. Acta* 709 (2012) 81–90.  
 [17] A. Barroso, E. Giménez, F. Benavente, J. Barbosa, V. Sanz-Nebot, Improved tryptic digestion assisted with an acid-labile anionic surfactant for the separation and characterization of glycopeptide glycoforms of a proteolytic-resistant glycoprotein by capillary electrophoresis time-of-flight mass spectrometry, *Electrophoresis* 37 (2016) 987–997.  
 [18] L. Oliveira-Ferrer, K. Legler, K. Milde-Langosch, Role of protein glycosylation in cancer metastasis, *Semin. Cancer Biol.* 44 (2017) 141–152.