



UNIVERSITAT DE  
BARCELONA

## Separación y caracterización de glicopéptidos y glicanos en fluidos biológicos. Aplicación al diagnóstico del cáncer y otras patologías

Montserrat Mancera Artau



Aquesta tesi doctoral està subjecta a la llicència Reconeixement- NoComercial – Compartir Igual 4.0. Espanya de Creative Commons.

Esta tesis doctoral está sujeta a la licencia Reconocimiento - NoComercial – Compartir Igual 4.0. España de Creative Commons.

This doctoral thesis is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0. Spain License.

2020

TESIS DOCTORAL

# SEPARACIÓN Y CARACTERIZACIÓN DE GLICOPÉPTIDOS Y GLICANOS EN FLUIDOS BIOLÓGICOS

APLICACIÓN AL DIAGNÓSTICO DEL CÁNCER Y OTRAS PATOLOGÍAS

Montserrat Mancera Arteu

*Montserrat Mancera Arteu*



**SEPARACIÓN Y CARACTERIZACIÓN DE  
GLICOPÉPTIDOS Y GLICANOS EN FLUIDOS BIOLÓGICOS**

APLICACIÓN AL DIAGNÓSTICO DEL CÁNCER Y OTRAS PATOLOGÍAS



*Montserrat Mancera Artau*

---



UNIVERSITAT DE  
BARCELONA



FACULTAD DE QUÍMICA

DEPARTAMENTO DE INGENIERÍA QUÍMICA Y QUÍMICA ANALÍTICA

Programa de Doctorado:

Química Analítica y Medio Ambiente

---

**Separación y caracterización de glicopéptidos y glicanos en fluidos biológicos. Aplicación al diagnóstico del cáncer y otras patologías**

---

Memoria presentada por:

**Montserrat Mancera Arteu**

para optar al grado de Doctor por la Universidad de Barcelona

Bajo la dirección de:

Dra. Victoria Sanz Nebot

Dra. Estela Giménez López

del Departamento de Ingeniería Química y Química Analítica de la Universidad de Barcelona



La Dra. Victoria Sanz Nebot, profesora titular del Departamento de Ingeniería Química y Química Analítica de la Universidad de Barcelona, y la Dra. Estela Giménez López, profesora agregada del mismo Departamento,

HACEN CONSTAR,

Que la presente memoria titulada: **“Separación y caracterización de glicopéptidos y glicanos en fluidos biológicos. Aplicación al diagnóstico del cáncer y otras patologías”**, ha sido realizada bajo su dirección por la Sra. **Montserrat Mancera Arteu** y que todos los resultados presentados son fruto de las experiencias realizadas por la citada doctoranda.

Y para que así conste, expiden el presente certificado.

Barcelona, septiembre 2020



Dra. Victoria Sanz Nebot



Dra. Estela Giménez López



***"I was taught that the way of progress was neither swift nor easy"***

**Marie Curie**





## **Agradecimientos**

---



Ya ha llegado el fin de esta maravillosa etapa, en la que he sufrido cuando las cosas no salían como esperaba pero también he disfrutado con los avances conseguidos y me ha permitido conocer a personas estupendas. No podría empezar de otra manera que agradeciendo a todos aquellos que me han apoyado durante este largo camino.

En primer lugar, quiero agradecer el esfuerzo y dedicación de mis directoras de tesis, la Dra. Victòria Sanz Nebot y la Dra. Estela Giménez López. Vicki, gràcies per estar-hi sempre, per tots els consells que m'has donat, tan personals com científics, per intentar sempre ajudar-me i que tot m'anés bé i per ser comprensiva i atenta amb mi. Estela, gràcies per poder sempre parlar amb tu, per ajudar-me en tot el que has pogut i més, per tranquil·litzar-me quan les coses no són com esperava i per ensenyar-me l'existència del fullet que ens canvia les coses per les nits jejeje. Perquè com un cop em vas dir som un equip i les victòries al igual que les derrotes son compartides i així ho hem fet fins ara. Us dec molt a les dues! Tambien estoy muy agradecida al Dr. José Barbosa por estar siempre pendiente para que todo siga su curso, por ayudar-me en todo lo que estaba a su alcance y por sus sabios consejos siempre que los he necesitado. Al Dr. Fernando Benavente, gracias por tu ayuda durante todos estos años, por tu facilidad para encontrar alternativas interesantes y, sobretodo, soluciones a los problemas que se presentan.

A mis compañeros de fatigas e incondicionales amigos que esta tesis me ha dado: Laura, Albert y Roger, les debo un gran apartado en esta sección. Laura, te diría tanto que no sé por dónde empezar... gracias por todo lo que me has ayudado y sigues ayudándome hoy en día, por todas nuestras conversaciones y risas que nos hemos pegado, por esa dulzura tan bonita que te caracteriza, por ser el juicio para mi locura y la calma para mis nervios. ¡Te quiero! A ti, Albert, gracias por ser mi guía cuando empecé este camino y enseñarme tantas cosas con esa paciencia que sólo tú tienes, por todo lo que hemos reído juntos y por esos vídeos épicos que aun conservamos. Me llevo un gran amigo. Roger, gràcies per ensenyar-me que la verdadera

amistat no entén de caràcters ni de formes de ser, per estar sempre disposat a escoltar-me i ajudar-me en tot lo necessari, per tots els congressos que em compartit, per tenir l'oportunitat de compartir l'estrès de totes les etapes d'aquesta tesi amb tu, m'emporto rècords molt macos! Gracias también a Hiba, Rocío y Oumaima por acompañarme en esta última etapa de la tesis, escucharme, ayudarme siempre que lo he necesitado y amenizar los momentos en el laboratorio con las músicas tradicionales de sus países.

También me gustaría dar las gracias a los compañeros con los que he tenido la oportunidad de compartir laboratorio durante periodos de esta tesis y han dejado huella en mí: Irena, Laura Pey, Lorena, Marina, Nejsi, Víctor y Aymed. Gracias por vuestro apoyo siempre y por todos esos buenos momentos que hemos pasado juntos y seguimos pasando cuando podemos vernos. Nejsi, I hope that someday we will be able to meet again.

Tampoco puedo olvidarme de mis compañeros de departamento y amigos que se han encargado de hacer mucho más llevaderos los días con nuestros raritos de risas a la hora de comer o desayunar: Juanfran, Ane, Guillem, Noe, Àlex, Sara, Daria, Alejandro y Clara. Gracias por los momentos compartidos, que espero que sigan, y por abrirme vuestro laboratorio siempre que lo necesitaba.

I would also like to thank Dr. Kevin Pagel for giving me the opportunity of working in his laboratory and for his kindness and support during my stay in Berlin. Thanks also to Christian, Rayoon, Michele and Waldemar for your hospitality, for trying always to make me feel at home, for taking care of me during that period and for sharing all together great moments. I am still waiting your visit in Barcelona hahaha.

Gracias a mis buenas amigas, Paula, Judit y Lea, por estar ahí siempre que lo necesito y hacer que los problemas se vean más pequeños. ¡Os quiero!

También gracias a Sílvia i Martí de Ahead Therapeutics, por confiar en mí y darme la oportunidad en esta última etapa de seguir haciendo lo que más me gusta.

Me gustaría agradecer especialmente a toda mi familia. Gracias mama, yaya y Carlos por cuidarme y escucharme siempre, por entender mi mal humor cuando estaba nerviosa, por cuidar de la pitufa cuando lo necesito y por hacerme la vida más fácil haciéndome la comida para el día siguiente jejeje. Albert, gracias por estar siempre ahí, a pesar de nuestras discusiones de hermanos jejeje, por hacerme la mejor portada que esta tesis podría tener porque ahora también tiene un trocito de ti. Avi, gràcies per ajudar-me des del cel sempre que t'ho he demanat, perquè encara que desgraciadament no puguis estar físicament, em fas sentir segura en tot moment i m'arriba el teu suport des d'allà. Gracias también a mis suegros y cuñados por su apoyo, por escuchar también mis problemas e intentar aportar soluciones. ¡Us estimo molt a tots!

Gracias infinitas a mi compañero de vida. Javi, gracias por animarme a empezar este proyecto, por escucharme tan atentamente siempre, aunque no entiendas de glicanos, y aportar la mejor de las soluciones. Gracias por los innumerables consejos a lo largo de estos años, por encargarte de todo en casa para que yo tenga tiempo para escribir la tesis, por compartir mi alegría cuando las cosas salían bien y, sobretodo, mi mal humor en épocas de estrés. Esta tesis, como todo lo que hemos conseguido juntos, es de ambos. Y, finalmente, agradecer a mi hija, quien vino al final de esta etapa para revolucionarlo todo y darme la mayor felicidad. Gràcies Arlet per estar al meu braçet mentre escrivia la tesi, la hem escrit juntes, i per rebre'm sempre amb aquest somriure immens que fa oblidar qualsevol problema quan arribo a casa. ¡Us estimo infinit als dos!

**Gràcies a tots / Gracias a todos / Thanks to everyone**



## **Abbreviations and acronyms**

---





**Symbol / Number**

|             |                   |
|-------------|-------------------|
| $\beta$ -ME | 2-mercaptoethanol |
| 1Ant        | Monoantennary     |
| 2Ant        | Biantennary       |
| 3Ant        | Triantennary      |

**A**

|             |  |
|-------------|--|
| Ab          | Antibody                                 |
| ACN         | Acetonitrile                             |
| ACPAs       | Anti-citrullinated protein antibodies    |
| ADCC        | Antibody-dependent cellular cytotoxicity |
| Amino acids |  |

| Amino acid    | Three letters code | One letter code |
|---------------|--------------------|-----------------|
| Alanine       | Ala                | A               |
| Arginine      | Arg                | R               |
| Asparagine    | Asn                | N               |
| Aspartic acid | Asp                | D               |
| Cysteine      | Cys                | C               |
| Glutamic acid | Glu                | E               |
| Glutamine     | Gln                | Q               |
| Glycine       | Gly                | G               |
| Histidine     | His                | H               |
| Isoleucine    | Ile                | I               |
| Leucine       | Leu                | L               |
| Lysine        | Lys                | K               |
| Methionine    | Met                | M               |
| Phenylalanine | Phe                | F               |
| Proline       | Pro                | P               |
| Serine        | Ser                | S               |
| Threonine     | Thr                | T               |
| Tryptophan    | Trp                | W               |
| Tyrosine      | Tyr                | Y               |
| Valine        | Val                | V               |

|        |                           |
|--------|---------------------------|
| AN     | Aniline                   |
| Ant    | Antenna                   |
| APO-C3 | Apolipoprotein C-III      |
| ATD    | Arrival time distribution |

**B**

|      |                                      |
|------|--------------------------------------|
| bAGP | Bovine $\alpha$ -1-acid glycoprotein |
| BGE  | Background electrolyte               |

**C**

---

|                    |  |
|--------------------|--|
| CA19-9             | Cancer antigen 19-9  |
| CapLC              | Capillary liquid chromatography  |
| CapLC-MS           | Capillary liquid chromatography mass spectrometry  |
| CapLC-TOF-MS       | Capillary liquid chromatography time-of-flight mass spectrometry                                   |
| CapZIC-HILIC-MS    | Capillary liquid chromatography with zwitterionic hydrophilic interaction mass spectrometry        |
| CapZIC-HILIC-MS/MS | Capillary liquid chromatography with zwitterionic hydrophilic interaction tandem mass spectrometry |
| CCS                | Collision cross-section  |
| CDGs               | Congenital disorders of glycosylation  |
| CE                 | Capillary electrophoresis  |
| CE-ESI-MS          | Capillary electrophoresis electrospray ionization mass spectrometry                                |
| CE-MS              | Capillary electrophoresis mass spectrometry  |
| CE-TOF-MS          | Capillary electrophoresis time-of-flight mass spectrometry   |
| CFA                | Complete Freund's adjuvant   |
| CFG                | Consortium for Functional Glycomics  |
| CHO                | Chinese hamster ovary cells  |
| ChrP               | Chronic pancreatitis   |
| CIA                | Collagen induced arthritis   |
| CID                | Collision induced dissociation   |
| Col-II             | Chicken collagen type II   |
| CZE                | Capillary zone electrophoresis   |

**D**

---

|     |                |
|-----|----------------|
| Da  | Dalton         |
| DTT | Dithiothreitol |

**E**

---

|        |  |
|--------|--|
| EIC    | Extracted ion chromatogram                     |
| EIE    | Extracted ion electropherogram                 |
| emPAI  | Exponentially modified protein abundance index |
| EOF    | Electroosmotic flow                            |
| EPO    | Human erythropoietin                           |
| ESA    | Erythropoiesis-stimulating agent               |
| ESI    | Electrospray ionization                        |
| ESI-MS | Electrospray ionization mass spectrometry      |

---

**F**

|        |   |
|--------|---|
| FAIMS  | Field asymmetric waveform ion mobility spectrometry |
| FDA    | Food and drug administration                        |
| FT-ICR | Fourier transform ion cyclotron resonance           |

---

**G**

|       |                                    |
|-------|------------------------------------|
| GRAVY | Grand average of hydropathy        |
| GRIL  | Glycan reductive isotope-labelling |

---

**H**

|               |  |
|---------------|--|
| HAc           | Acetic acid  |
| hAGP (or AGP) | Human $\alpha$ -1-acid glycoprotein  |
| HC            | Healthy control  |
| HCD           | Higher-energy collisional dissociation   |
| hEPO          | Human erythropoietin   |
| HFor          | Formic acid  |
| HILIC         | Hydrophilic interaction liquid chromatography                                    |
| HPLC          | High performance liquid chromatography   |
| HPLC-ESI-MS   | High performance liquid chromatography electrospray ionization mass spectrometry |
| HPLC-MS       | High performance liquid chromatography mass spectrometry                         |
| hTf           | Human transferrin  |

---

**I**

|       |   |
|-------|---|
| IAA   | Iodoacetamide                                 |
| IAC   | Immunoaffinity chromatography                 |
| id    | Inner or internal diameter                    |
| IEF   | Isoelectric focusing                          |
| Ig    | Immunoglobulin                                |
| IM    | Ion mobility                                  |
| IMAC  | Immobilized metal ion affinity chromatography |
| IMS   | Ion mobility spectrometry                     |
| IM-MS | Ion mobility mass spectrometry                |
| iPrOH | 2-propanol or isopropanol                     |
| IT    | Ion trap                                      |
| ITP   | Isotachopheresis                              |

---

**L**

|        |                       |
|--------|-----------------------|
| LacNAc | N-acetyllactosamine   |
| LC     | Liquid chromatography |

|       |   |
|-------|---|
| LC-MS | Liquid chromatography mass spectrometry |
| LOD   | Limit of detection                      |
| $L_T$ | Total length                            |
| LTQ   | Linear ion trap                         |
| LV    | Latent variable                         |

## M

---

|                             |   |
|-----------------------------|---|
| M or $M_r$                  | Molecular mass                              |
| $m_e$ ( $\mu_e$ )           | Electrophoretic mobility                    |
| $M_{exp}$                   | Experimental molecular mass                 |
| $M_{glycan} / M_{glicano}$  | Glycan molecular mass                       |
| $M_{peptide} / M_{péptido}$ | Peptide molecular mass                      |
| $M_{theo} / M_{teo}$        | Theoretical molecular mass                  |
| MALDI                       | Matrix-assisted laser desorption/ionization |

Monosaccharides

| Monosaccharide                 | Abbreviation | One letter code |
|--------------------------------|--------------|-----------------|
| Galactose                      | Gal          | H               |
| Glucose                        | Glc          | H               |
| Mannose                        | Man          | H               |
| Hexose                         | Hex          | H               |
| N-acetylgalactosamine          | GalNAc       | N               |
| N-acetylglucosamine            | GlcNAc       | N               |
| N-acetylhexosamine             | HexNAc       | N               |
| Fucose                         | Fuc          | F               |
| N-acetylneuraminic acid        | NeuAc        | S               |
| N-glycolylneuraminic acid      | NeuGc        | S               |
| Sialic acid or neuraminic acid | SiA          | S               |

|       |                          |
|-------|--------------------------|
| MS    | Mass spectrometry        |
| MS/MS | Tandem mass spectrometry |
| mTf   | Mouse transferrin        |
| MW    | Molecular weight         |
| $m/z$ | Mass-to-charge ratio     |

## N

---

|               |  |
|---------------|--|
| n             | Number of replicates                                       |
| nanoESI       | Nanoelectrospray ionization                                |
| nanoESI-IM-MS | Nanoelectrospray ionization ion mobility mass spectrometry |
| nanoLC        | Nano liquid chromatography                                 |
| NanoLC-MS/MS  | Nano liquid chromatography tandem mass spectrometry        |
| NaOH          | Sodium hydroxide   |
| $NH_4Ac$      | Ammonium acetate   |

NH<sub>4</sub>OH Ammonium hydroxide

## O

oa Orthogonal acceleration  
 oa-TOF Orthogonal acceleration time-of-flight mass analyzer  
 od Outer diameter  
 OVA Ovalbumin

## P

PAGE Polyacrylamide gel electrophoresis  
 PC Principal component  
 PCA Principal component analysis  
 PDAC Pancreatic ductal adenocarcinoma  
 PGC Porous graphitic carbon  
 pI Isoelectric point  
 PLS Partial least squares  
 PLS-DA Partial least squares discriminant analysis  
 PNGase-F N-glycosidase F  
 ppm Part per million  
 PTM Posttranslational modification

## Q

q Electric charge  
 Q Quadrupole  
 qTOF Hybrid quadrupole time-of-flight analyzer  
 q/M Charge-to-mass ratio

## R

RA Rheumatoid arthritis  
 Rf Rheumatic factor  
 rhEPO Recombinant human erythropoietin  
 RP-LC Reversed phase liquid chromatography  
 RSD (or %RSD) Relative standard deviation (%)  
 R<sup>2</sup> Regression coefficient

## S

s Standard deviation  
 SDS Sodium dodecyl sulfate  
 SDS-PAGE Sodium dodecyl sulfate polyacrylamide gel electrophoresis  
 SLe<sup>a</sup> Sialyl lewis A

|                  |  |
|------------------|--|
| SLe <sup>x</sup> | Sialyl lewis X   |
| SPE-CE           | On-line solid-phase extraction capillary electrophoresis                   |
| SPE-CE-MS        | On-line solid-phase extraction capillary electrophoresis mass spectrometry |
| S/N              | Signal-to noise ratio  |

## T

---

|                             |  |
|-----------------------------|--|
| Tf                          | Transferrin  |
| TFA                         | Trifluoroacetic acid   |
| TiO <sub>2</sub>            | Titanium dioxide   |
| TiO <sub>2</sub> -SPE-CE-MS | Solid phase extraction capillary electrophoresis with titanium dioxide sorbent mass spectrometry |
| TOF                         | Time-of-flight mass analyzer   |
| TOF-MS                      | Time-of-flight mass spectrometry   |
| t <sub>R</sub>              | Retention time   |
| TWIMS                       | Traveling wave ion mobility spectrometry   |

## V

---

|     |                                   |
|-----|-----------------------------------|
| VIP | Variable importance on projection |
|-----|-----------------------------------|

## W

---

|    |               |
|----|---------------|
| WH | Wave height   |
| WT | Wild-type     |
| WV | Wave velocity |

## Z

---

|           |  |
|-----------|--|
| z         | Electric charge  |
| ZIC-HILIC | Zwitterionic hydrophilic interaction liquid chromatography |

**Index**

---





|   |                |
|---|----------------|
| <b>ABSTRACT</b>   | <b>- 1 -</b>   |
| <b>AIMS</b>   | <b>- 7 -</b>   |
| <b>CHAPTER 1. INTRODUCTION</b>  | <b>- 11 -</b>  |
| <b>1.1. GLYCOSYLATION</b>   | <b>- 13 -</b>  |
| <b>1.2. GLYCOPROTEINS AS BIOMARKERS AND BIOPHARMACEUTICALS</b>                          | <b>- 21 -</b>  |
| <b>1.2.1. CANCER. PANCREATIC DUCTAL ADENOCARCINOMA</b>                                  | <b>- 22 -</b>  |
| 1.2.1.1. HUMAN ALPHA-1-ACID GLYCOPROTEIN  | - 24 -         |
| <b>1.2.2. ARTHRITIS. COLLAGEN-INDUCED ARTHRITIS</b>                                     | <b>- 27 -</b>  |
| 1.2.2.1. MOUSE TRANSFERRIN  | - 28 -         |
| <b>1.2.3. BIOPHARMACEUTICALS</b>  | <b>- 29 -</b>  |
| 1.2.3.1. RECOMBINANT HUMAN ERYTHROPOIETIN   | - 31 -         |
| <b>1.3. GLYCOPROTEIN ANALYSIS</b>   | <b>- 33 -</b>  |
| <b>1.3.1. PURIFICATION TECHNIQUES</b>   | <b>- 37 -</b>  |
| <b>1.3.2. LIQUID CHROMATOGRAPHY</b>   | <b>- 40 -</b>  |
| <b>1.3.3. ELECTROPHORESIS</b>   | <b>- 42 -</b>  |
| 1.3.3.1 GEL ELECTROPHORESIS   | - 42 -         |
| 1.3.3.2 CAPILLARY ELECTROPHORESIS   | - 43 -         |
| <b>1.3.4. MASS SPECTROMETRY</b>   | <b>- 46 -</b>  |
| 1.3.4.1 ELECTROSPRAY IONIZATION   | - 47 -         |
| 1.3.4.2 MASS ANALYZERS  | - 49 -         |
| <b>1.3.5. ION MOBILITY-MASS SPECTROMETRY</b>  | <b>- 51 -</b>  |
| <b>1.4. DATA ANALYSIS. CHEMOMETRIC METHODS</b>  | <b>- 54 -</b>  |
| <b>1.4.1. PRINCIPAL COMPONENT ANALYSIS</b>  | <b>- 55 -</b>  |
| <b>1.4.2. PARTIAL LEAST DISCRIMINANT ANALYSIS</b>                                       | <b>- 56 -</b>  |
| <b>CHAPTER 2. GLYCOPEPTIDE PURIFICATION METHODS</b>                                     | <b>- 59 -</b>  |
| <b>CHAPTER 3. GLYCAN ISOMER CHARACTERIZATION</b>  | <b>- 97 -</b>  |
| <b>CHAPTER 4. GLYCAN-BASED BIOMARKER DISCOVERY</b>                                      | <b>- 139 -</b> |
| <b>CAPÍTULO 5. RESULTADOS Y DISCUSIÓN</b>   | <b>- 177 -</b> |
| <b>5.1. ESTABLECIMIENTO DE METODOLOGÍAS PARA LA PURIFICACIÓN DE GLICOPÉPTIDOS</b>       | <b>- 179 -</b> |
| <b>5.1.1. PURIFICACIÓN DE GLICOPÉPTIDOS MEDIANTE PRECIPITACIÓN CON ACETONA</b>          | <b>- 179 -</b> |
| 5.1.1.1. ESTUDIO DE LA PRECIPITACIÓN DE PÉPTIDOS Y GLICOPÉPTIDOS                        | - 180 -        |
| 5.1.1.2. ESTUDIO DE LA PRECIPITACIÓN DE O-GLICOPÉPTIDOS                                 | - 183 -        |
| 5.1.1.3. OPTIMIZACIÓN DE LA PRECIPITACIÓN DEL GLICOPÉPTIDO O <sub>126</sub> DE LA RHEPO | - 185 -        |
| <b>5.1.2. PURIFICACIÓN DE GLICOPÉPTIDOS MEDIANTE TiO<sub>2</sub>-SPE-CE-MS</b>          | <b>- 187 -</b> |
| 5.1.2.1. CONSTRUCCIÓN DEL PRECONCENTRADOR   | - 188 -        |

|  |                |
|--|----------------|
| 5.1.2.2. OPTIMIZACIÓN DE LA PURIFICACIÓN DEL O-GLICOPÉPTIDO DE LA RHEPO POR TiO <sub>2</sub> -SPE-CE-MS    | - 190 -        |
| 5.1.2.3. ANÁLISIS DE LOS GLICOPÉPTIDOS DE LA RHEPO POR TiO <sub>2</sub> -SPE-CE-MS                         | - 194 -        |
| 5.1.2.4. ANÁLISIS DE LOS GLICOPÉPTIDOS DE OTRAS GLICOPROTEÍNAS POR TiO <sub>2</sub> -SPE-CE-MS             | - 199 -        |
| <b>5.1.3. COMPARACIÓN DE LOS MÉTODOS DE PURIFICACIÓN DE GLICOPÉPTIDOS</b>                                  | <b>- 199 -</b> |
| <b>5.2. SEPARACIÓN Y CARACTERIZACIÓN DE GLICANOS Y SUS CORRESPONDIENTES ISÓMEROS</b>                       | <b>- 201 -</b> |
| <b>5.2.1. OPTIMIZACIÓN DE LAS CONDICIONES CROMATOGRÁFICAS POR CAPZIC-HILIC-MS</b>                          | <b>- 202 -</b> |
| <b>5.2.2. CARACTERIZACIÓN DE GLICANOS MEDIANTE DIGESTIÓN CON EXOGLICOSIDASAS</b>                           | <b>- 205 -</b> |
| 5.2.2.1. CARACTERIZACIÓN DE LOS ENLACES DE LOS ÁCIDOS SIÁLICOS   | - 207 -        |
| 5.2.2.2. CARACTERIZACIÓN DE LOS ENLACES DE LAS FUCOSAS   | - 211 -        |
| <b>5.2.3. CARACTERIZACIÓN DE GLICANOS POR ESPECTROMETRÍA DE MASAS EN TÁNDEM</b>                            | <b>- 214 -</b> |
| 5.2.3.1. CARACTERIZACIÓN DE LOS ENLACES DE LOS ÁCIDOS SIÁLICOS   | - 217 -        |
| 5.2.3.3. CARACTERIZACIÓN DE LOS ENLACES DE LAS FUCOSAS   | - 222 -        |
| <b>5.2.4. CARACTERIZACIÓN DE GLICANOS POR ESPECTROMETRÍA DE MASAS DE MOVILIDAD IÓNICA (IM-MS)</b>          | <b>- 225 -</b> |
| <b>5.2.5. COMPARACIÓN DE LOS MÉTODOS DE CARACTERIZACIÓN DE GLICANOS ESTABLECIDOS</b>                       | <b>- 233 -</b> |
| <b>5.3. ESTUDIO E IDENTIFICACIÓN DE GLICANOS BIOMARCADORES</b>   | <b>- 234 -</b> |
| <b>5.3.1. ANÁLISIS DE LOS GLICANOS DE LA HAGP EN MUESTRAS DE CÁNCER DE PÁNCREAS Y PANCREATITIS CRÓNICA</b> | <b>- 235 -</b> |
| 5.3.1.1. ANÁLISIS DE LOS GLICANOS INTACTOS   | - 235 -        |
| 5.3.1.2. ANÁLISIS DE LOS GLICANOS DESIALILADOS   | - 238 -        |
| <b>5.3.2. ANÁLISIS DE LOS GLICANOS DE LA MTF EN MUESTRAS DE ARTRITIS INDUCIDA POR COLÁGENO</b>             | <b>- 243 -</b> |
| 5.3.2.1. PURIFICACIÓN DE LA MTF MEDIANTE CROMATOGRAFÍA DE INMUNOAFINIDAD                                   | - 244 -        |
| 5.3.2.2. ANÁLISIS DE LOS GLICANOS DE LA MTF  | - 246 -        |
| <b>5.3.3. COMPARACIÓN DE LOS RESULTADOS OBTENIDOS EN PDAC, CHRP Y CIA</b>                                  | <b>- 252 -</b> |
| <b>CONCLUSIONS</b>   | <b>- 253 -</b> |
| <b>REFERENCES</b>  | <b>- 259 -</b> |

## **Abstract**

---



Protein glycosylation of proteins is the covalent addition of carbohydrates, also known as **glycans**, to the polypeptide core and it is a **co-** or **posttranslational modification (PTM)** mediated by enzymes. Protein glycosylation is considered the most frequent modification and it has been estimated that over half of the proteins in nature are glycosylated. Glycosylated proteins, commonly known as **glycoproteins**, play an important role in major biological processes, such as cell-cell interaction, signalling and protein folding, which are mainly modulated and controlled by glycans. Glycans can be formed by several different monosaccharide units and adopt several structures generating structural diversity in proteins. Therefore, each glycoprotein is composed by several **glycoforms**, heterogeneous mixtures of glycoconjugate species that differ in the amount, size and/or structure of the attached glycans. Additionally, glycans with the same composition could present different configuration or connectivity of the glycosidic bonds between monosaccharides, which has led to the existence of isomeric glycan structures. The main difference between **glycan isomers** has been associated with different sialic acid or fucose linkage-types.

Alterations in protein glycans have been described in many diseases such as important inflammatory processes and many types of cancer. Moreover, certain glycan structures as well as linkage-types have been related to specific diseases. Most of glycan alterations have been associated with under or overexpression of glycosyltransferases that generally results in sialylation, fucosylation and/or branching changes. In the case of cancer, a close relationship between altered glycosylation and tumour initiation, progression and metastasis was also found. Thus, in last years, special attention has paid to find novel **glycan-based biomarkers** that could be used for early diagnosis, disease monitoring or prognosis of many pathologies. In this regard, **acute-phase proteins**, which concentration is adaptively regulated in response to most forms of inflammation, infection and tissue injury, could have great importance. On the other hand, the glycosylation pattern of **recombinant glycoproteins**, which are frequently used as biopharmaceuticals, usually differ with respect to the endogenous one and it can cause an

adverse immune response. In addition, differences in the biological activity and pharmacokinetic among glycoforms have been observed. Hence, an in-depth characterization of recombinant protein glycosylation is crucial to ensure their proper therapeutic function.

In this thesis, methods to selectively enrich glycopeptides from glycoprotein digests were developed using **recombinant human erythropoietin (rhEPO)** as model glycoprotein. Although the analysis of glycopeptides is a commonly used strategy for glycan characterization of biopharmaceuticals, the detection of low abundant glycopeptides that could be essential to differentiate endogenous and recombinant variants of certain proteins continues to be a challenge. Precipitation with acetone was firstly studied and optimized in this thesis to maximize glycopeptide enrichment with the minimal coprecipitation of peptides. Principal component analysis (**PCA**) and partial least squares discriminant analysis (**PLS-DA**) were used to investigate the physico-chemical parameters and properties that make peptides, N- and O-glycopeptides present a different trend in front of acetone precipitation. The best conditions for rhEPO O-glycopeptide purification were selected due to its interest in biopharmaceutical analysis and doping control. The method was also successfully applied for the purification of glycopeptides from other glycoproteins.

On-line solid-phase extraction capillary electrophoresis mass spectrometry (**SPE-CE-MS**) using titanium dioxide (TiO<sub>2</sub>) as a sorbent was also evaluated in this thesis for glycopeptide purification and enrichment. Under the optimal conditions, repeatability and microcartridge lifetime were acceptable and limit of detection (LOD) for rhEPO O-glycopeptide was significantly decreased. Glycopeptide enrichment factors up to 100 times were achieved. The applicability of the method for the purification of glycopeptides with different compositions and structures was also demonstrated.

On the other hand, different analytical methodologies to separate glycan isomers and characterize their structures and linkage-types between monosaccharides were developed in

this thesis. A **CapZIC-HILIC-MS** method previously established by our research group was firstly optimized to improve isomer separation and sensitivity of **human alpha-1-acid glycoprotein (hAGP)** glycans. Using this approach, exoglycosidase digestion in combination with **glycan relative isotope labelling (GRIL)** strategy using  $[^{12}\text{C}_6]/[^{13}\text{C}_6]$ aniline was successfully used for sialic acid and fucose linkage-type assignments of most hAGP glycan isomers. Likewise, the CapZIC-HILIC separation coupled to **tandem mass spectrometry** was also applied in this thesis for the characterization of hAGP glycan isomers. In this regard, the diagnostic value of some ion fragments reported in the literature for sialic acid and fucose linkage-type characterization was confirmed or discarded using the hAGP glycan assignments performed previously. In addition, this approach allowed deepening into glycan characterization identifying in some cases the sialic acid position on each antenna. On the other hand, an **ion mobility mass spectrometry (IM-MS)** method was also established in this thesis for separation and characterization of glycan isomers due to its described potential to separate isomeric compounds. Separation between glycan isomers was achieved in both intact and fragment levels of hAGP glycans. However, it was observed that isomer separation by IM-MS in complex glycan structures can be related to differences in antenna localization rather than to different sialic acid linkage-types.

Finally, the established CapZIC-HILIC-MS method in combination with the GRIL strategy with  $[^{12}\text{C}_6]/[^{13}\text{C}_6]$ aniline was also used in this thesis to evaluate possible glycan isomer modifications in two acute-phase proteins, hAGP and **mouse transferrin (mTf)**, caused by certain inflammatory diseases and cancer. The importance of separating and identifying glycan isomers was demonstrated in recent years since only certain isomers or linkage-types seem to be associated to specific pathologies. Moreover, multivariate data tools, such as PCA and PLS-DA, were also employed in these studies to identify glycan-based biomarker candidates. In the case of hAGP, serum samples from patients with **pancreatic ductal adenocarcinoma (PDAC)** and **chronic pancreatitis (ChrP)**, as well as healthy controls, were studied in order to find



glycan isomers that allowed the differentiation between these two pathologies. The hAGP glycan structures and linkage-types more related to PDAC were identified using this approach and the potential of certain isomers to discriminate between PDAC and ChrP patients was demonstrated. Similarly, the modifications occurred in mTf glycan isomers were evaluated in mice suffering from **collagen-induced arthritis (CIA)**. CIA is an inflammatory and autoimmune disease that resembles human **rheumatoid arthritis (RA)** in terms of disease course, histological findings and its response to commonly used anti-arthritic drugs. Certain glycan isomers were identified as meaningful to distinguish CIA from control mice, being most of them highly sialylated and branched structures, and their overexpression seemed also to be related to CIA progression. Both studies pointed out the great utility of this combined methodology in patho-glycomics to find novel glycan-based biomarkers.

**Aims**

---



The main purpose of this thesis is the development of analytical methodologies for the separation and characterization of **glycans and glycopeptides** from biologically relevant glycoproteins, using **capillary liquid chromatography**, as well as **capillary electrophoresis**, both hyphenated to mass spectrometry (**CapLC-MS** and **CE-MS**, respectively). Additionally, ion mobility mass spectrometry (IM-MS) has been also evaluated, as an alluring platform for glycan isomer separation.

In this thesis we have focused on the study of human **alpha-1-acid glycoprotein (hAGP)** and **mouse transferrin (mTf)**, whose normal glycosylation pattern has been described to be altered in several pathological processes. Moreover, **recombinant human erythropoietin (rhEPO)** has been selected as model glycoprotein, not only due to its relevance in biopharmaceutical field and doping analysis, but also due to the presence of N- and O-glycosylation and its high microheterogeneity.

A more extended view of the main points of this thesis can be found below:

- Evaluation of a simple and rapid acetone precipitation method for glycopeptide purification by CE-MS and study of the physicochemical parameters and properties that make peptides, O- and N-glycopeptides to show a different precipitation trend.
- Establishment of an **on-line solid-phase extraction capillary electrophoresis mass spectrometry (SPE-CE-MS)** method, using a titanium dioxide sorbent, for the analysis of glycoprotein digests in order to clean-up and enhance glycopeptide sensitivity.
- Optimization of a CapLC-MS method with the aim of improving the separation and sensitivity of hAGP glycans and their corresponding isomers.

- Sialic acid and fucose linkage-type characterization of hAGP glycan isomers by exoglycosidase digestion in combination with **glycan reductive isotope labeling (GRIL)** strategy using [ $^{12}\text{C}_6$ ]/[ $^{13}\text{C}_6$ ]aniline. Optimization of each digestion procedure to perform a reliable assignment.
- Development of a **CapLC tandem mass spectrometry (CapLC-MS/MS)** approach as a confirmatory and complementary method for the characterization of hAGP glycan isomers. Identification of the ion fragments with higher diagnostic value for structural and linkage-type assignments.
- Assessment of the potential of IM-MS for the separation of isomeric glycans differing in the sialic acid linkage-type using hAGP as model glycoprotein.
- Study of hAGP glycans in healthy controls and patients with **pancreatic ductal adenocarcinoma (PDAC)** and **chronic pancreatitis (ChrP)** using the developed CapLC-MS method and the GRIL strategy for relative quantification. Evaluation of **principal component analysis (PCA)** and **partial least squares discriminant analysis (PLS-DA)** as powerful tools for the identification of hAGP glycan isomers that could be considered biomarker candidates of PDAC.
- Optimization of a mTf purification method by **immunoaffinity chromatography (IAC)** and application of the previous CapLC-MS method and chemometric analysis for the study of mTf glycan isomers in mice with **collagen induced arthritis (CIA)**, an inflammatory disease homologous to human **rheumatoid arthritis (RA)**.

## **Chapter 1. Introduction**

---



## 1.1. Glycosylation

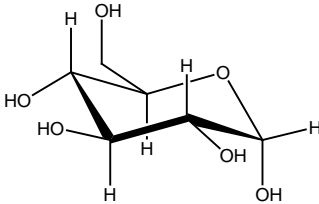

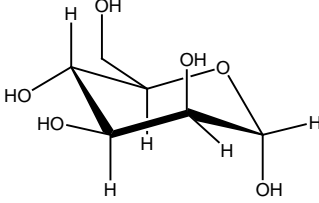

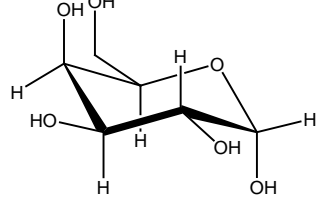

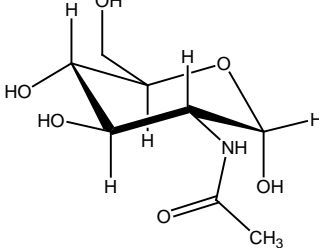

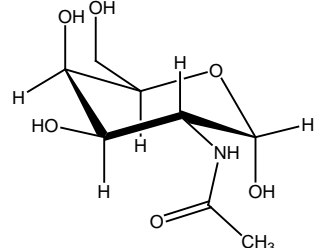

Glycosylation is the covalent addition of carbohydrates, also known as glycans, to the polypeptide chain of a protein, which can take place co-translationally or as a post-translational modification mediated by enzymes. Glycosylation is considered the most frequent modification in proteins, estimating that over half of the proteins in nature are glycosylated [1,2]. Moreover, it plays a crucial role in determining the function and physicochemical properties of proteins as glycans are involved in several important biological processes such as cell-cell interaction, signalling and protein folding [3–5]. It is also worth mentioning that glycans can adopt several structures, from a simple monosaccharide unit to a branched polysaccharide, generating structural diversity in proteins. Glycosylated proteins, commonly known as glycoproteins, are synthesised in the endoplasmic reticulum and the Golgi apparatus in a complex process mediated by the sequential activity of at least thirteen glycosyltransferases, enzymes that catalyse the transfer of sugar moieties from activated donor nucleotides or lipids to acceptor glycan structures. [6–8]. Whereas protein synthesis follows a well-defined, genetically encoded linear process, glycosylation is a non-template-driven process. Various competing reactions in the processing pathways, plus the need for enzyme, acceptor and substrate concurrence, as well as other physiological factors contribute to glycoprotein microheterogeneity. In this sense, each glycoprotein is composed by several *glycoforms*, heterogeneous mixtures of glycoconjugate species that differ in the amount, size and/or structure of the attached glycans. Glycoforms of a given glycoprotein can range from less than five to more than one hundred.

Glycan chains are composed of well-known monosaccharide units covalently bound to each other by glycosidic bonds. The full diversity of mammalian glycans (estimated to be  $\geq 7000$  structures) are assembled from only ten different monosaccharides [2]. **Table 1.1** shows the most common monosaccharide units that can be found in protein glycan structures, with the

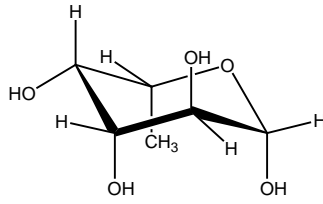


standardized symbol nomenclature proposed by the Consortium for Functional Glycomics (CFG) [9].

**Table 1.1.-** Most common monosaccharide units found in protein glycans. Symbol nomenclature follows the rules of the Consortium for Functional Glycomics (CFG) [9]. <sup>a</sup>In parenthesis the abbreviation of the main group in which the monosaccharide belongs to is indicated: H: hexose; N: N-acetylhexosamine; F: fucose; S: sialic acid.

| Monosaccharide        | Structure   | Abbreviation <sup>a</sup> | Symbol  |
|-----------------------|---|---------------------------|---|
| Glucose               |    | Glc (H)                   |    |
| Mannose               |   | Man (H)                   |   |
| Galactose             |  | Gal (H)                   |  |
| N-acetylglucosamine   |  | GlcNAc (N)                |  |
| N-acetylgalactosamine |  | GalNAc (N)                |  |

Fuco

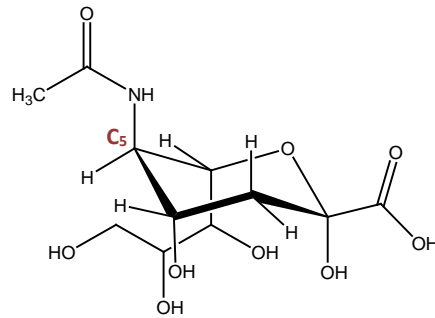


Fuc (F)



Sialic acids (SiA)

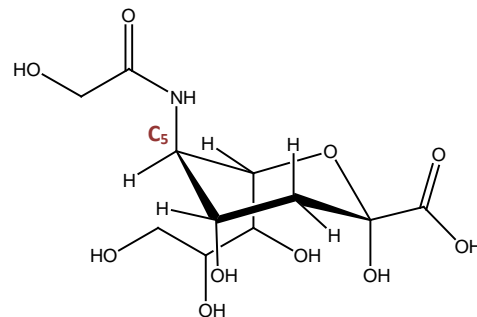
N-acetylneuraminic acid



NeuAc (S)



N-glycolylneuraminic acid



NeuGc (S)



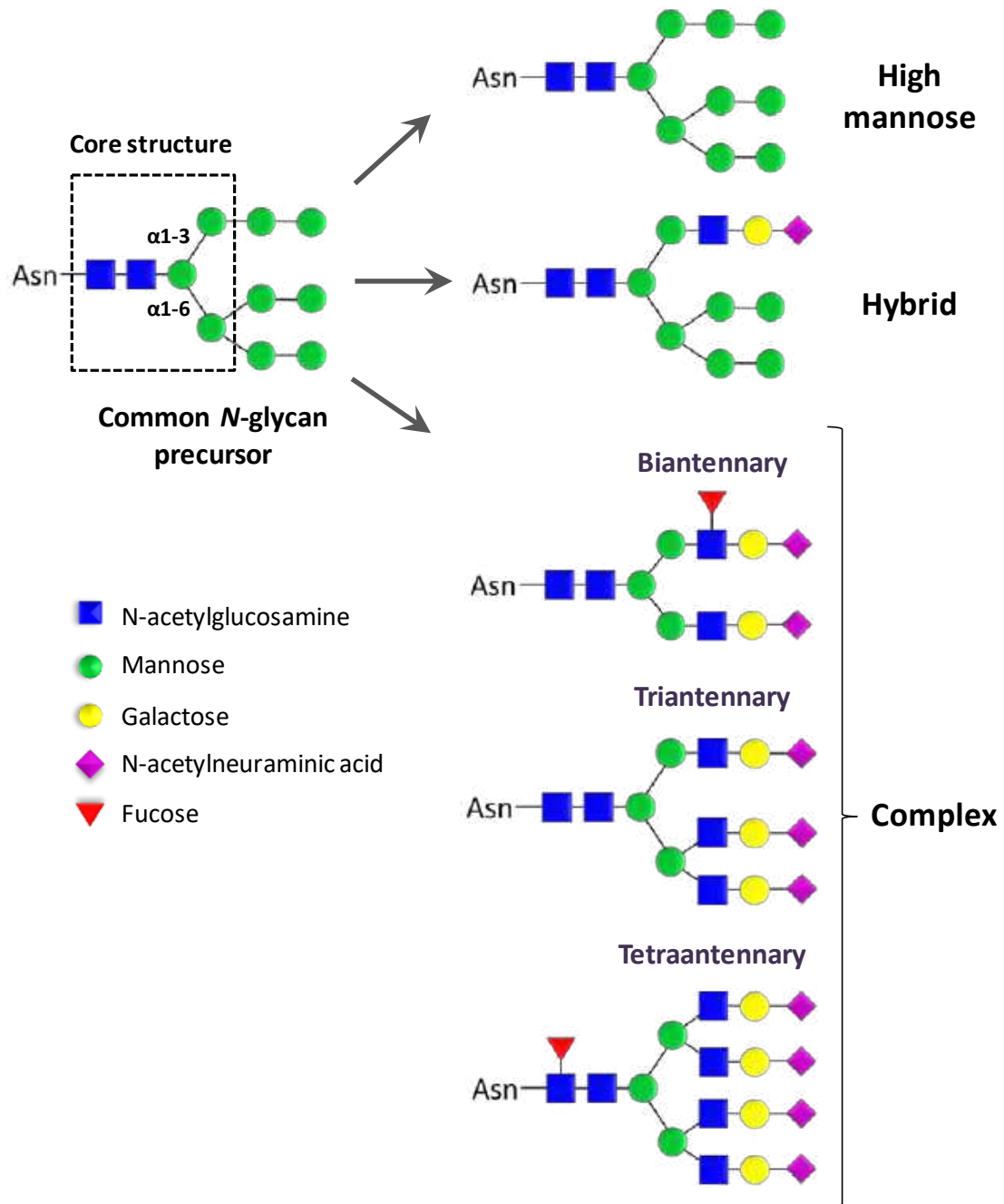
Glycosylation can be mainly classified in O- and N-glycosylation depending on the amino acid where the glycan is bound (*glycosylation site*) [10,11]. In O-glycosylation, glycans are linked to a serine (Ser) or a threonine (Thr) in the polypeptide chain of a protein through the oxygen atom of the hydroxyl group (O-glycosidic bond). O-linked glycans can range from a single monosaccharide unit to highly branched structures [12,13]. In the case of N-glycosylation, peptide sequence determines the potential glycosylation sites, and glycans are only attached to asparagines (Asn) in the sequon containing Asn-X-Ser/Thr, where X corresponds to any amino acid except proline (Pro), by the nitrogen atom of the amide group (N-glycosidic bond)

[14,15]. All N-glycans share a common structure, commonly known as *core* (**Figure 1.1**), and several monosaccharide units can be linked to this core to build different branched structures.

The different *N*-glycan structures can be classified in three main groups [16]:

- a) **High mannose**: only mannose residues are attached to the core of the glycan.
- b) **Complex**: N-acetyllactosamine units (LacNAc: N-acetylglucosamine plus galactose) are bound to the mannoses of the core, resulting in further branched structures, commonly denominated antennas. These antennas usually end with a sialic acid residue and can form bi-, tri- and tetraantennary structures.
- c) **Hybrid**: can be considered a combination of the two other N-glycan types, formed by only mannose residues in the  $\alpha$ 1-6 arm of the core and addition of LacNAc units, with or without sialic acid, in the  $\alpha$ 1-3 arm of the core.

The different types of *N*-glycans are presented in **Figure 1.1**, following the symbols recommended by the CFG [9].

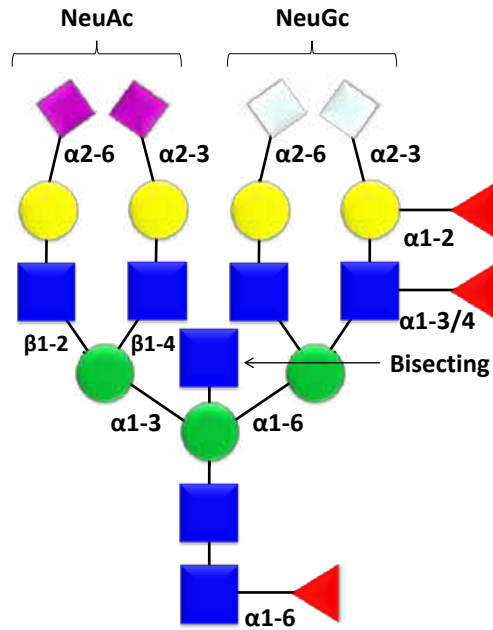


**Figure 1.1.-** Different types of *N*-glycan structures. Symbol nomenclature follows the rules of the Consortium for Functional Glycomics (CFG) [9].

As the main type of glycans studied in this thesis correspond to complex *N*-glycans, special attention to their structural diversity is paid in this section. As mentioned before, complex *N*-glycans can form highly branched structures with multiple antennae. Moreover, they usually show a sialic acid residue (SiA), also known as neuraminic acid, at the end of each antenna.

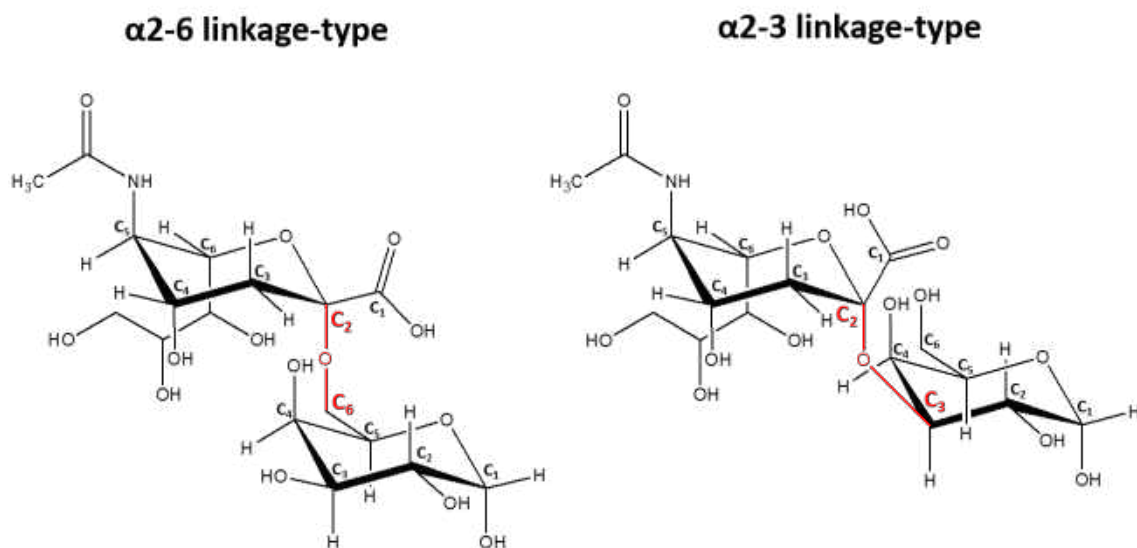
Sialic acids are widely distributed as terminal monosaccharides that coat the eukaryotic cell surface. By virtue of their negative charge at physiological pH and hydrophilicity, sialic acids have many structural and modulatory roles, getting involved in various biological processes, including cell-cell communications, inflammation, immune defense, and cancer metastasis [17–19]. Sialic acids are commonly modified in most glycans, presenting several natural variations. The most common sialic acid in humans is the so-called N-acetylneuraminic acid (NeuAc), where an N-acetyl group is added in carbon 5 to the sialic acid (**Table 1.1**), whereas the N-glycolylneuraminic acid (NeuGc), with the addition of a hydroxylated N-acetyl group, is the typical sialic acid present in other animals [20].

Apart from the composition of complex N-glycans, the configuration as well as the connectivity of the glycosidic bonds between the monosaccharides increases even more their heterogeneity [21]. Configuration refers to the stereochemistry of the glycosidic bonds, which can be  $\alpha$  or  $\beta$ . With regard to connectivity, different linkage-types can be formed between monosaccharides. By way of example, a tetraantennary complex *N*-glycan with different linkage possibilities is depicted in **Figure 1.2**. As can be observed, terminal monosaccharides, such as sialic acids and fucoses, can be bound with different linkage-types.



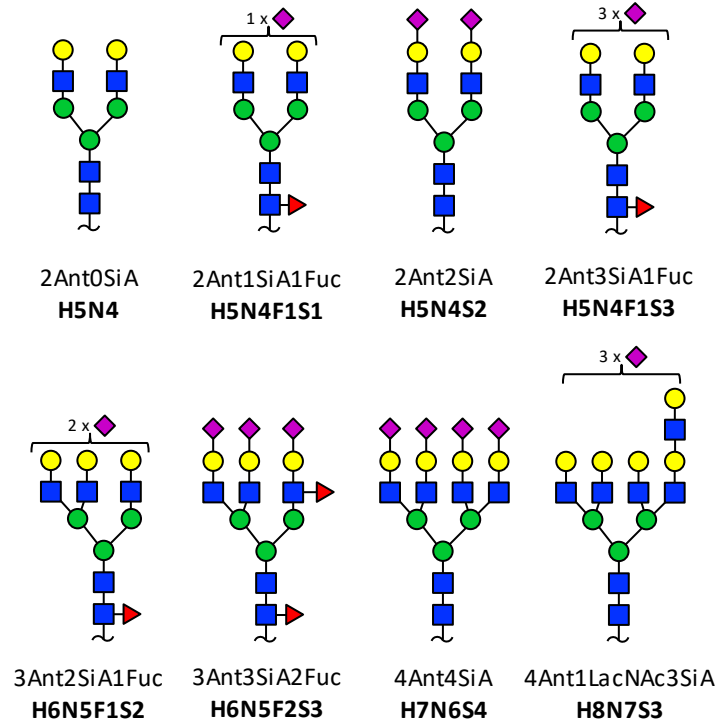
**Figure 1.2.-** Possible structural diversity of complex N-glycans depicted in a tetraantennary glycan (H7N7F3S4) as an example.

In the case of sialic acids, they are usually  $\alpha 2-3$  or  $\alpha 2-6$  linked to galactose (see **Figure 1.3**), or  $\alpha 2-8$  linked to another sialic acid in polysialic acid structures. Regarding fucoses, they can be  $\alpha 1-3$  and  $\alpha 1-4$  linked to N-acetylglucosamine on an external branch (Fuc Antenna),  $\alpha 1-2$  linked to galactose or  $\alpha 1-6$  linked to the core N-acetylglucosamine (Fuc Core), as shown in **Figure 1.2**. On the other hand, an additional N-acetylglucosamine residue can be bound to the central mannose of the core, resulting in a bisecting glycan (see **Figure 1.2**). All these possible variations in complex N-glycans contribute to substantially increase the microheterogeneity of glycoproteins and lead to different glycan isomers with identical atomic composition and mass (i.e. *isobaric isomers*).



**Figure 1.3.-** Main sialic acid linkage-types present in complex type N-glycans.

Regarding glycan nomenclature, the classification indicating the number of antennas (branching) as well as the number of sialic acids and fucoses (e.g. 2Ant2SiA1Fuc) has been widely used. However, in the last few years, this nomenclature has fallen into disuse, and has been replaced by a more standardized one that only refers to the number of each monosaccharide present in the glycan structure (e.g. 2Ant2SiA1Fuc glycan would be replaced by H5N4F1S2). For the sake of consistency throughout this thesis, only the simplified nomenclature will be used. By way of an example, **Figure 1.4** presents some complex N-glycans expressed in both nomenclatures.



**Figure 1.4.-** Different examples of complex type N-glycans with the former and novel (bold) nomenclature. For the sake of simplicity, the SiA depicted in this picture is only the NeuAc.

## 1.2. Glycoproteins as biomarkers and biopharmaceuticals

In the last decades, the analysis of glycoproteins has aroused great interest because they mediate many important biological processes being their glycosylation mainly responsible of the stability and proper function of the proteins [5]. Moreover, the glycosylation profile of a given protein can be altered as response of several biological and pathological processes. Alterations in protein glycans have been described in many diseases such as congenital disorders of glycosylation, inflammatory processes and cancer [22–24]. Similarly, specific glycan structures as well as linkage-types, which led to isobaric isomers, have been related to certain diseases. Therefore, glycoproteins have been widely used with diagnostic purposes in different pathological processes. Special importance should be paid to acute phase proteins, plasma proteins which concentration is adaptively regulated in response to most forms of inflammation, infection and tissue injury [25,26]. On the other hand, recombinant



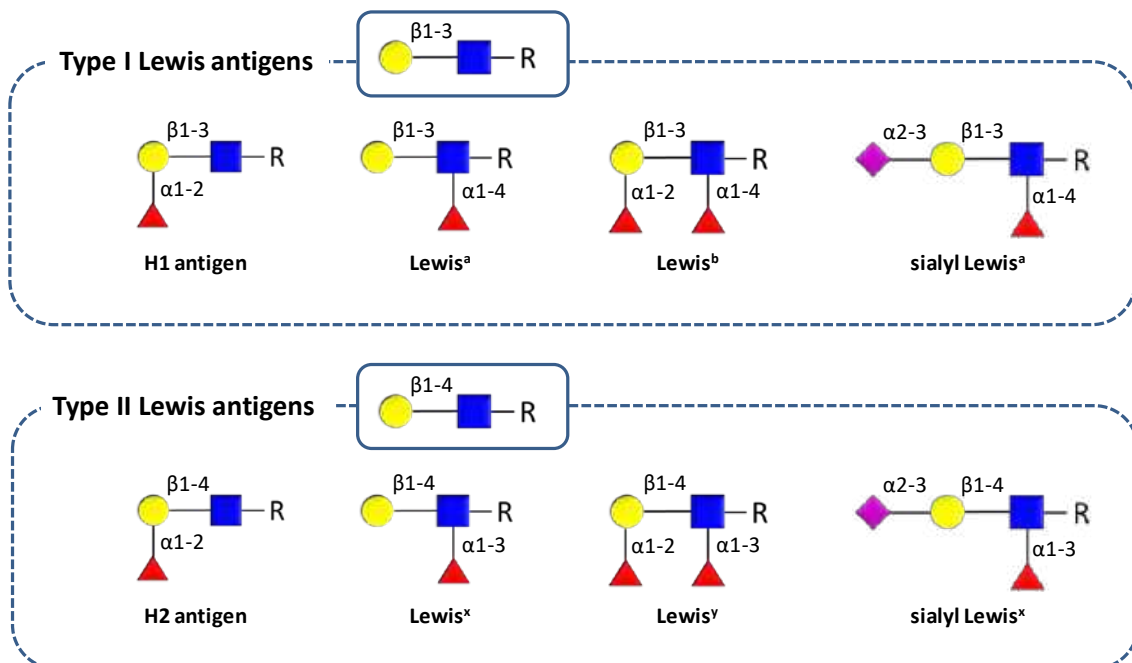
glycoproteins are frequently used as biopharmaceuticals. The glycosylation pattern of recombinant glycoproteins usually differs with respect to the endogenous one because could be affected by several production aspects such as the culture media or the purification method used. As some authors have demonstrated differences in the biological activity and pharmacokinetic of certain glycoforms, control over the production process have been demanded [27,28]. Hence, taking into account all afore-mentioned circumstances, characterization of the glycan profile in both biomarker and biopharmaceutical fields is crucial to properly assess the glycoprotein role in the pathological state or in the therapeutic treatment.

This thesis is mainly focused on the development and optimization of different analytical methodologies for the analysis and characterization of protein glycans and their corresponding isomers described to be altered in important diseases such as inflammatory processes and cancer. Specifically, human alpha-1-acid glycoprotein (hAGP) and mouse transferrin (mTf) have been studied to find novel glycan-based biomarkers of pancreatic ductal adenocarcinoma (PDAC) and collagen-induced arthritis (CIA), respectively. Furthermore, methods to enrich the glycopeptide glycoforms of recombinant human erythropoietin (rhEPO) were also developed in this thesis due to its great interest as biopharmaceutical and doping agent.

### **1.2.1. Cancer. Pancreatic ductal adenocarcinoma**

Cancer is considered the leading cause of death in wealthy countries. In recent years, several glycomic studies have been addressed in cancer research given the glycosylation role in various cancer mechanisms and as glycosylated species provides a set of targets for diagnostic application and therapeutic strategies [22,29,30]. In human tumours, the cancer-specific N-glycan alterations include: accumulation of high-mannose glycans, due to premature termination of glycan processing, reduction in bisecting glycans, increased branching and

terminal modifications such as sialylation and fucosylation. One of the most widely occurring cancer associated changes is abnormal sialylation, which is often driven by the altered expression of sialyltransferase enzymes and is related to poor patient prognosis and metastasis [31,32]. Although several sialyltransferase enzymes are implicated in cancer, ST6GAL1, which catalyses the addition of  $\alpha$ 2-6 linked sialic acids onto terminal *N*-glycans, has become increasingly dominant in the literature [33,34]. ST6GAL1 was described to be upregulated in many cancer types, including pancreatic, prostate, breast and ovarian cancer, having a key role in tumour aggression and metastasis. Similarly, overexpression of fucosylation has been also associated with cancer [35,36]. Several studies have suggested that monitoring serum or tissue fucose levels could be a promising approach for the early diagnosis and prognosis of several malignancies. The addition of fucose residues to complex glycans, catalysed by fucosyltransferase enzymes, led to the formation of Lewis antigens [37,38], **Figure 1.5**.



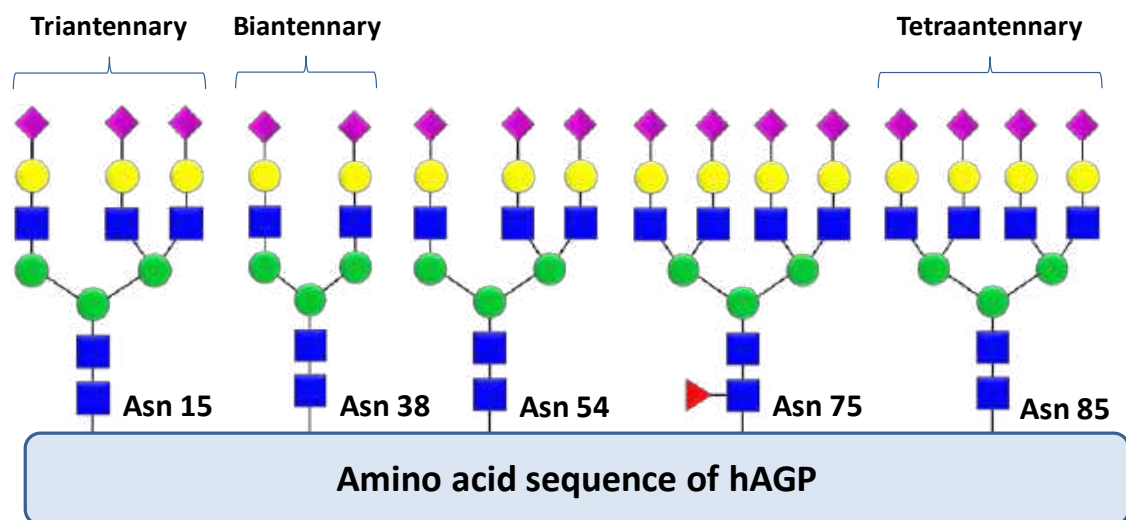
**Figure 1.5.-** Terminal oligosaccharide structures determining the Lewis antigens. Symbol nomenclature follows the rules of the Consortium for Functional Glycomics (CFG).

As these structures promote extravasation of circulating tumour cells, they are considered highly related to metastasis. In addition, the expression of some of these structures has been described to increase considerably during acute and chronic inflammatory processes. In particular, sialyl Lewis<sup>a</sup> (SLe<sup>a</sup>) epitope, which is detected by the serological assay CA19-9, is a cancer-associated marker widely used in the clinical practice, especially in patients with an established diagnosis of pancreatic, colorectal, gastric or biliary cancer [39,40]. In the case of pancreatic ductal adenocarcinoma (PDAC), CA19-9 is currently the only clinical biomarker used, which, although widely used for disease monitoring, does not provide adequate accuracy for early detection and diagnosis [41,42]. PDAC is one of the world's most aggressive malignancies with a five year survival rate of less than 5%, the worst prognosis among all cancers [43]. The poor survival rate could be attributed to its own tendency to metastasise at an early stage, but also to the lack of a reliable early detection method. In addition to the increase in SLe<sup>a</sup> levels, a wide range of alterations in other glycan structures have been described in PDAC, including increases of sialyl Lewis<sup>x</sup> (SLe<sup>x</sup>) antigen and overexpression of branched and fucosylated N-glycans [44]. These studies reveal the biological significance of glycans in PDAC and therefore, the clinical value of exploiting aberrant glycosylation to improve the diagnosis and treatment of this deadly disease.

#### **1.2.1.1. Human alpha-1-acid glycoprotein**

Human alpha-1-acid glycoprotein (hAGP), named as orosomucoid, is a positive acute-phase protein that is mainly biosynthesized and secreted by hepatocytes. It is the main component of seromucoid proteins fraction at concentrations between 0.55 to 1.4 mg·mL<sup>-1</sup>, although hAGP serum levels increase up to four times as a result of an infection or inflammation process [45]. hAGP is composed by a single polypeptide chain of 183 amino acids, with a molecular weight of 41-43 kDa, and 45% carbohydrate content as complex-type N-glycans attached to five glycosylation sites (Asn-15, -38, -54, -75, -85). It exhibits an unusually high degree of glycan

branching, being one of the few plasma proteins possessing not only bi- and triantennary, but also tetraantennary glycans [46]. Moreover, this degree of branching indirectly results in a high sialic acid composition, which contributes significantly to the low isoelectric point (pI) of the protein, which ranges from 2.8 to 3.8 due to the existence of multiple isoforms [47]. hAGP glycans can also present fucose units, which can be linked to an external branch, as well as to the core. Despite the high number of potential glycoforms of hAGP, due to the huge structural variability of glycans, only 12-20 different glycoforms are expressed in non-pathological conditions [46]. The number of feasible glycoforms is reduced since the Asn residues are selective for the type of glycans they express. It was also found that the majority of hAGP glycans correspond to tri- or tetraantennary structures (85-90%) and only the remaining 10-15% are biantennary glycans. **Figure 1.6** shows a general schematic illustration of a possible hAGP glycoform.



**Figure 1.6.-** General schematic illustration of an hAGP glycoform with possible glycan structures attached to each glycosylation site.

Although the specific biological role of hAGP has yet to be clearly defined, it is considered a natural anti-inflammatory and immunomodulatory agent, which concentration and glycosylation profile can change under various physiological and pathophysiological conditions

[45,48]. It has been suggested this glycoprotein to play an important role in the tumour microenvironment, through affecting the immune modulation, physiology, drug resistance and cancer progression and metastasis [49]. Although the concentration of hAGP alone is not diagnostic for a particular pathological condition, the altered glycosylation profile of hAGP in different diseases, could provide an alternative biomarker target. Therefore, investigations into changes in its glycan profile during different diseases have been, and are still, of great interest for the research community. In this regard, an increase of sialylation, fucosylation and branching of hAGP glycans was reported in several carcinomas, including breast, pancreas, lung and liver cancer [50–52]. Similarly, formation of SLe<sup>x</sup> epitope, which contributes to tumour cell migration towards distant tissues and metastasis, was also described in hAGP [45,49,53].

With regard to PDAC, previous studies analysing hAGP glycans from healthy controls and patients suffering from PDAC and chronic pancreatitis (ChrP) demonstrated that some fucosylated glycans could be possible biomarker candidates of this cancer type, being upregulated only in PDAC and thus differentiating this pathology from the inflammatory benign disease (ChrP) [54]. Increased fucosylation levels in PDAC samples were attributed to an overexpression of  $\alpha$ 1-3 fucosylation, considering this linkage-type crucial in the discrimination of PDAC patients [55]. In addition, different expression of certain glycan isomers was observed in cancer samples, which could also differentiate this pathology from healthy controls and ChrP. In this thesis, exhaustive characterization of hAGP glycan isomers was performed in order to further investigate if certain structures or linkage-types could be specifically related to PDAC and therefore be used as biomarkers of this disease.

### 1.2.2. Arthritis. Collagen-induced arthritis

Inflammatory arthritic diseases are autoimmune disorders in which the host immune system self-invades the host defence mechanism, resulting in a gradual degeneration of the normal immune response and thus inflammation [56]. In particular, rheumatoid arthritis (RA) is the most common inflammatory arthritis in human population, affecting up to 1.0% of population worldwide and doubling mortality rate of RA patients compared to healthy individuals. The disease is characterized by chronic joint inflammation, swelling, tenderness and destruction of bone and cartilage [57,58]. An early diagnosis and initiation of disease therapy is necessary in order to minimize irreversible joint destruction in RA. Currently, clinical diagnosis of RA is based on standardised classification criteria, as set up by the American College of Rheumatology and the European League against Rheumatism [59], using serological tests for rheumatic factor (Rf) and anti-citrullinated protein antibodies (ACPAs). Rf and ACPAs are autoantibodies that can be present in serum long before onset of clinical symptoms in RA. However, positive Rf serum levels can also result from other diseases or occur in healthy individuals. Thus, the sensitivity and selectivity of Rf and ACPAs in the diagnosis of RA is rather low and complementary disease-specific biomarkers are required. In this regard, glycosylation changes have been reported for several acute-phase proteins isolated from the sera of RA patients, including haptoglobin, transferrin (Tf) and AGP, which present interesting and promising basis for the discovery of new biomarkers and novel therapies for this disease [60–62]. Increase of sialylation, fucosylation and branching as well as glycans enriched in SLe<sup>x</sup> epitopes are typical characteristics for acute-phase proteins during inflammation. Therefore, further investigations addressed to study the glycan profile of such proteins in patients suffering from RA and the differences with respect to healthy individuals are necessary.

Mice models are widely used to study human diseases due to their availability, size, low cost, ease of handling, fast reproduction rate and a high identity between their genes and the

human orthologues [63,64]. Several mice models of arthritis have proven to be useful for studying the course of RA and testing new therapies. In particular, collagen-induced arthritis (CIA) is the most commonly studied autoimmune model of RA [65]. A chronic form of CIA is induced in C57BL/6 wild-type (WT) mice by immunization with chicken type II collagen in complete Freund's adjuvant (Col-II/CFA) [66]. This autoimmune arthritis in mice closely resembles human RA in terms of disease course, histological findings and also in its response to commonly used anti-arthritic pharmaceuticals. Therefore, it could be considered an adequate model to study the efficacy of novel drugs and to evaluate the glycosylation changes derived from arthritis.

#### **1.2.2.1. Mouse transferrin**

Transferrin (Tf) is a glycoprotein mainly responsible of the iron transport through the blood plasma. In humans, it is a negative acute-phase protein and its glycosylation was described to be altered in congenital disorders of glycosylation (CDGs), chronic alcoholism and other inflammatory processes [24,67,68]. In this sense, it has been reported that the serum profile of human Tf (hTf) isoforms was altered in patients with RA and levels of its trisialylated glycoforms were found to be useful as biochemical marker of the RA activity [69].

In the case of mice, mouse Tf (mTf) presents a molecular weight of around 80 kDa and shows only one N-glycosylation site at Asn-494 with complex N-glycans attached. The carbohydrate content of mTf is about 3% and sialic acids in its glycan structures correspond to NeuGc instead of NeuAc, which is mainly found in humans. **Figure 1.7** shows the peptide sequence of mTf, with the N<sub>494</sub> glycosylation site marked in blue.

| Mouse transferrin (mTf) |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|-------------------------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1                       | V | P | D | K | T | V | K | W | C | A | V | S | E | H | E | N | T | K | C | I | S | F | R | D | H | M | K | T | V | L | P | P | D | G | P | R | L | A | C | V | K | K | T | S | Y | P | D | C | I | K | A | I | S | A | S | E | A | D | A | M |
| 61                      | T | L | D | G | G | W | V | Y | D | A | G | L | T | P | N | N | L | K | P | V | A | A | E | F | Y | G | S | V | E | H | P | Q | T | Y | Y | A | V | A | V | V | K | K | G | T | D | F | Q | L | N | Q | L | E | G | K | K | S | C | H | T |   |
| 121                     | G | L | G | R | S | A | G | W | V | I | P | I | G | L | L | F | C | K | L | S | E | P | R | S | P | L | E | K | A | V | S | S | F | F | S | G | S | C | V | P | C | A | D | P | V | A | F | P | K | L | C | Q | L | C | P | G | C | G | C | S |
| 181                     | S | T | Q | P | F | F | G | Y | V | G | A | F | K | C | L | K | D | G | G | G | D | V | A | F | V | K | H | T | T | I | F | E | V | L | P | E | K | A | D | R | D | Q | Y | E | L | L | C | L | D | N | T | R | K | P | V | D | Q | Y | E | D |
| 241                     | C | Y | L | A | R | I | P | S | H | A | V | V | A | R | K | N | N | G | K | E | D | L | I | W | E | I | L | K | V | A | Q | E | H | F | G | K | G | K | S | K | D | F | Q | L | F | S | S | P | L | G | K | D | L | L | F | K | D | S | A | F |
| 301                     | G | L | L | R | V | P | P | R | M | D | Y | R | L | Y | L | G | H | N | Y | V | T | A | I | R | N | Q | Q | E | G | V | C | P | E | G | S | I | D | N | S | P | V | K | W | C | A | L | S | H | L | E | S | R | T | K | C | D | E | W | S | I |
| 361                     | S | E | G | K | I | E | C | E | S | A | E | T | T | E | D | C | I | E | K | I | V | N | G | E | A | D | A | M | T | L | D | G | G | H | A | Y | I | A | G | Q | C | G | L | V | P | V | M | A | E | Y | Y | E | S | S | N | C | A | I | P | S |
| 421                     | Q | Q | G | I | F | P | K | G | Y | Y | A | V | A | V | V | K | A | S | D | T | S | I | T | W | N | N | L | K | G | K | K | S | C | H | T | G | V | D | R | T | A | G | W | N | I | P | M | G | M | L | Y | N | R | I | N | H | C | K | F | D |
| 481                     | E | F | F | S | Q | G | C | A | P | G | Y | E | K | N | S | T | L | C | D | L | C | I | G | P | L | K | C | A | P | N | N | K | E | E | Y | N | G | Y | T | G | A | F | R | C | L | V | E | K | G | D | V | A | F | V | K | H | Q | T | V | L |
| 541                     | D | N | T | E | G | K | N | P | A | E | W | A | K | N | L | K | Q | E | D | F | E | L | L | C | P | D | G | T | R | K | P | V | K | D | F | A | S | C | H | L | A | Q | A | P | N | H | V | V | V | S | R | K | E | K | A | A | R | V | K | A |
| 601                     | V | L | T | S | Q | E | T | L | F | G | G | S | D | C | T | G | N | F | C | L | F | K | S | T | T | K | D | L | L | F | R | D | D | T | K | C | F | V | K | L | P | E | G | T | T | P | E | K | Y | L | G | A | E | Y | M | Q | S | V | G | N |
| 661                     | M | R | K | C | S | T | S | R | L | E | A | C | T | F | H | K | H |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |

**N** N-glycosylation site

**Figure 1.7.-** Mouse transferrin (mTf) amino acid sequence (UniProtKB / Swiss-Prot Accession Number: Q92111).

Previous studies of our research group demonstrated that mTf glycopeptide glycoforms were modified in presence of CIA, mainly observing an increase of fucosylation and branching in glycan structures [70]. Hence, further investigations are needed to deeply identify the potential alterations occurred in the glycan profile of mTf isolated from mice with CIA. These findings could be an important previous step in order to discover in the future novel glycan-based biomarkers for RA diagnosis and prognosis.

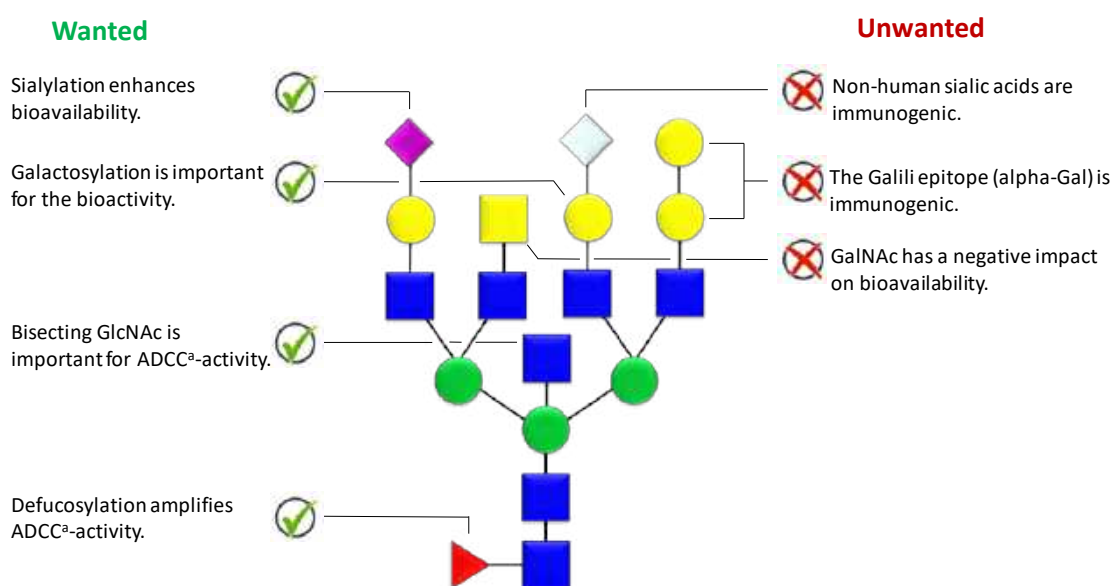
### 1.2.3. Biopharmaceuticals

Recombinant glycoproteins have become one of the most intensely studied and commercialized class of biopharmaceuticals, being the vast majority of approved drugs by the Food and Drug Administration (FDA) over the last years [71]. Due to human-compatible and consistent glycosylation is required for a safe and efficacious drug product, the glycoform profile of an approved therapeutic recombinant glycoprotein is considered a parameter that must be controlled to ensure biopharmaceutical quality [27,28,72–74]. **Figure 1.8** shows



several glycan features that would be desired or not for glycoprotein-based drug products in biopharmaceutical industry.

As can be observed, among other characteristics, special attention should be paid to sialic acids as the abundance and type of sialylation influences the clinical performance of therapeutic glycoproteins (serum half-life, immunogenicity, activity, etc) [18]. While NeuAc is found in both human and non-human cells, NeuGc is synthesized by all mammalian cells except human cells. Despite NeuGc has only one oxygen atom more than its homologue, this difference is enough to cause immunogenic response in humans. Moreover, acetylation of sialic acids could also impact clinical efficacy of biopharmaceuticals [76]. Hence, characterization of sialylation throughout the product lifecycle is also considered a regulatory requirement.



**Figure 1.8.-** Desirable and undesirable glycan features in glycoprotein-based biopharmaceuticals. <sup>a</sup>ADCC (antibody-dependent cellular cytotoxicity) is a mechanism of cell-mediated immune defense whereby an effector cell of the immune system actively lyses a target cell, whose membrane-surface antigens have been bound by specific antibodies [75].

For all these reasons, the development of methodologies for glycan characterization of recombinant glycoproteins is of great importance. In this regard, analysis of glycopeptides, obtained from enzymatic digestion of glycoproteins, is a commonly used strategy for glycan characterization of biopharmaceuticals. Nevertheless, glycopeptide purification prior to analysis continues to be a challenge due to the difficulties to detect low abundant glycopeptide glycoforms, which could be essential to differentiate endogenous and recombinant variants of certain proteins, in presence of more abundant peptides. Hence, development of reliable enrichment strategies to selectively isolate glycopeptides from protein digests before analysis is required.

### 1.2.3.1. Recombinant human erythropoietin

Human erythropoietin (hEPO) is a glycoprotein hormone of about 30 kDa composed of a single polypeptide chain of 165 amino acid residues. hEPO is a highly glycosylated protein (around 40% w/w) which shows three N-glycosylation sites with complex-type N-glycans, at Asn-24, -38 and -83, and only one O-glycosylation site in serine 126 (Ser-126) [77,78]. The peptide sequence of rhEPO is shown in **Figure 1.9** where its N- and O-glycosylation sites are marked in red and green, respectively. This glycoprotein is mainly synthesized by the kidneys and is responsible of stimulating red blood cell production (erythropoiesis) in the bone marrow, increasing oxygen supply to the tissues [79].

| Recombinant human erythropoietin (rhEPO) |                          |            |                         |            |             |            |
|--|--------------------------|------------|-------------------------|------------|-------------|------------|
| 1  | APRLICDSR                | VLERYLLEAK | EAENITTGCA              | EHCSLNENIT | VPDTKVNIFYA | WKRMEVGQQA |
| 61                                       | VEVWQGLALL               | SEAVLRGQAL | LVN <sup>SS</sup> QPWEP | LQLHVDKAVS | GLRSLTLLR   | ALGAQKEAIS |
| 121                                      | PPDAAS <sup>A</sup> AAPL | RTITADTFRK | LFRVYSNFLR              | GKCLKYTGEA | CRTGDR      |            |

N N-glycosylation site    O O-glycosylation site

**Figure 1.9.-** Recombinant human erythropoietin (rhEPO) amino acid sequence (UniProtKB / Swiss-Prot Accession Number: P01588).

Recombinant human erythropoietin (rhEPO) was firstly produced in mammalian cells using recombinant DNA technology in 1985. rhEPO is a widely used therapeutic agent for the treatment of anemias and represents one of the largest biopharmaceutical markets, which is usually produced in Chinese hamster ovary (CHO) cells [80–82]. Exogenous erythropoietins encompass a group of synthetic glycoproteins collectively called erythropoiesis-stimulating agents (ESA). ESAs are used in the treatment of anemia in chronic kidney disease, myelodysplasia and cancer chemotherapy. The rhEPOs most used as biopharmaceuticals are epoetin alfa and beta, the first ones to be synthesized. The recombinant protein is homogeneous with respect to the peptide sequence of natural hEPO, but heterogeneous with respect to the carbohydrate moieties, since the glycosylation profiles appear to differ between preparations. As mentioned before, these differences could play an important role in determining its biological activity, thus influencing the final biopotency, and are considered to be mainly related to the number of sialic acid residues at the end of the tri- and tetraantennary sugar chains [80].

On the other hand, rhEPO has also become particularly popular in the last decades due to its unlawful use as a performance-enhancing drug in endurance sport disciplines [83,84]. Its use as a doping agent was forbidden by sport authorities since 1989. The presence of about 2% of NeuGc in the glycans of rhEPO O<sub>126</sub>-glycosylation site could distinguish between recombinant and endogenous forms of hEPO in doping control. Therefore, control and characterization of the glycosylation profile of rhEPO is needed to guarantee high quality and therapeutic efficacy in biopharmaceutical industry, as well as to distinguish recombinant variants of hEPO in urine in doping analysis.

Apart from its high importance in biopharmaceutical and doping fields, rhEPO can be considered an excellent glycoprotein model to establish and optimize analytical methodologies in glycoproteomic studies as it presents both O- and N-glycosylation forms with different

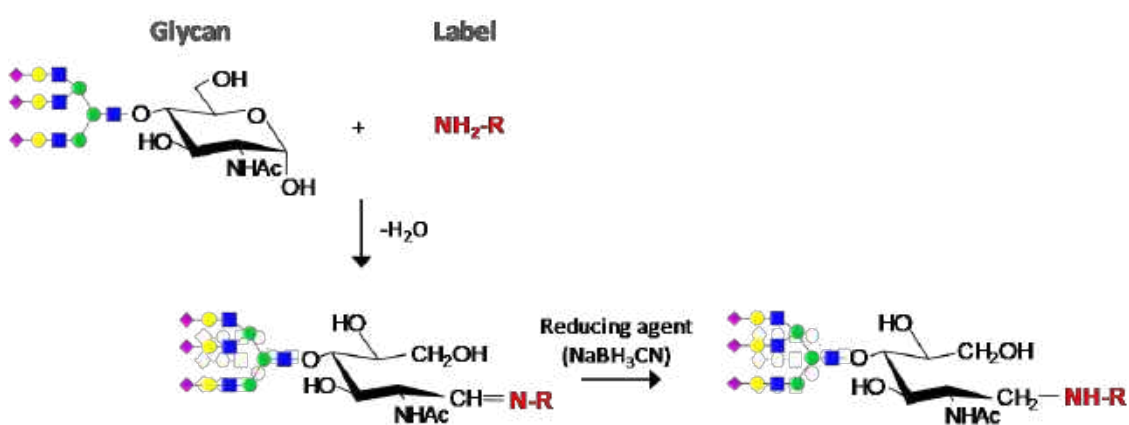
degrees of branching and sialylation. In this sense, several studies have been addressed in our research group to characterize rhEPO glycosylation, analysing the intact glycoprotein as well as the glycopeptides obtained from its tryptic digestion [85–87]. In this thesis, different strategies to purify and enrich rhEPO glycopeptides were developed due to the need to decrease their limits of detection.

### **1.3. Glycoprotein analysis**

The analysis of protein glycosylation can be conducted by three different approaches: intact glycoprotein, glycopeptides and glycans. The analysis of the intact glycoprotein can be considered the more direct and faster way as practically no sample treatment is required. However, obtaining information about the glycosylation sites and their degree of occupancy is up to now almost impossible, due to the high microheterogeneity of this type of proteins [88,89]. In addition, the ionization of such large molecules in mass spectrometry, the most common detection technique used in glycoproteomic studies, is not an easy task [90,91]. Therefore, the use of enzymes or specific reagents to obtain glycosylated structures of lower molecular mass, such as glycopeptides or glycans, has become popular in the last years. Glycopeptides, or glycosylated peptides, are obtained after digestion of the intact glycoprotein with a protease, enzyme that catalyses the hydrolysis of the peptide bonds between amino acids. Several proteases can be employed for this purpose, being trypsin the most widely used in a multitude of different studies. Trypsin is a pancreatic serine protease that cleaves peptides on the carboxyl side of lysine and arginine amino acid residues. The obtained glycopeptides after digestion provide information about the glycan structures, the glycosylation sites to which they are attached and thus about their degree of occupancy [92].

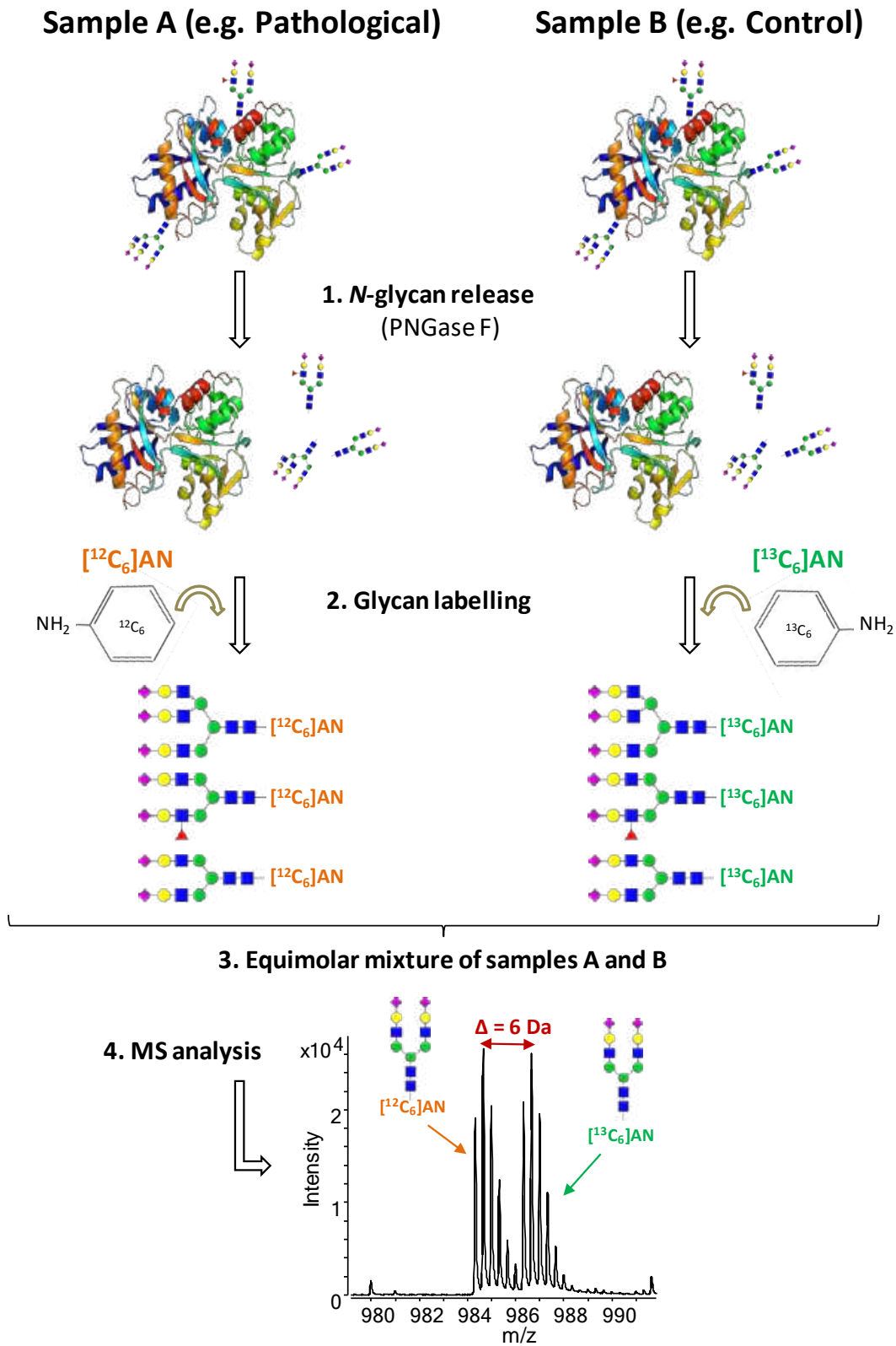
Lastly, but no less important, the analysis of glycans, which has allowed separation between isomers that differ in certain monosaccharide linkage-types, has aroused great interest in

biomarker discovery as one isomer could have substantially more biological activity than the others or could be related to an specific pathology [93–95]. N-glycans are usually released from the glycoprotein by enzymatic digestion, using an amidase such as peptide-N-glycosidase F (PNGase F), while O-glycans are commonly obtained by chemical release through reductive alkaline  $\beta$ -elimination. Although analysing glycans the information about the glycosylation sites they come from or the carrier protein is lost, their considerably different polarity allows a better isolation from the other components of the sample (i.e. peptides or deglycosylated protein) resulting in better detection sensitivity. Moreover, their detection can be improved by applying a wide variety of derivatization strategies with a large number of different labels [96,97]. The most common reaction employed to add a labelling agent to the reducing end of glycans is the *reductive amination* because is usually simple, fairly rapid and preserves glycan modifications such as sulfation and phosphorylation, which could present potential biological significance. In this reaction, a label containing a primary amine group reacts with the aldehyde group of the glycan, resulting in an imine, which is then reduced to a stable secondary amine by a reducing agent such as sodium cyanoborohydride ( $\text{NaBH}_3\text{CN}$ ), **Figure 1.10**.



**Figure 1.10.**- Representation of reductive amination reaction to add a labelling agent to the reducing end of glycans.

Several labels can be used allowing derivatization of glycan with a chromophore or fluorophore, to improve detection after chromatographic or electrophoretic separation, or to link charged or hydrophobic groups at the reducing end enhancing glycan separation and mass-spectrometric detection. The use of stable isotopic labels, such as [ $^{12}\text{C}_6$ ]-aniline/[ $^{13}\text{C}_6$ ]-aniline, for relative glycan quantitation provides qualitative and quantitative means for comparative glycomic analyses of different samples [98]. In the so-called Glycan Reductive Isotope Labelling (GRIL) strategy, equimolar mixtures of two glycan samples labeled by reductive amination with either [ $^{12}\text{C}_6$ ]-aniline or [ $^{13}\text{C}_6$ ]-aniline were analysed by mass spectrometry (MS), as illustrated in **Figure 1.11**. This method, using a 6 Da mass difference, easily resolves the glycans obtained from the two different samples (e.g. healthy control vs. pathological) and allows comparing and quantifying the changes resulted in the glycan profile of the target glycoprotein. In addition, as both glycan samples are ionised at the same time, possible variations between runs due to ionization differences, ion-suppression effects or tuning of the mass spectrometer are avoided and thus a reliable quantitation of glycans is performed. This GRIL strategy, using [ $^{12}\text{C}_6$ ]-aniline/[ $^{13}\text{C}_6$ ]-aniline, was previously established in our research group for relative quantitation of *N*-glycans obtained from different glycoproteins and afterwards applied to the analysis of patient samples suffering from ChrP and PDAC [54,99]. These studies demonstrated the potential value of this technique to detect major and, more importantly, minor modifications in the glycosylation pattern of a glycoprotein caused by certain pathologies, thus allowing the identification of potential glycan-based biomarker candidates. In this thesis, this method was widely used for glycan analysis to compare and quantify either the glycan profile before and after the digestion of glycans with specific exoglycosidases or the under- or overexpression of glycan isomers in presence of certain diseases.



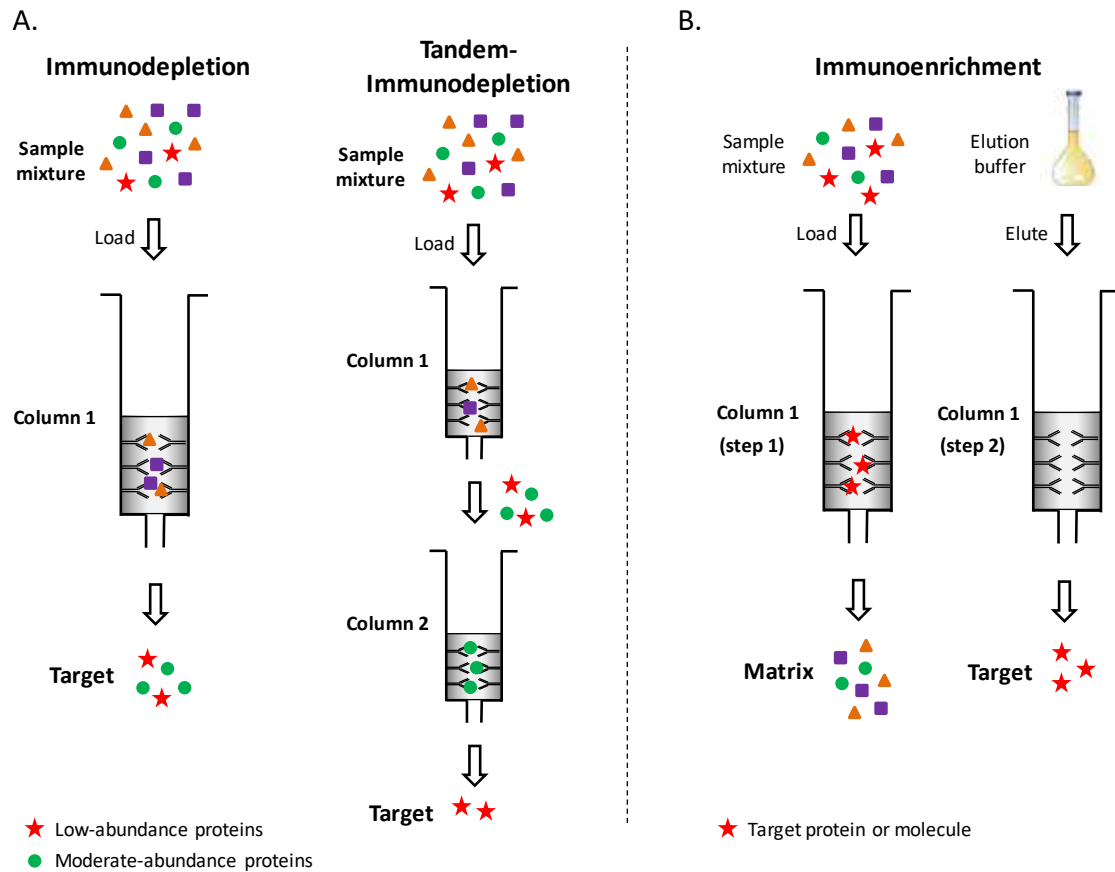
**Figure 1.11.-** Schematic illustration of the GRIL strategy steps, using  $^{12}\text{C}_6$ -aniline/ $^{13}\text{C}_6$ -aniline, for relative glycan quantitation of different samples.

### 1.3.1. Purification techniques

Human biofluids, especially blood plasma/serum and urine, can contain disease-associated proteins secreted or leaked from pathological tissues across the body and thus they are considered the most promising sources for the discovery of novel biomarkers for disease diagnosis and prognosis [100]. Moreover, biofluids are often easily obtainable through noninvasive procedures, in contrast to tissues, making them particularly attractive for large-scale clinical studies. Over the last decade, there has been tremendous interest in profiling the biofluid proteomes with the aim of finding glycoprotein-based biomarkers for various diseases [101,102]. Nonetheless, glycoproteins can be present in only minute quantities in the analysed biofluids together with other more abundant proteins, becoming mandatory their isolation and enrichment prior to analysis. Numerous attempts have been made in this sense in order to develop methods aimed at the enrichment of glycoproteins present in complex biological samples.

Immunoaffinity chromatography (IAC) approaches have become the most commonly used strategies for digging deeper into the biofluid proteomes by both global profiling and targeted analysis [103,104]. IAC represents a specific type of affinity chromatography where the stationary phase is composed of immobilized antibodies or antibody-related agents on a solid support matrix. The underlying principle of IAC is based on the selective non-covalent interaction between antibodies and their specific binding targets or antigens. The purpose of IAC separations is to enrich low abundant proteins of interest by either removing the high abundant ones from the complex samples through immunoaffinity depletion (immunodepletion) or capturing directly the low-abundance targets of interest through immunoaffinity enrichment (immunoenrichment), as illustrated in **Figure 1.12**.





**Figure 1.12.-** The main workflows of immunodepletion (A) and immunoenrichment (B). For immunodepletion, stars and circles represent low-abundance and moderate-abundance proteins, respectively. For immunoenrichment, stars represent the target molecules and other symbols are sample matrix.

The utility of IAC is highly dependent on the quality of antibodies, which should present a high intrinsic affinity towards the target protein and a reversible antibody-antigen interaction that can be easily de-stabilized favouring afterwards protein elution [105]. Two main types of antibodies are commonly used in IAC, namely polyclonal and monoclonal antibodies. Polyclonal antibodies are produced as a heterogeneous population of antibodies from multiple clones of B-cells, which can recognize and bind a variety of epitopes on a single antigen with diverse affinity. In contrast, monoclonal antibodies constitute a homogeneous population that has monovalent affinity since they bind to the same epitope on an antigen. In this thesis, IAC

has been the main affinity technique used for the purification of the studied glycoproteins due to the high isolation obtained with some few and simple steps.

On the other hand, enrichment of glycoconjugates such as glycopeptides, obtained after glycoprotein enzymatic digestion, is also considered of crucial importance in order to facilitate and improve protein glycosylation analysis in complex samples. In recent years, continuous efforts have been devoted to the development of glycopeptide enrichment and separation strategies, due to the difficulties to detect low abundant glycopeptide glycoforms and the ion suppression occurred in mass spectrometry analysis by the copresence of peptides in the glycoprotein digests. [106,107]. One of the most commonly used methodologies for glycopeptide enrichment is lectin affinity chromatography. Lectins are an extensive family of proteins of non-immune origin which present high binding capacity against certain carbohydrates and can recognize entire glycans, specific fragments and even different linkage-types between monosaccharides or glycan branching [108–111]. However, because of their high specificity, only a subset of glycopeptide glycoforms can be enriched, and a combination of different lectins is usually required. Other methods, such as hydrophilic interaction chromatography, boronate affinity chromatography and size exclusion chromatography have been proposed for glycosylation analysis. However, these methodologies lack selectivity for glycopeptides [106,112]. On the other hand, the use of titanium dioxide ( $\text{TiO}_2$ ) sorbents is one of the most promising and noteworthy methods for the purification and enrichment of glycopeptides.  $\text{TiO}_2$  has been used for a long time for phosphopeptide enrichment from proteolytic digests [113,114] but more recently it has been described for the enrichment of sialic acid containing glycopeptides [115,116]. The mechanism of glycopeptides retention is mainly attributed to interaction between the negatively charged carboxyl groups with the positively charged  $\text{TiO}_2$  surface [117]. In this regard, sialic acids can form a multipoint binding to  $\text{TiO}_2$  through both hydrophilic interactions and ligand-exchange mechanisms. Nonetheless, its main drawback is that non-sialylated glycopeptides could be lost using this purification

method. Likewise, other simple methods, based on the different solubility between non-glycosylated and glycosylated peptides in acetone, have also been described for glycopeptide enrichment [118]. Some of the above mentioned glycopeptide enrichment strategies have been explored in this thesis with the aim of establishing a rapid, simple and highly selective methodology to detect minor glycopeptide glycoforms that could be important in biopharmaceutical assays and doping control.

### **1.3.2. Liquid chromatography**

High-performance liquid chromatography (HPLC), commonly referred as liquid chromatography (LC), has become the most important technique for the highly efficient separation, identification and quantitation of several compounds in complex samples and has experienced a worldwide expansion with thousands of applications in the last decades [119–122]. LC mainly utilizes a column that holds packing material (stationary phase) and a pump that moves pressurized liquid solvent, known as mobile phase. The difference in the distribution of species between the two phases leads to the separation of the sample components and it depends on the physicochemical properties of the compounds being analysed (e.g. polarity, size, charge, among others) and the stationary and mobile phases used.

Currently, there are several LC stationary phases commercially available, including reversed phase, normal phase, hydrophilic interaction, size exclusion, ion exchange and affinity LC, which show different mechanisms of interaction. Reversed phase liquid chromatography (RP-LC) and, in particular, C<sub>18</sub> and C<sub>8</sub> columns, were the most commonly used chromatographic stationary phases in glycoproteomic studies, because of their commercial availability in different sizes and their relative low cost [123,124]. However, highly hydrophilic and uncharged species, such as glycans, are not properly retained using RP-LC. Hence, the use of hydrophilic interaction liquid chromatography (HILIC) stationary phases have been proposed as

a powerful alternative for glycomic studies [125–127]. Glycan separation by HILIC is achieved through partitioning between mobile phase rich in organic solvent, mostly acetonitrile, and a thin water layer immobilized on the polar stationary phase. In addition, dipole-dipole and electrostatic interactions may also contribute to the separation mechanism. Several quite different HILIC stationary phases have been introduced including silica, amino, amide, cellulose, and cyclodextrin columns. More recently, zwitterionic type HILIC (hereafter called ZIC-HILIC) columns, with sulfobetaine functional groups, were described to be useful and applicable for separations of sialylated N-glycopeptide isomers as well as sialylated N-glycan isomers [128–130]. The mechanism underlying the isomeric separation of such glycoconjugates is thought to be based on hydrophilic and electrostatic interactions between the glycans and the sulfobetaine stationary phase.

New significant advances in instrumentation and column technology have given rise to different types of LC where column dimensions, particle sizes, sample handling, pressure ranges and flow rates are considerably modified to meet with the new challenges and demands of users in different fields, including glycomics [131–133]. In this regard, capillary liquid chromatography (CapLC) and nano liquid chromatography (nanoLC) have been extensively used for sample-constrained applications such as biomarker discovery in proteomic studies [134–136]. CapLC have been the main LC mode used during this thesis for the analysis of glycan isomers using a ZIC-HILIC column. The most important advantage of using micro-flow through capillary columns is the higher sensitivity that can be achieved since the reduction in the column i.d. results in reduced dilution of the chromatographic band during analysis. This is very important when determining compounds present at low concentrations in limited sample volumes. Moreover, capLC provides many other advantages including less mobile phase consumption, good separation efficiency, short analysis time, instrumental simplicity, full automation and greater robustness than nanoLC or microchip technologies. It is also worth mentioning that the microliter flow rates used in capLC are ideally suited for direct, splitless

coupling with electrospray ionisation mass spectrometry (ESI-MS), in contrast to the special nanoElectrospray (nanoESI) interface often required to couple nanoLC with mass spectrometry.

### **1.3.3. Electrophoresis**

Electrophoresis is a general term that describes the migration and separation of charged species through a matrix under the influence of an electric field. Separation by electrophoresis is based on the different mobility of ions depending on their charge to mass ratio, the higher charge per unit of mass the faster the migration. Electrophoresis encompasses a wide group of analytical techniques, which uses different matrices and conditions for separation of charged compounds, some of them with relevant application in glycoconjugate analysis.

#### **1.3.3.1 Gel electrophoresis**

Gel electrophoresis is used to separate and identify macromolecules as DNA, RNA or proteins based on size, form or isoelectric point. Gel electrophoresis with a polyacrylamide matrix, commonly called polyacrylamide gel electrophoresis (PAGE), is undoubtedly one of the most widely used techniques to characterize complex protein mixtures [137]. It is a convenient, fast and inexpensive method, which only requires the order of micrograms quantities of protein. Polyacrylamide forms a gel matrix that serves as a sieve, slowing the rate of migration of larger molecules when they travel through the pores in the matrix in response to an electrical field. Other influences on the rate of migration include the structure and charge of the proteins. The use of sodium dodecyl sulfate (SDS) in the polyacrylamide gel (SDS-PAGE) largely eliminates the influence of the structure and charge, and proteins are solely separated based on polypeptide chain length. SDS is a surfactant with strong protein-denaturing effect that unfold proteins into a linear shape giving them a net negative charge, which allows proteins to

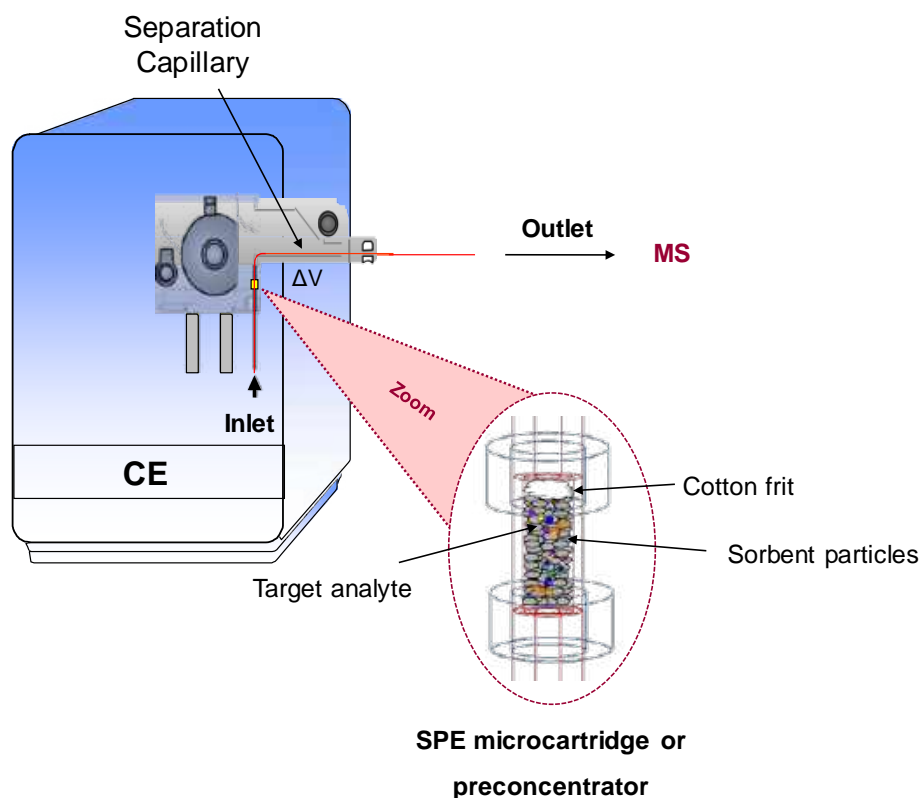
migrate through the gel in direct relation to their size. SDS-PAGE, first described by Laemmli [138], can be used to estimate the relative molecular mass of proteins, the relative abundance of proteins in a sample and the distribution of proteins among fractions, as well as the purity of protein samples [139].

### **1.3.3.2 Capillary electrophoresis**

Capillary electrophoresis (CE) has proved to be a very attractive alternative separation technique to LC due to its high-resolution capabilities. Owing to the multiple separation modes that can be used in CE, it can be considered a versatile and suitable technique for numerous applications [140]. Among the different modes of operation, Capillary Zone Electrophoresis (CZE) is the most common due to its simplicity and versatility. For this reason, the use of only CE to refer to CZE has been extensively accepted. In CE, in contrast to gel electrophoresis, separation takes place in a narrow fused silica capillary filled with a background electrolyte (BGE). Charged species migrate through the capillary on the basis of their different charge to mass ratio by the effect of an applied electric field. CE offers many advantages over a variety of other analytical methods, including an extremely high separation efficiency, simple operation, short analysis time, automated and reproducible analysis, very low consumption of sample and solvents and on-line direct coupling with ESI-MS [141,142]. The flow of the BGE represents only a few tenths nL/min and the reduced internal diameter of the capillary (usually between 50 and 75  $\mu\text{m}$ ) enables the formation of thin droplets at the capillary outlet. Such properties are particularly favourable to the ESI process, providing optimal ionization efficiency and a good sensitivity in mass spectrometry. Furthermore, injected sample volumes are in the femtoliter to nanoliter range, which make CE suitable for biological sample applications. Taking into account all aforementioned advantages, CE has emerged in last years as a powerful approach also for glycoconjugate analyses [143–145].

#### 1.3.3.2.1 On-line preconcentration

Despite all advantages that CE presents, the limits of detection (LODs) are constrained by the dimensions of the capillary, whose small volume limits the total volume of sample that can be injected. Moreover, when using optical detectors, the optical pathway is limited. Many different techniques have been developed to improve the concentration sensitivity in CE [146–150]. In general, the concentration sensitivity of CE can be enhanced by electrophoretic or chromatographic preconcentration. Electrophoretic preconcentration techniques are designed to compress analyte bands within the capillary, thereby increasing the volume of sample that can be injected without losing CE efficiency. This on-line sample preconcentration, generally referred to as stacking, is based on either the manipulation of differences in the electrophoretic mobility of analytes at the boundary of two buffers with differing resistivities. Stacking or isotachopheresis (ITP) approaches for electrophoretic preconcentration have been widely used, obtaining in some cases highly preconcentration factors [147,148]. Nonetheless, these methods generally depend on the analysed compounds and the sample matrix physicochemical properties. Therefore, their performance in many applications can be limited. As an alternative, chromatographic preconcentration techniques using on-line solid-phase extraction capillary electrophoresis (SPE-CE) have shown wider applicability and better reproducibility without compromising the high preconcentration factors [150–152]. Moreover, SPE-CE can be used for on-line sample clean-up and purification, with minimum sample handling. In the typical SPE-CE configuration, a microcartridge (**Figure 1.13**), filled with a sorbent that shows affinity for the compounds of interest, is integrated in-line near the inlet of the separation capillary, in order to clean-up and preconcentrate the target analytes from a large volume of sample (typically ~50-100  $\mu\text{L}$ ). After washing to remove non-retained molecules, the retained analytes are eluted in a small volume of an appropriate solution (~25-50 nL), resulting in sample clean-up and concentration enhancement before electrophoretic separation and detection [149,153].



**Figure 1.13.-** On-line SPE-CE setup used in this thesis.

Selection of the appropriate sorbent for optimum performance in SPE-CE is not an easy task. Sorbents not only should show a specific affinity for the target analyte but they also have to be adapted to the reduced dimensions of the microcartridges and to the fact that the extraction is undertaken on-line with a voltage-driven separation. Several sorbents commonly used for off-line SPE or affinity chromatography have been applied to SPE-CE, including reversed phase, ionic exchange, size-exclusion, molecular imprinted polymers, antibodies, aptamers or lectins [149,152–154]. Among them, silica-based sorbents (e.g.  $C_8$  and  $C_{18}$ ) are widely recognized for their high efficiency and good extraction capacities, being  $C_{18}$  the most common chromatographic sorbent used for SPE-CE applications [154]. However, the major drawback of these conventional chromatographic sorbents is their limited selectivity, which has prompted to explore the use of sorbents with higher extraction selectivity such as immobilized metal ion affinity chromatography (IMAC), aptamers, lectin-based sorbents, among others.



In particular, this thesis will present a study by SPE-CE coupled to mass spectrometry (SPE-CE-MS) using TiO<sub>2</sub> beads as a sorbent to selectively analyse glycopeptides from the typical enzymatic digests of glycoproteins prepared in bottom-up proteomics approaches. This sorbent has been selected due to its described affinity for phosphopeptides [113,114] as well as its recent application for the purification of sialylated glycopeptides in off-line protocols [116]. TiO<sub>2</sub> sorbent characteristics and mechanisms of interaction with glycopeptides have already been discussed in section 1.3.1 about purification techniques.

#### **1.3.4. Mass spectrometry**

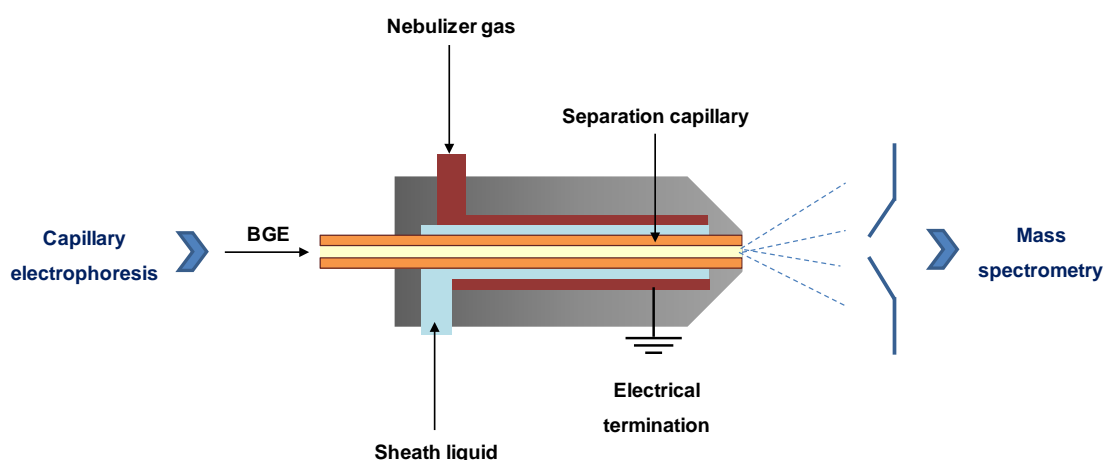
Over the years, mass spectrometry (MS) has become a powerful tool in glycoproteomics due to its outstanding 4S features (sensitivity, stoichiometry, specificity and speed), which facilitates the identification of the glycoproteins, glycosylation sites and structures of glycans [155–158]. There are two general MS-based strategies for glycoprotein analysis. One is the “top-down” MS-based strategy in which the intact glycoproteins are directly subjected to MS and tandem MS analysis to provide the protein sequencing and glycosylation site localization without extensive separation or digestion. Although minimal sample preparation is required, this strategy has been limited up to now to small or low glycosylated glycoproteins, being not used for clinical glycoproteomic analysis due to the complex MS spectrum interpretation. The other is the “bottom-up” MS-based strategy that is the most widely applied for glycoprotein analysis. In this strategy, glycans or glycopeptides, obtained by chemical or enzymatic digestion methods as previously described, are often purified and subsequently characterised by MS. Both “top-down” and “bottom-up” strategies are considered complementary and, in order to obtain all the possible structural information of glycoproteins, the three approaches (i.e. intact glycoprotein, glycopeptides and glycans) should be taken into account.

Regarding instrumentation, a mass spectrometer is a complex and sophisticated analytical instrument that consists of several parts, being the most important ones the ionization source, the analyzer and the detector [159–161]. Briefly, the ionization source allows molecules in the liquid phase to be transferred directly into ions in the gas phase. Then, in the analyzer, ions are separated on the basis of their mass-to-charge ratios ( $m/z$ ). The mechanism behind this separation is highly dependent on the type of analyzer, which could also affect the sensitivity and resolution of the mass spectra. Finally, the detector records either the charge induced or the current produced when an ion passes by or hits a surface.

#### **1.3.4.1 Electrospray ionization**

Electrospray ionization (ESI) and matrix-assisted laser desorption ionization (MALDI) are two ionization technologies that are widely applied to the analysis of glycoproteins and glycoconjugates. These two revolutionary techniques greatly expanded the use of mass spectrometry to almost all fields in science, as they allowed analysis by MS to be extended to non-volatile and thermolabile compounds. Both ionization processes are referred as soft ionization methods because almost no energy is retained by the analytes and, in general, no fragmentation happens during the ionization process [162,163]. Although both ionization methods permitted extensive studies on proteomics field, ESI can be easily coupled to LC or CE and thus is the most widely used ionization technique in glycoproteomics [164–166]. Moreover, ESI generates multiple charged ions that allow both small and large molecules to be transformed into gas-phase ions, only requiring a mass analyzer with a limited range of mass-to-charge ratios ( $m/z$ ). Nonetheless, in the case of CE, its connection to MS (CE-ESI-MS) can be problematic and two key requirements have to be considered: the CE circuit needs to be electrically closed, and very low and BGE-dependent flows need to be handled. To overcome these problems, Smith et. al. [167] developed the so-called sheath-flow interface, which allowed for the first time successful hyphenation of CE to MS. In this interface, a coaxial sheath

liquid is used to increase the flow rate and also close the electric contact in the terminus end of the CE capillary [167] (see **Figure 1.14**). Moreover, an inert gas is usually employed to assist the ionization and improve spray stability. Nowadays, this interface is commercially available, being Agilent Technologies® the leading company in this regard. Although the use of the sheath liquid produces certain dilution of the capillary eluent and detection sensitivity may be reduced, this interface provides increased reproducibility, robustness, ease of use and versatility and has resolved the main problems faced when coupling CE with MS.



**Figure 1.14.**- Schematic representation of the sheath-flow interface for CE-MS coupling.

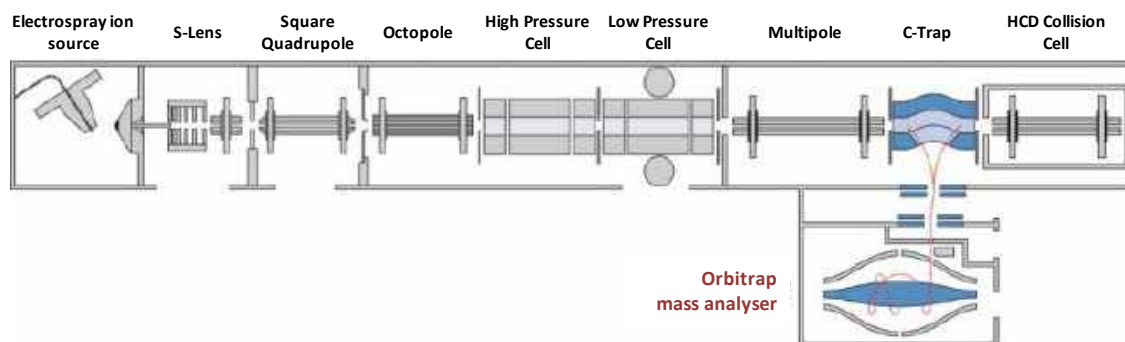
In the last few years, a new variant of ESI has raised considerably interest among the scientific community, the noteworthy nanoESI [168,169]. This interface is ideally suited for the ionization of analytes in nanoflow range systems, such as nanoLC or CE. NanoESI shows high sensitivity and low sample consumption due to the initial size of the formed droplets is considerably reduced. Furthermore, less droplet fissions are needed to ionize the analytes and solvent evaporation is much faster, apart from the reduced adduct formation.

#### 1.3.4.2 Mass analyzers

Nowadays, five main mass analyzers are widely used for the analysis of protein glycosylation, namely, ion trap (IT), quadrupole (Q), time-of-flight (TOF), Orbitrap and Fourier transform ion cyclotron resonance (FT-ICR) [170]. These analyzers vary in terms of size, price, resolution, scanning range and speed, dynamic range and the ability to perform MS/MS experiments. The main analyzer used in this thesis has been an orthogonal acceleration TOF (oa-TOF), due to its seemingly unlimited  $m/z$  range, high-speed acquisition capabilities, high mass accuracy and resolving power, with a relative affordable price [170]. The different direction for the separation of ions, orthogonally introduced with oa-TOF provided several advantages, such as better efficiency in gating ions from an external continuous source (e.g. ESI), reduction of velocity and spatial dispersion and the concomitant increase in mass resolving power, mass accuracy and signal-to-noise ratio (S/N).

In recent years, hybrid analyzer instruments are extensively applied to characterize glycoproteins, not only because such hybrid species could combine the merits of different analyzers but, particularly, for the fact that more structural information can be obtained as MS/MS experiments are easily carried out [170]. Linear ion trap quadrupole (LTQ) and quadrupole-time-of-flight (qTOF) instruments have been already used for glycoprotein analysis [156]. In such hybrid analyzers, collision-induced dissociation (CID) is commonly employed as fragmentation technique to produce tandem mass spectra. The process involves the acceleration of molecules, which are collided with a neutral gas (e.g. nitrogen), resulting in the breaking of molecular bonds and the generation of tandem mass spectra [171,172]. Orbitrap analyzer combined with a linear ion trap (LTQ Orbitrap) provides rapid and accurate tandem MS analysis of complex compounds and has been already employed to identify glycoconjugates with high resolution and mass accuracy [173–175]. In LTQ Orbitrap, a CID-based approach termed higher-energy collisional dissociation (HCD) was developed as a

fragmentation process that overcomes the shortfalls of CID in the linear ion trap, such as low mass resolution and accuracy [176]. In this case, fragmentation spectra are produced outside the ion trap, namely in the C-trap. A schematic representation of LTQ Orbitrap Velos (Thermo Scientific®) mass spectrometer used in this thesis is shown in **Figure 1.15**.



**Figure 1.15.-** Schematic representation of a Linear Trap Quadrupole (LTQ) Velos Orbitrap mass spectrometer.

Furthermore, new advances in instrumentation have led to the introduction of more powerful Orbitrap mass spectrometers, such as Orbitrap Fusion Lumos™ Tribrid (Thermo Scientific®) that combines advanced quadrupole technology, dual-pressure linear ion trap and ultra-high field orbitrap mass analyzer. This instrument has had a great impact in glycoproteomics field as reported by several authors [177–179].

Tandem mass spectrometry is currently an efficient and powerful technique for the structural characterization of glycans [172]. Although exoglycosidase digestion has classically been used for the structural elucidation of glycans, the use of tandem mass spectrometry approaches to obtain structural information about the sequence and linkage-type of glycans has gained interest in last years as commercially available enzymes can only cleave certain linkage-types and their specificity might not be absolutely guaranteed. In this regard, Harvey and co-workers extensively studied the fragmentation of high-mannose, hybrid and complex N-glycans, establishing characteristic fragments of specific features such as sialic acid and fucose linkage-

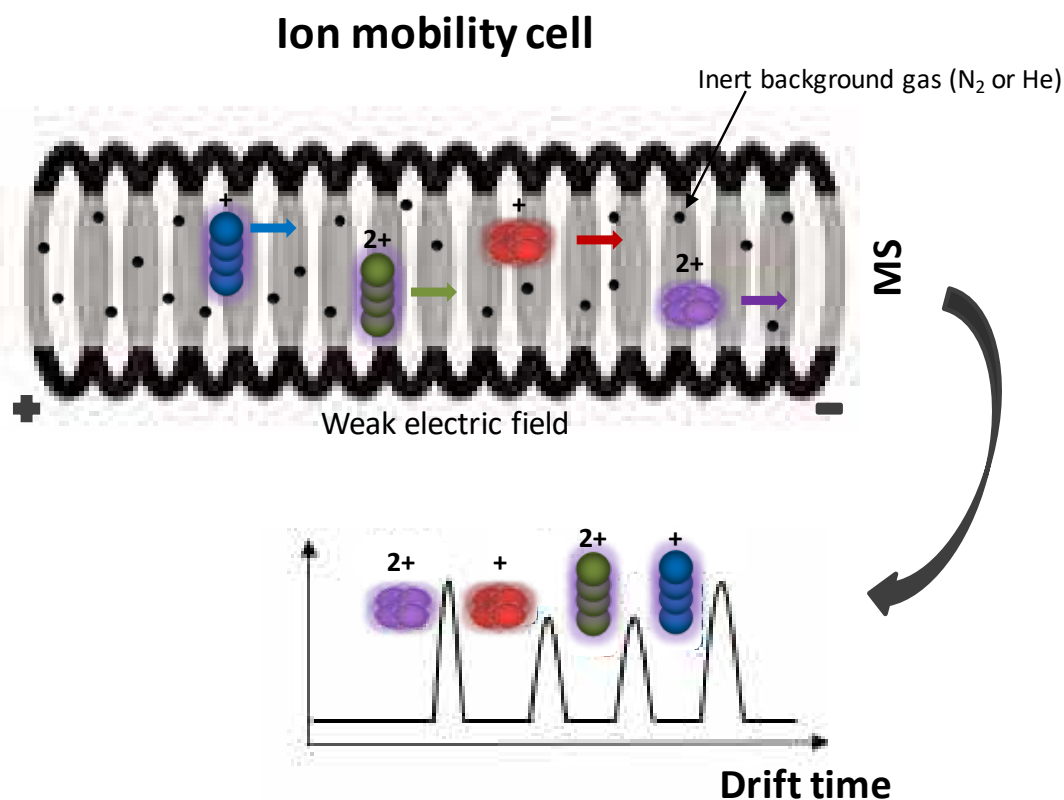
types [180,181]. In this thesis, LTQ Orbitrap has been employed to determine the reliable diagnostic fragments useful for sialic acid and fucose linkage-type assignments using the information of hAGP glycan isomers obtained with exoglycosidase digestions. On the other hand, Orbitrap Fusion Lumos™ Tribrid was also used in this thesis for protein identification of IAC eluted fractions.

### 1.3.5. Ion mobility-mass spectrometry

Ion mobility-mass spectrometry (IM-MS) is a versatile two-dimensional analytical technique for rapid separation and simultaneous detection of the compounds of interest by conducting a gas-phase separation prior to mass analysis [182]. IM-MS provides a new dimension in the separation of compounds, where ions are not only separated due to their mass and charge, but also on the basis of their shape and size. For this reason, its usefulness in various fields has become apparent, especially for glycan analysis, resolving ions that would be otherwise indistinguishable solely by MS, such as isomers [183–186].

In IM-MS, gas ions flow cross a cell filled with a background inert neutral gas (usually N<sub>2</sub> or He) under the joint action of a weak electric field and resistance to reversed gas flow before they are trapped for mass analysis [187]. The time that a particular ion takes to cross the IM cell is called *drift time* and it is mainly determined by the ion-gas collisions. The drift time of an ion is dependent on a multitude of experimental parameters, which makes a direct comparison difficult, and it can be used to calculate the rotationally averaged collision cross section (CCS) of an ion in a specific drift gas. This CCS is a molecular property, which under controlled conditions is independent of instrument parameters and can be correlated to the overall shape and topology of an ion. As such, CCSs can be used as an additional identification parameter, which can be stored in databases, allowing an easier and more reliable structural assignment [188,189]. Separation occurs because ions with high charge and more compact structures

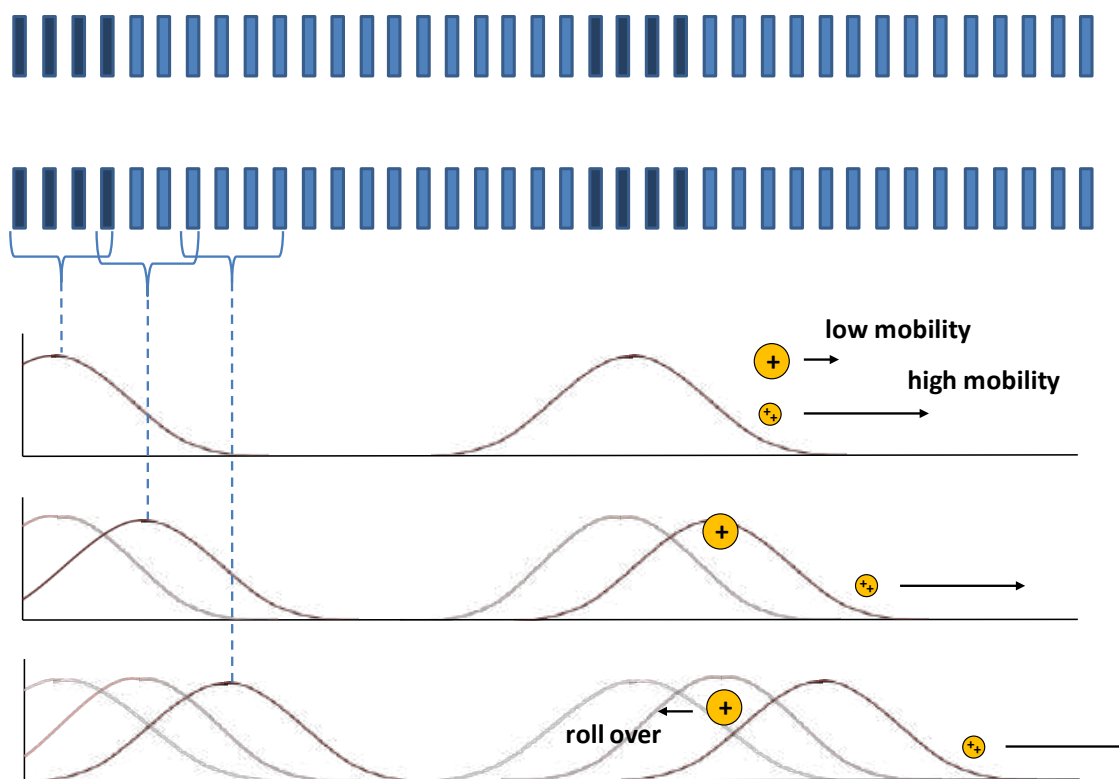
(small CCS) have shorter drift times than those with lower charge and higher CCS. **Figure 1.16** illustrates the fundamentals of classical IM-MS separation.



**Figure 1.16.-** Separation in classical IM-MS. Ions with smaller CCS and higher charge are separated before those with higher CCS and smaller charge.

There are several types of IM instruments, that mainly differ in the nature of the electric field that is used to propel the ions through the cell [187]. Among them, traveling-wave ion mobility spectrometer (TWIMS) can be considered the one that has seen a major growth in the last years and it is the one used in this thesis. In TWIMS, ions are propelled through the cell thanks to a sequence of symmetric potential waves, each ion with its own mobility depending on its charge and CCS. Those ions with higher mobility (i.e. high charge and small CCS) will be pushed along with the wave, thus, they travel the cell faster. Whereas, low mobility ions (i.e. low charge and high CCS) roll over the top of the wave and, consequently, they stay longer in the

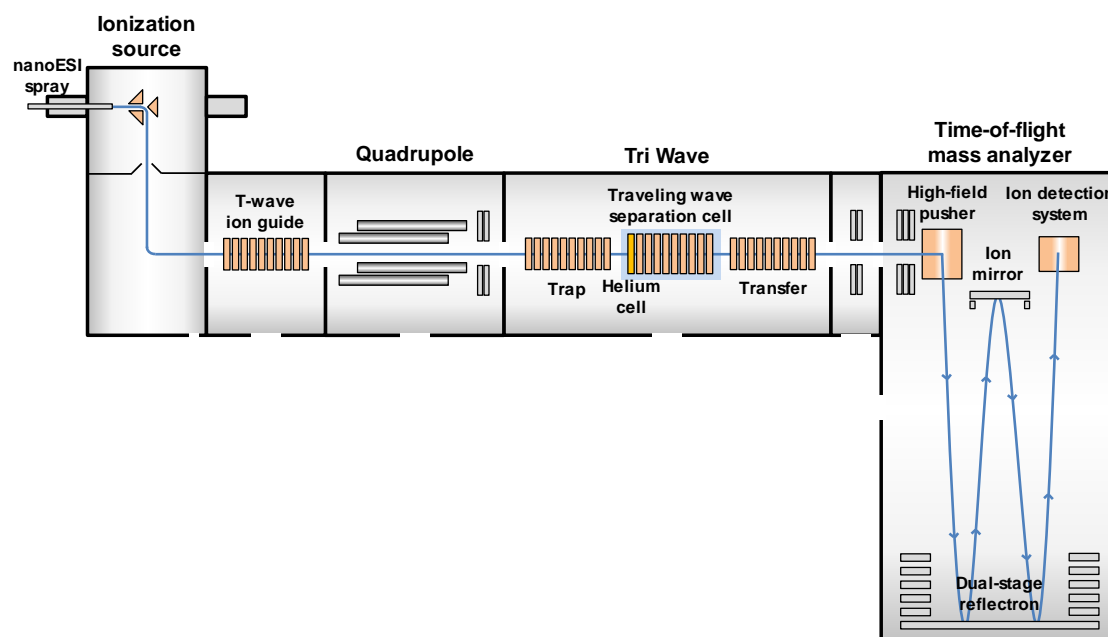
cell. Therefore, different species transit the cell in different times [190,191]. **Figure 1.17** shows an schematic representation of method operation in TWIMS.



**Figure 1.17.-** Fundamentals of TWIMS separation. The top section represents the electrodes that are selectively turned on to emulate a wave-like effect of traveling potentials as represented by the bottom part. Low mobility ions are impeded and roll over the wave whereas high mobility ions are moved along with the wave allowing for temporal separation.

One of the main advantages of TWIMS is that it disperses ion mixtures, allowing the simultaneous measurement of multiple species. This, in conjunction with a high sensitivity obtained when TWIMS is coupled to certain MS analyzers, such as time-of-flight (TOF), has made this platform an alluring option for structural analysis and isomer separation. This platform, along with other IM methods, has been recently explored for the analysis of glycans or glycoconjugates by several authors [192–194]. **Figure 1.18** shows a representation of an ion mobility mass spectrometer with a traveling wave ion mobility cell.





**Figure 1.18.-** Schematic representation of an ion mobility mass spectrometer (IM-MS) with a traveling wave ion mobility cell (Synapt G2 from WATERS® Corp.).

#### 1.4. Data analysis. Chemometric methods

Chemometric methods play a crucial role in data processing, exploration and classification of the massive and complex datasets generated in proteomics and metabolomics studies, especially in untargeted analysis [195,196]. The use of multivariate data analysis methods could be even more valuable when applied to glycomic studies due to the inherent complexity and large microheterogeneity of glycopeptides and glycans with the presence of several isomers. Therefore, efficient data processing and interpretation tools of such huge and complex datasets are required to obtain accurate information and extract reliable conclusions, especially in biomarker research. In this thesis, principal component analysis (PCA) and partial least discriminant analysis (PLS-DA) have been used for glycan-based biomarker discovery of two important diseases. The principles of both chemometric methods will be briefly described in the following section.

### 1.4.1. Principal component analysis

PCA is a widely applied mathematical tool for unsupervised data decomposition and dimensionality reduction, which helps to understand and interpret large and complex datasets [197,198]. PCA is a technique that condenses all the information into a few number of components (Principal Components, PCs) obtained from decomposition of dataset matrix. These PCs maximize explained variance in the data on each successive component under the constraint of orthonormality to each other [197,199]. As a result, a bilinear model is generated which is a product of scores (T) and loadings (P) matrices:

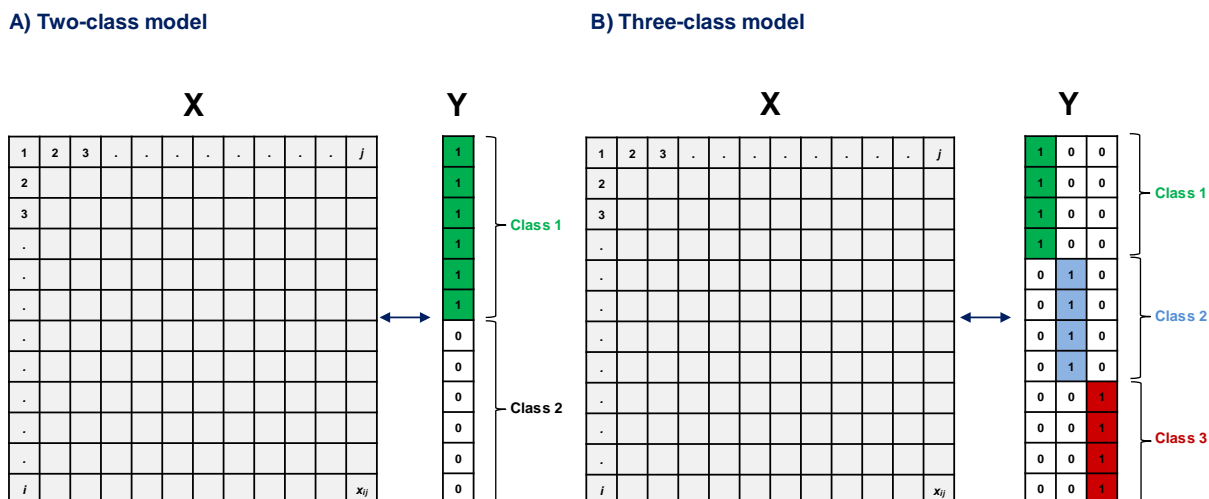
$$X = TP^T + E \quad \text{Eq. 1}$$

where X is the matrix of the dataset (M x N matrix; M, rows: number of samples; N, columns: measured variables), T is the scores matrix (M x A matrix; A: number of calculated PCs) and  $P^T$  is the loadings matrix (A x N matrix). T and P consist of orthogonal and orthonormal vectors, respectively, and E is the residuals (i.e. the variance not explained by the PCs).

PCA is considered a potent visualization technique in which each object (sample) gets a scores value on each PC and, similarly, each variable gets a loadings value on each PC. Thus, objects and variables can be presented in the so-called scores and loadings plots, respectively. Scores plots are quite useful for revealing patterns, such as clusters, trends and outliers, in the data. Additionally, loadings plots are mainly used to check whether there is covariance among variables or to explain and interpret the patterns observed in the scores plot [197,199]. Finally, it is worth mentioning that usually only PCs that explain or map the dominant variation patterns in the data are extracted and noise is left in the residuals matrix.

### 1.4.2. Partial least discriminant analysis

PLS-DA also reduces the dimensionality of a dataset matrix by means of decomposition into a set of components, in this case referred to as latent variables (LVs) [200–203]. In contrast to PCA, which is an unsupervised data decomposition method, PLS-DA is used for the supervised identification of trends and clustering of the data. The method is in fact an extension of partial least squares (PLS). In PLS, a matrix is composed of normalized weight vectors ( $W^T$ ), which are calculated as the covariance between the response matrix  $Y$  (i.e. groups, class membership) and the data matrix  $X$  (i.e. raw data). Scores for the PLS components are calculated by projecting the spectral variables  $X$  on  $W^T$ , whereas loadings are calculated by projecting  $X$  on the resulting scores vectors [201]. When PLS is used as a supervised classification method, the response variable is just a binary vector of zeros and ones (in contrast to PCA, where only the matrix  $X$  is present), which describes the class membership for each sample in the studied groups. In this case, the method is referred to as PLS-DA [204,205]. **Figure 1.19** shows a representation of matrices  $X$  and  $Y$  for PLS-DA models including two and three classes.



**Figure 1.19.-** Representation of partial least squares-discriminant analysis (PLS-DA) for models including two classes (A) and three classes (B).

This method provides several statistics such as the loading weight, the regression coefficient and the variable importance in the projection (VIP), which can be used to identify the most

important variables and their importance to explain differences between classes. As in PCA, this technique provides a visual interpretation of complex datasets through a low-dimensional, easily interpretable scores plot that illustrates the separation between different groups. Comparison of loadings and scores plots supports investigations in terms of the relationship between important variables that can be specific to the group of interest [204,205].



## **Chapter 2. Glycopeptide purification methods**

---



Among the three strategies used for the analysis of protein glycosylation (intact glycoprotein, glycopeptides and glycans), glycopeptide analysis provides valuable information not only about the glycan structures but also about the glycosylation sites and thus about their degree of occupancy. Therefore, the analysis of glycopeptides, obtained after enzymatic digestion of the intact glycoprotein, is a bottom-up MS-based strategy widely applied in glycoproteomics. However, the copresence of peptides in the protein digest can hinder the ionization of the glycopeptides by MS and thus the detection of low abundant glycoforms. In this regard, several purification methods to selectively isolate glycopeptides from protein digests prior to MS, LC-MS or CE-MS analysis have been developed in last years to overcome these limitations. Nonetheless, these methodologies, including affinity chromatography, hydrophilic interaction chromatography or titanium dioxide (TiO<sub>2</sub>) particles, are usually off-line, expensive, laborious and require time-consuming sample treatments before analysis.

CE-MS is especially interesting in glycosylation studies as glycopeptide glycoforms that differ in the number of sialic acids can be separated in a relative easy manner. Therefore, this technique can be considered an alluring alternative to LC-MS for the separation and characterization of glycoprotein digests following a typical bottom-up strategy. Additionally, the on-line coupling of solid-phase extraction with CE (SPE-CE) has been successfully applied to the analysis of a broad variety of small molecules, peptides and proteins using different sorbents. This approach provides an on-line sample clean-up and preconcentration of the target analytes.

This chapter describes the development of two different methods for the purification of glycopeptides from protein digests. Precipitation with acetone is presented as a simple and rapid method for the peptide removal from protein digests. In addition, the influence of the different physicochemical parameters and properties that make peptides, N- and O-glycopeptides to present a different behavior in front of acetone precipitation are studied



using chemometric tools. This chapter also presents an SPE-CE-MS method using TiO<sub>2</sub> sorbent (TiO<sub>2</sub>-SPE-CE-MS) as a novel alternative for on-line sample clean-up and glycopeptide enrichment. The capacity of the above mentioned TiO<sub>2</sub>-SPE-CE-MS methodology to selectively purify glycopeptides with different composition and number of sialic acids is also evaluated.

This chapter includes the following publications:

- **Publication 2.1.-** Analysis of O-glycopeptides by acetone enrichment and capillary electrophoresis-mass spectrometry. Mancera-Arteu, M., Giménez, E., Benavente, F., Barbosa, J., Sanz-Nebot, V. *Journal of Proteome Research* (2017), 16, 4166-4176.
- **Publication 2.2.-** Analysis of glycopeptide biomarkers by on-line TiO<sub>2</sub> solid-phase extraction capillary electrophoresis-mass spectrometry. Mancera-Arteu, M., Lleshi, N., Sanz-Nebot, V., Giménez, E., Benavente, F. *Talanta* (2020), 209, 120563.

## Analysis of O-Glycopeptides by Acetone Enrichment and Capillary Electrophoresis-Mass Spectrometry

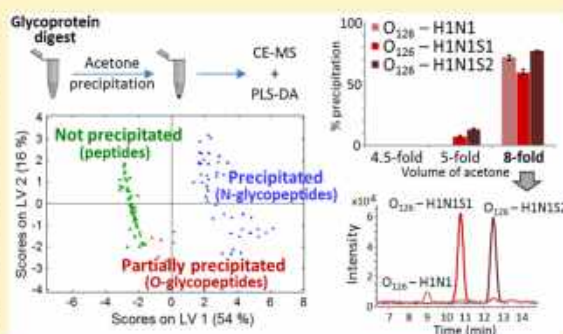
Montserrat Mancera-Arteu,<sup>†</sup> Estela Giménez,<sup>†</sup> Fernando Benavente,<sup>\*,†</sup> José Barbosa,<sup>†</sup> and Victòria Sanz-Nebot<sup>†</sup>

<sup>†</sup>Department of Chemical Engineering and Analytical Chemistry, University of Barcelona, 08028 Barcelona, Spain

### Supporting Information

**ABSTRACT:** Acetone precipitation was evaluated as a rapid, simple, low-cost, and efficient method for the selective purification of O-glycopeptides from enzymatic digests of glycoproteins. Ovalbumin (OVA), human and bovine  $\alpha_1$ -acid glycoprotein (hAGP and bAGP), human apolipoprotein C-III (APO-C3), and recombinant human erythropoietin (rhEPO) were used to obtain enzymatic digests with a broad and varied set of peptides, N-glycopeptides, and O-glycopeptides. After digestion and before capillary electrophoresis mass spectrometry (CE-MS) analysis, the amount of ice-cold acetone added to the digests was optimized to maximize recoveries of O-glycopeptides. Furthermore, the different behavior of peptides, N- and O-glycopeptides was explained by studying with multivariate data analysis methods the influence of several physicochemical parameters and properties related to their composition and structure. Principal component analysis (PCA) and, afterward, partial least-squares discriminant analysis (PLS-DA) were used to identify the most significant variables and their importance to differentiate between peptides, N-glycopeptides and O-glycopeptides, or within these classes. This information was useful to understand precipitation of these compounds after addition of acetone and for the selection of the optimal conditions for purification of specific O-glycopeptide biomarkers. Special attention was paid to O<sub>126</sub>-glycopeptide glycoforms of rhEPO because of their applicability in biopharmaceutical quality control and doping analysis.

**KEYWORDS:** acetone, capillary electrophoresis, enrichment, mass spectrometry, O-glycopeptides



### 1. INTRODUCTION

Glycosylation is one of the most common post-translational modifications in proteins and plays essential roles in biochemical and physiological functions including cell–cell interaction, cell signal regulation, and antigen recognition.<sup>1</sup> Aberrant N- and O-glycosylation has been also related to numerous pathologies such as congenital disorders and cancer,<sup>2–6</sup> or it can cause immune response to certain biopharmaceuticals.<sup>7,8</sup> Therefore, the interest toward the characterization of the carbohydrate moieties of glycoproteins has aroused in biotechnological and biomedical research. In recent years, as O-glycosylation is in general simpler than N-glycosylation,<sup>9,10</sup> it is being regarded as an excellent target for finding novel specific biomarkers. An example is the case of the O<sub>126</sub>-glycopeptide of human erythropoietin (hEPO), a glycoprotein hormone that is involved in the physiological feedback mechanism that maintains red blood cell number and tissue oxygen supply at adequate levels. The presence of nonhuman N-glycolylneuraminic acid (NeuGc) in the carbohydrates of the O<sub>126</sub>-glycosylation site of recombinant human erythropoietin (rhEPO) is being investigated to characterize biopharmaceuticals and discriminate between

endogenous and recombinant erythropoietins in doping control.<sup>11–13</sup>

Tryptic-digested glycoproteins are commonly used in bottom-up proteomics approaches for glycosite characterization. Nevertheless, glycopeptide purification prior to mass spectrometry (MS) analysis continues to be a challenge. The difficulties to detect low abundant glycopeptides and the ion suppression occurred by the copresence of peptides have promoted the development of several enrichment strategies to selectively isolate glycopeptides from protein digests before MS, liquid chromatography–mass spectrometry (LC–MS) or capillary electrophoresis mass spectrometry (CE–MS) analysis.<sup>14</sup> In this regard, because of the advantages and versatility of CE (e.g., high efficiency, short analysis time, low sample, reagent and solvent consumption, and online coupling to MS detection), CE–MS has become an excellent alternative to LC–MS for the analysis of proteins and peptides.<sup>15–17</sup>

One of the most used methods for glycopeptide purification is lectin affinity chromatography.<sup>18–21</sup> However, in general, because of the broad selectivity of lectins toward sugar chains,

Received: July 24, 2017

Published: September 25, 2017

multidimensional approaches or a combination of several lectins are necessary to enrich various types of glycopeptides with different glycan compositions and structures. Another option is boronate affinity chromatography, but it could be affected by nonspecific binding of peptides as it is not specific for glycans.<sup>22</sup> Hydrophilic interaction chromatography and, more recently, extraction with titanium dioxide (TiO<sub>2</sub>) particles were also reported as good alternatives.<sup>23–27</sup> However, in the case of TiO<sub>2</sub> particles, only glycopeptides with sialic acid can be retained; hence, nonsialylated glycopeptides are lost during the purification step. Furthermore, in general, all these methodologies are expensive, labor-intensive, and time-consuming. Thereby, the development of rapid, simple, low-cost, and efficient purification methods for the selective isolation of glycopeptides is still necessary.

Recently, Takakura et al. proposed a method based on precipitation with five-fold volume excess of acetone for the selective enrichment of glycopeptides from glycoprotein digests.<sup>28</sup> That study was focused on the separation of glycopeptides from peptides, but no attention was paid to the potential of the different behavior of N- and O-glycopeptides. Nowadays, the main structural differences between peptides, N-glycopeptides and O-glycopeptides are well-known, but very often this empirical knowledge is not enough to reliably explain differences on experimental observations about solubility in different solvents, chromatographic retention, electrophoretic migration, or electrospray ionization before general principles or theories are derived. In this sense, we recently showed that the classical semiempirical relationships between the electrophoretic mobility and the charge-to-mass ratio ( $m_e$  vs  $q/M^a$ ) are useful to model in CE-MS the migration behavior of peptides and glycopeptides originated from the digestion of rhEPO.<sup>29</sup> Other studies have been also performed to simulate in CE-MS the migration of peptide hormones and peptides from protein digests,<sup>30–32</sup> but unfortunately, studies for complex post-translationally modified peptides are scarce.<sup>29</sup>

In this paper, precipitation with acetone of peptides, N- and O-glycopeptides from enzymatic digests of several glycoproteins of biomedical interest was explained by investigating with multivariate data analysis methods the influence of several physicochemical parameters and properties related to their composition and structure. The information about the most relevant features to differentiate between peptides, N- and O-glycopeptides, or within these compound classes was useful to understand their precipitation after addition of acetone as well as for the selection of the optimal conditions for purification of specific O<sub>126</sub>-glycopeptide glycoforms of rhEPO.

## 2. EXPERIMENTAL SECTION

### 2.1. Chemicals

All chemicals used in the preparation of buffers and solutions were of analytical reagent grade. Acetic acid (HAc, glacial), formic acid (FA 98–100%), sodium hydroxide, and acetone were supplied by Merck (Darmstadt, Germany). D,L-Dithiothreitol (DTT, ≥ 99%), iodoacetamide (IAA, ≥ 98%), and ammonium hydrogen carbonate (≥99.9%) were purchased from Sigma-Aldrich (St. Louis, MO, USA). Isopropyl alcohol and water LC-MS quality grade were provided by Scharlab (Barcelona, Spain) and Sigma-Aldrich, respectively. Trypsin (sequencing grade modified, 16 000 U mg<sup>-1</sup>) and chymotrypsin were purchased from Promega (Madison, WI, USA). ESI low concentration (ESI-L) tuning mix was supplied by Agilent

Technologies (Waldbronn, Germany) for tuning and calibration of the mass spectrometer.

### 2.2. Glycoprotein Samples

Ovalbumin from chicken egg (OVA, 98%), human and bovine  $\alpha$ 1-acid-glycoproteins (hAGP and bAGP, 99%), and apolipoprotein C-III from human plasma (APO-C3) were obtained from Sigma-Aldrich. Stock solutions of 1000 mg L<sup>-1</sup> were prepared in water. rhEPO produced in a chinese hamster ovary (CHO) cell line was provided by the European Pharmacopeia as a Biological Reference Product (BRP-lot4). Each sample vial contained 100  $\mu$ g of rhEPO (a mixture of alpha and beta rhEPO), 24 mg of D-(+)-trehalose, 2.4 mg of arginine hydrochloride, 0.08 mg of Tween-20, 3.6 mg of NaCl, and 2.5 mg of NaH<sub>2</sub>PO<sub>4</sub>·2H<sub>2</sub>O. The content of each vial was dissolved in water to obtain a 1000 mg L<sup>-1</sup> protein solution. Excipients of low-molecular mass were removed by ultracentrifugation using Microcon YM-10 centrifugal filters from Millipore (M<sub>r</sub> cutoff 10 000, Bedford, MA, USA) as described elsewhere.<sup>13</sup> Centrifugations were carried out in a Mikro 20 centrifuge (Hettich, Tuttlingen, Germany) at room temperature. Finally, aliquots of 1000 mg L<sup>-1</sup> stock protein solutions were evaporated to dryness using a Savant SPD-111 V SpeedVac concentrator (Thermo-Fisher Scientific, Waltham, MA, USA) and stored at -20 °C until enzymatic digestion.

OVA, hAGP, and bAGP were first reduced and alkylated to break the disulfide bridges and facilitate digestion. Briefly, an aliquot of 25  $\mu$ g of dried glycoprotein was dissolved in 25  $\mu$ L of 50 mM NH<sub>4</sub>HCO<sub>3</sub> (pH 7.9), and 2.5  $\mu$ L of 0.5 M DTT in the same buffer were added. The mixture was incubated in a thermoshaker at 56 °C for 30 min and then alkylated by adding 7  $\mu$ L of 50 mM IAA in 50 mM NH<sub>4</sub>HCO<sub>3</sub> (pH 7.9) and shaking for 30 min at room temperature in the dark. Excess of low-molecular mass reagents was removed with Microcon YM-10 centrifugal filters as described elsewhere.<sup>13</sup> The final protein residue was dissolved in 50 mM NH<sub>4</sub>HCO<sub>3</sub> (pH 7.9) to obtain a final concentration of 1000 mg L<sup>-1</sup>. Trypsin digestion: a 25  $\mu$ L aliquot of 1000 mg L<sup>-1</sup> protein solution in 50 mM NH<sub>4</sub>HCO<sub>3</sub> (pH 7.9) was digested. APO-C3 and rhEPO tryptic digests were prepared from aliquots of 25  $\mu$ g of dried glycoprotein, while hAGP and bAGP digests were prepared from aliquots of 25  $\mu$ g of reduced and alkylated dried glycoprotein. Trypsin was added in an enzyme to protein ratio of 1:25 m/m (1  $\mu$ L of 1  $\mu$ g/ $\mu$ L solution). The mixture was vortexed and subsequently incubated at 37 °C for 18 h. Digestion was stopped by heating at 100 °C for 10 min, and glycoprotein sample was stored at -20 °C until analysis.<sup>13</sup> Chymotrypsin digestion: a 25  $\mu$ L aliquot of 1000 mg L<sup>-1</sup> OVA solution in 50 mM NH<sub>4</sub>HCO<sub>3</sub> (pH 7.9) was digested. OVA chymotryptic digest was prepared from an aliquot of 25  $\mu$ g of reduced and alkylated dried glycoprotein. Chymotrypsin was added in an enzyme to protein ratio of 1:25 m/m (10  $\mu$ L of 0.1  $\mu$ g/ $\mu$ L solution), and the mixture was vortexed and incubated at 37 °C for 18 h. Digestion was stopped by heating in a thermoshaker at 100 °C for 10 min, and sample was stored at -20 °C until analysis.<sup>13</sup> All digestions were performed in triplicate. Incubations were performed in a TS-100 thermoshaker (Biosan, Riga, Latvian Republic).

### 2.3. Acetone Enrichment

An amount of ice-cold acetone, from 4.5-fold to nine-fold volume excess, was added to an aliquot of 25  $\mu$ L of the enzymatic digest and the mixture was incubated in the freezer at -25 °C overnight.<sup>28</sup> Afterward, the supernatant and the

precipitate were separated by centrifugation at 16 000g for 15 min at 20 °C. Centrifugations were carried out in a cooled Rotanta 460 centrifuge (Hettich, Tuttlingen, Germany). The supernatant and the precipitate fractions were collected in new tubes and evaporated to dryness by SpeedVac. Both sample fractions were reconstituted in 25  $\mu$ L of background electrolyte (BGE: 50 mM HAc and 50 mM HFor, pH 2.2) before CE-MS analysis.

#### 2.4. CE-MS

CE-MS experiments were performed in a HP<sup>30</sup>CE system coupled to a 6220 oa-TOF LC/MS mass spectrometer with an orthogonal sheathflow interface (Agilent Technologies). The sheath liquid (50:50 (v/v) iPrOH/H<sub>2</sub>O with 0.05% (v/v) of HFor) was delivered at a flow rate of 3.3  $\mu$ L min<sup>-1</sup> by a KD Scientific 100 series infusion pump (Holliston, MS, USA) and degassed for 10 min by sonication before use. CE control and separation data acquisition (e.g., voltage, temperature and current) were performed using Chemstation software (Agilent Technologies) that was running in combination with the MassHunter workstation software (Agilent Technologies) for control, data acquisition, and processing of the mass spectrometer. The mass spectrometer was tuned and calibrated following the manufacturer's instructions. A "check tune" of the instrument was performed every day in positive mode to ensure accurate mass assignments. Instrumental parameters were optimized for the analysis of rhEPO O<sub>126</sub> and N<sub>33</sub> glycopeptides in a previous study.<sup>13</sup> The optimized operational conditions in positive electrospray ionization (ESI) mode were: capillary voltage 4000 V, drying gas (N<sub>2</sub>) temperature 200 °C, drying gas flow rate 4 L min<sup>-1</sup>, nebulizer gas (N<sub>2</sub>) 10 psig, fragmentor voltage 190 V, skimmer voltage 60 V, and OCT 1 RF Vpp voltage 300 V. Data were collected in profile (continuum) at 1 spectrum s<sup>-1</sup> (approximately 10 000 transients spectrum<sup>-1</sup>) between  $m/z$  100 and 3200 working in the highest resolution mode (4 GHz).

A bare fused-silica capillary of 70 cm total length ( $L_T$ )  $\times$  50  $\mu$ m internal diameter (I.D.)  $\times$  360  $\mu$ m outer diameter (O.D.) (Polymicro Technologies, Phoenix, AZ, USA) was used for CE-MS separations. Activation and conditioning procedures were carried out off-line to avoid contamination with NaOH of the mass spectrometer. New capillaries were activated by flushing (930 mbar) sequentially for 30 min each with 1 M NaOH, water, and BGE (50 mM HAc and 50 mM HFor, pH 2.2). Capillaries were conditioned every day by flushing with NaOH (5 min), water (7 min), and BGE (10 min). Samples were injected for 15 s at 50 mbar, and electrophoretic separations were performed at 25 °C and 25 kV under normal polarity (cathode in the outlet). Between runs, capillaries were flushed with water (1 min), 1 M HAc (3 min), water (1 min), and BGE (5 min). Capillaries were stored overnight filled with water. Before CE-MS, all solutions were passed through a 0.45  $\mu$ m nylon filter (MSI, Westboro, MS, USA).

pH measurements were carried out with a Crison 2002 potentiometer and a Crison electrode 52-03 (Crison Instruments, Barcelona, Spain).

#### 2.5. Data Analysis

CE-MS data, collected for the glycoprotein digests, as well as for the supernatant and the precipitate fractions obtained after the addition of acetone, were processed to obtain the extracted ion electropherograms (EIE) of all the peptides and glycopeptides. Peak areas were measured from the EIEs and used to calculate a percentage of precipitation (i.e., area in the

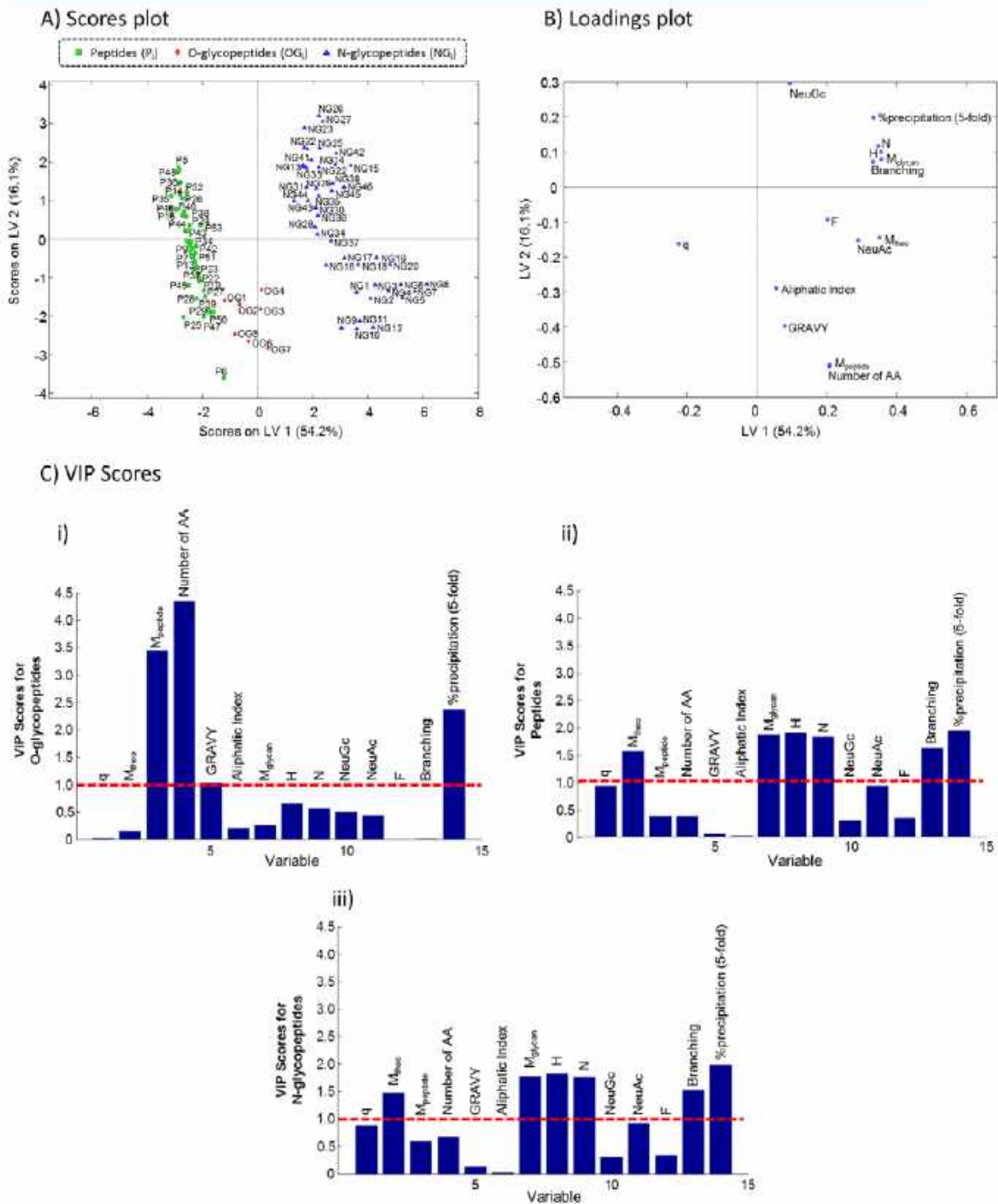
precipitate fraction divided by the total area (precipitate plus supernatant)). The percentage of precipitation adding five-fold and eight-fold volumes of acetone and several other parameters related to the structure and composition of peptides and glycopeptides were used to build a matrix for multivariate data analysis. Table S-1 summarizes all the variables considered for the peptides and glycopeptides of the analyzed glycoprotein digests. The theoretical monoisotopic  $M_r$  of the complete molecule ( $M_{theo}$ ), the peptide ( $M_{peptide}$ ), and the glycan ( $M_{glycan}$ ) backbones were taken into account. Net molecular charge ( $q$ ) was calculated as described in our previous work using the Sillero and Ribeiro expression,<sup>29,33</sup> which is based on the Henderson–Hasselbalch equation. Parameters related to the structure of the glycan moiety, such as the number and type of monosaccharide units (i.e., hexose (H), N-acetylhexose (N), fucose (F), and sialic acids (SiA (S): N-acetylneuraminic acid, NeuAc, and N-glycolylneuraminic acid, NeuGc)) and branching, as well as of the peptide part, such as number of amino acids (AAs), were also included. The hydrophobicity of the peptide chain was estimated calculating the grand average of hydrophobicity (GRAVY) and the aliphatic index. The GRAVY values for the peptide chains were calculated as the sum of hydrophobicity values of all the AAs, divided by the number of residues in the chain.<sup>34</sup> The aliphatic index, defined as the relative volume occupied by the aliphatic side chains (alanine, valine, isoleucine, and leucine), was calculated as described by Ikaai.<sup>35</sup>

All variables obtained for each peptide/glycopeptide were autoscaled (mean centered and scaled to unit standard deviation). Principal component analysis (PCA) was performed to explore the data for different classes and outliers.<sup>36,37</sup> Partial least-squares discriminant analysis (PLS-DA) was applied afterward to maximize class separation and identify which parameters were the most significant to discriminate between classes taking into account the variable importance in the projection (VIP) scores.<sup>38,39</sup> Leave-one-out ( $n < 20$ ) or venetian blinds ( $n \geq 20$ ) cross-validations of the PLS-DA model were performed during calibration, depending on the number of peptides/glycopeptides of the data set ( $n$ ).<sup>40,41</sup> SOLO (Version 8.2, student edition, eigenvector Research Inc., Wenatchee, WA, USA) was used for PCA, PLS-DA, and VIP calculations.

## 3. RESULTS AND DISCUSSION

### 3.1. Evaluation of Peptide and Glycopeptide Precipitation with Acetone

First, precipitation of peptides, N- and O-glycopeptides with acetone were investigated by CE-MS using a five-fold volume of acetone and five enzymatic digests obtained from different glycoproteins of biomedical interest (rhEPO, APO-C3, hAGP, bAGP, and OVA). In a previous study, Takakura et al. recommended this amount of acetone for the quantitative separation of peptides from glycopeptides.<sup>28</sup> Our set of peptides, N- and O-glycopeptides was broad and varied, including a total of 53 peptides, 46 N-glycopeptides, and 7 O-glycopeptides with different structures and compositions (see Table S-1). Only in the case of OVA, digestion was performed with chymotrypsin, as the N-glycopeptide obtained with trypsin presented a longer peptide chain and it was not possible to detect it. It is also worth mentioning, that the set of O-glycopeptide glycoforms was smaller than the other two sets of compounds. In humans, with the exception of mucins, O-



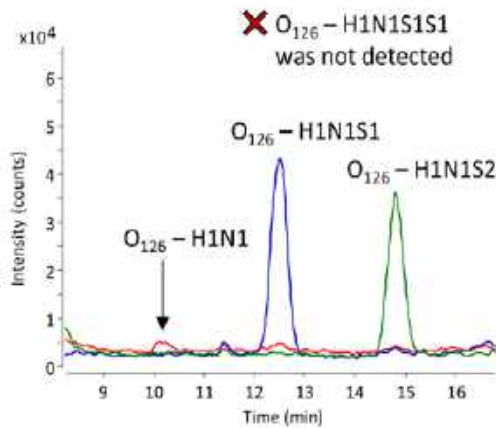
**Figure 1.** (A) Scores plot and (B) loadings plot of the PLS-DA model applied to the percentage of precipitation adding five-fold volume of acetone and the different physicochemical parameters and properties of peptides and glycopeptides of the enzymatic digests of ovalbumin (OVA), human and bovine  $\alpha_1$ -acid glycoprotein (hAGP and bAGP), human apolipoprotein C-III (APO-C3) and rhEPO (see Table S-1). (C) VIP scores of the different variables when considering the separation of (i) O-glycopeptides from peptides and N-glycopeptides, (ii) peptides from N- and O-glycopeptides, and (iii) N-glycopeptides from peptides and O-glycopeptides.

glycosylation is simpler (smaller and less branched carbohydrates) than N-glycosylation; hence, O-glycopeptides show less amount of glycoforms than N-glycopeptides.<sup>9,10</sup> The percent-

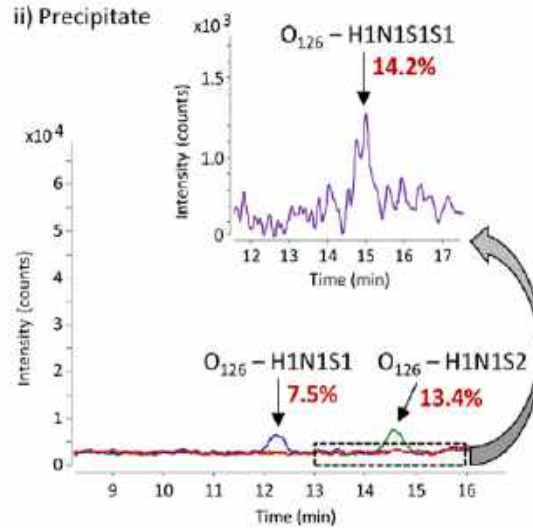
age of precipitation of the different compounds adding a five-fold volume of acetone and the physicochemical parameters and properties related to their structure and composition are

## A) 5-fold volume of acetone

## i) Supernatant

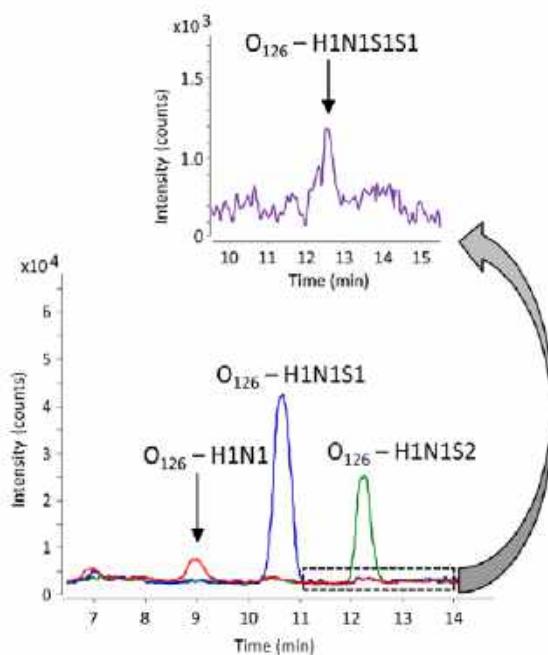


## ii) Precipitate

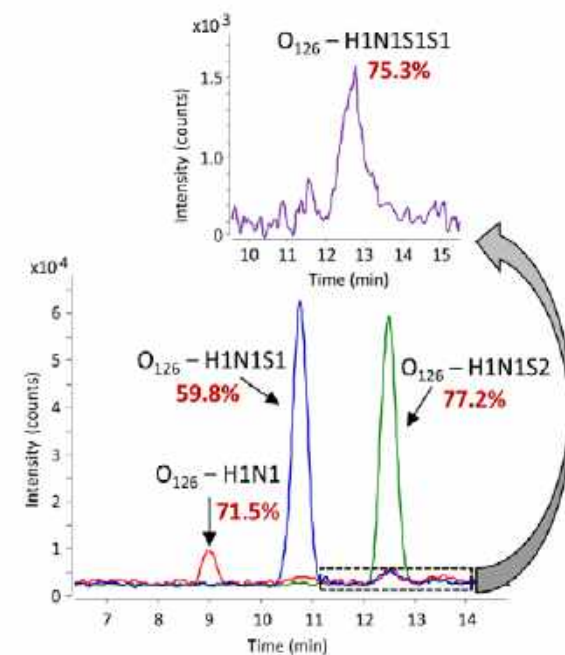


## B) 8-fold volume of acetone

## i) Supernatant



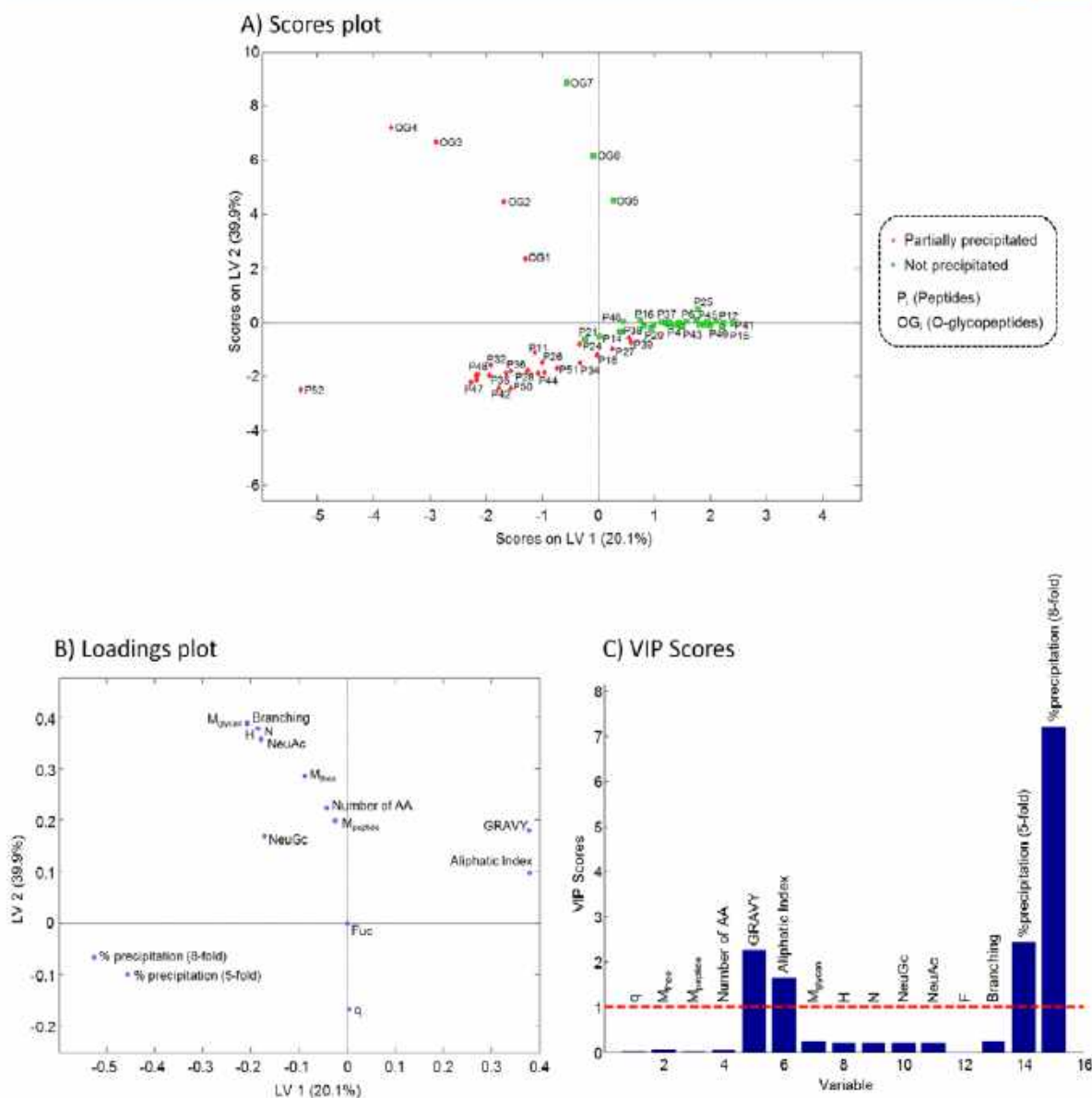
## ii) Precipitate



**Figure 2.** Extracted ion electropherograms (EIEs) of the most abundant  $O_{126}$ -glycopeptide glycoforms in rhEPO tryptic digest using (A) five-fold and (B) eight-fold volume of acetone. (i) Supernatant and (ii) precipitate fractions. A zoom of the EIE of H1N1S1S1  $O_{126}$ -glycopeptide, which contained one NeuAc and one NeuGc, is shown because of its very low abundance. The percentage of precipitation of each glycoform is shown in red color.

shown in Table S-1. Specifically, variables associated with the whole molecule ( $M_{\text{hbo}}$  and  $\eta$ ), the peptide chain ( $M_{\text{peptide}}$ , number of AAs, GRAVY, and aliphatic index), and the glycan moieties ( $M_{\text{glycan}}$ , number and type of monosaccharide units (H, N, F, S, or SiA) and branching) were taken into account. With regard to the precipitation with a five-fold volume of acetone, it was observed that N-glycopeptides were mostly

enriched in the precipitate, peptides remained in the supernatant, and O-glycopeptides were partially precipitated. These results agreed with the observations of Takakura et al.,<sup>28</sup> who proposed adding a five-fold volume of acetone for the selective enrichment of glycopeptides from glycoprotein digests. However, they were studying mostly N-glycopeptides, without



**Figure 3.** (A) Scores plot, (B) loadings plot, and (C) VIP scores of the PLS-DA model applied to the percentage of precipitation adding a five-fold and eight-fold volume of acetone and the different physicochemical parameters and properties of peptides and O-glycopeptides. Partial precipitation was considered when the percentage of precipitation of the peptides or glycoforms was higher than 10% (see Table S-1).

exploring the potential separation of O-glycopeptides from N-glycopeptides.

Multivariate data analysis was used to study the influence of all these variables on five-fold volume acetone precipitation in an easy, rapid, and systematic manner. First, PCA was used to explore the data for the unsupervised identification of trends, classes, and outliers.<sup>36,37</sup> In Figure S-1A of the Supporting Information, the scores plot for the first two principal components, which explained a total of 70.1% of the variance (53.8% by PC1 and 16.3% by PC2), is shown. The first principal component was the most useful to differentiate between classes of compounds, while PC2 separated to some extent compounds that belong to the same class. As can be observed, three classes corresponding to peptides (P), N-

glycopeptides (NG<sub>i</sub>), and O-glycopeptides (OG<sub>i</sub>) were clearly separated. The loadings plot (Figure S-1B of the Supporting Information) showed interesting information about the parameters that were related to each class. It is worth mentioning that two N-glycopeptides from the OVA digest (NG39 and NG40), which were expected to precipitate, behaved as peptides, and thus, they remained in the supernatant (see Table S-1). To explain their abnormal behavior, an independent PCA was carried out for N-glycopeptides (see scores and loadings plots in Figure S-2 of the Supporting Information). All the OVA N-glycopeptides (NG39-NG46) were separated from the rest because they did not present SiA, as well as  $q$ ,  $M_{theor}$  and  $M_{peptide}$  were in general lower than the rest of N-glycopeptides, while hydrophobicity (GRAVY and

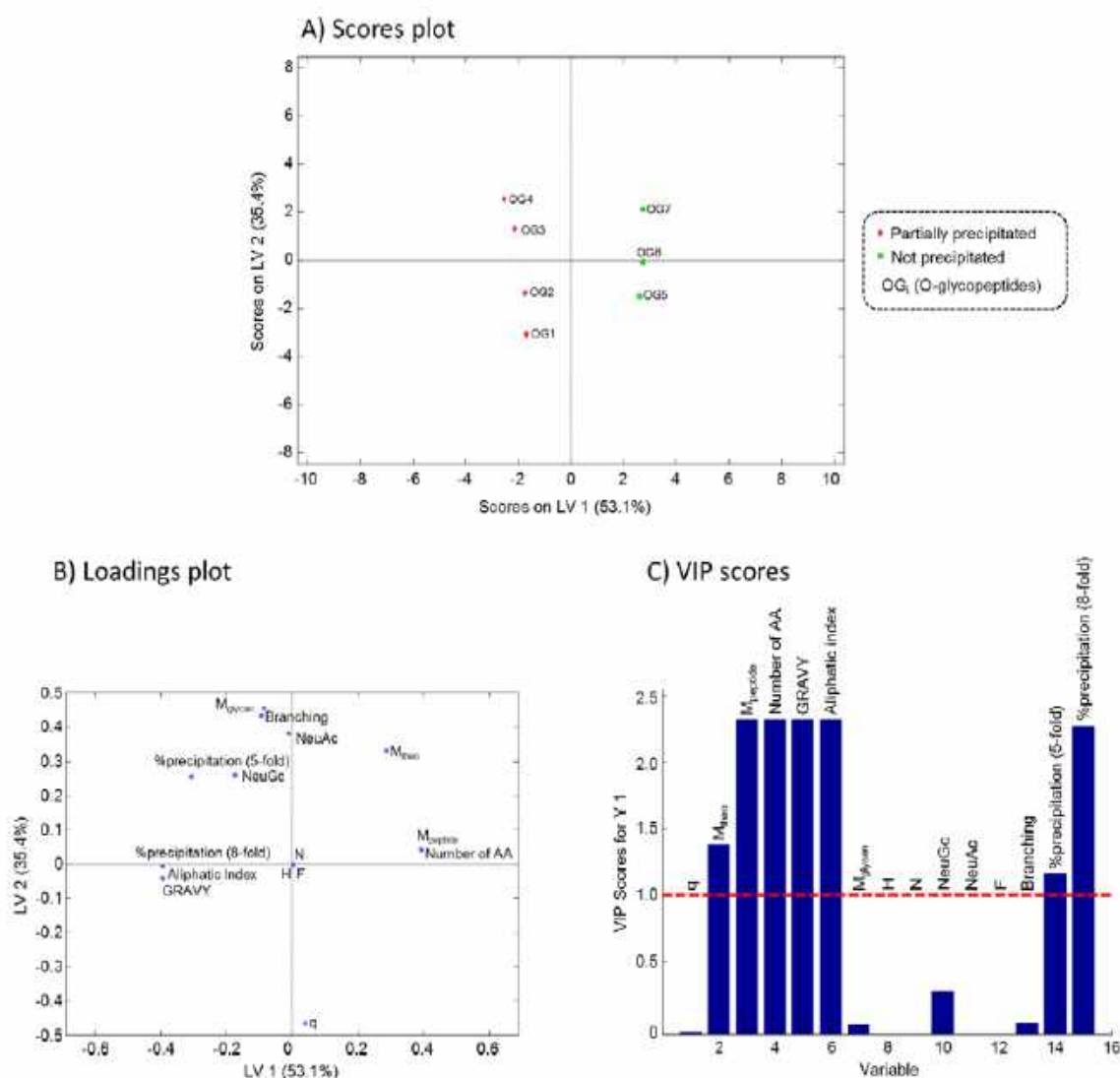


Figure 4. (A) Scores plot, (B) loadings plot, and (C) VIP scores of the PLS-DA model applied to the percentage of precipitation adding a five-fold and eight-fold volume of acetone and the different physicochemical parameters and properties of O-glycopeptides (see Table S-1).

aliphatic index), sugar content and branching were higher (see scores and loadings plots in Figure S-2 of the Supporting Information). These differences were extreme in NG39 and NG40 N-glycopeptides, which presented the lowest  $M_{iso}$  (see Table S-1) and would explain why they did not precipitate with a five-fold volume of acetone as expected by the conclusions of Takakura et al.<sup>25</sup> Therefore, NG39 and NG40 were considered as outliers and were not taken into account for PLS-DA.

As PCA does not reveal the importance of the variables to differentiate between the observed classes in the scores plot, we explored PLS-DA. The PLS-DA model was built considering the three classes observed by PCA and separation was improved, which also facilitated the identification of the most relevant variables that characterize the different precipitation of peptides and glycopeptides after adding five-fold volume of acetone. Figure 1A shows in the scores plot that two latent variables (LV) were enough to differentiate between the three classes of compounds and a total of 70.3% of variance was

explained. The loadings plot (Figure 1B) was very similar to the one obtained before with PCA (Figure S-1B of the Supporting Information) and showed the contribution of each variable to the different LV. As can be seen in Figure 1B, together with the percentage of precipitation, variables related to glycan structure and composition such as  $M_{iso}$ ,  $M_{glycan}$ , number of H, N, and NeuAc, branching, and q were the most important for the separation along the LV1 x-axis. Additionally, the VIP scores allowed quantifying the significance of the different variables on the separation. Figure 1C shows the VIP scores of the different variables when considering (i) separation of the O-glycopeptides from peptides and N-glycopeptides, (ii) peptides from O- and N-glycopeptides, and (iii) N-glycopeptides from peptides and O-glycopeptides. As can be observed, similar bar plots were obtained to distinguish peptides and N-glycopeptides from others (Figure 1C-ii and C-iii), indicating that  $M_{iso}$ ,  $M_{glycan}$ , number of H and N and branching were the most relevant parameters in the precipitation of these compounds with a five-



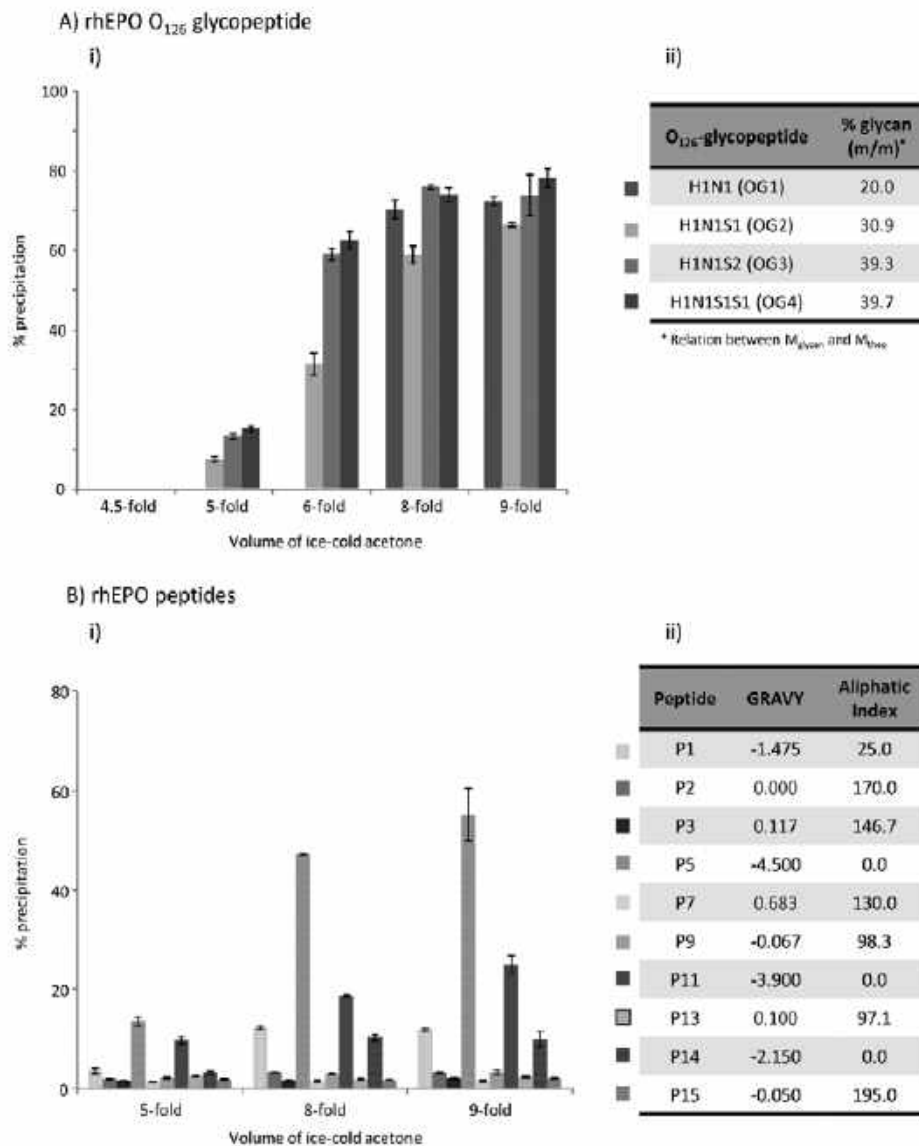


Figure 5. Bar graph with the percentage of precipitation ( $n = 3$ ) adding different amounts of acetone for (A) rhEPO O<sub>126</sub> glycopeptide glycoforms and (B) some representative rhEPO peptides. (ii) Tables given as insets summarize the most important parameters influencing precipitation.

fold volume of acetone (VIP values higher than 1). Concerning the differentiation of the O-glycopeptides from the rest of compounds (Figure 1C-i),  $M_{peptide}$ , the number of AAs of the peptide chain and, in less extent, GRAVY were the most significant. By taking into account these observations, it could be concluded that practically all the N-glycopeptides precipitated quantitatively with a five-fold volume of acetone because they showed a large and structurally rich glycan moiety ( $\geq 45\%$  (m/m) in all cases) and hence a larger  $M_{theor}$ . Moreover, as can be noted, N-glycopeptides precipitated regardless of the charge or of the presence of SiA and their type (NeuAc or NeuGc). Concerning O-glycopeptides, the number of AAs,  $M_{peptide}$  and hydrophobicity as well as the lower glycan content ( $\leq 40\%$  (m/m) in all cases) with regard to N-glycopeptides, which made them structurally closer to peptides, could explain why five-fold volume of acetone was not enough for a

quantitative precipitation, but allowed a little differentiation from peptides.

### 3.2. Study of O-Glycopeptide Enrichment by Precipitation with Acetone

With the aim of understanding the influence of the acetone amount on the precipitation of O-glycopeptides, the results with a five-fold and eight-fold volume of ice-cold acetone were compared for the tryptic digests of rhEPO and APO-C3. The percentage of precipitation obtained for each peptide and O-glycopeptide is shown in Table S-1. With a five-fold volume of acetone, several peptides precipitated to some extent and different behavior between O<sub>94</sub>-APO-C3 and O<sub>126</sub>-rhEPO glycopeptides was observed. In the case of APO-C3, all the glycopeptide glycoforms of the O<sub>94</sub> glycosylation site were detected in the supernatant (OG5-OG7 glycopeptides, see Table S-1), whereas the sialylated glycoforms of O<sub>126</sub>-rhEPO

started to precipitate (OG2-OG4 glycopeptides, see Table S-1). In contrast, with a eight-fold volume of acetone, rhEPO O<sub>126</sub>-glycopeptide was further enriched in the precipitate being able to detect all the glycopeptide glycoforms (OG1-OG4 glycopeptides, see Table S-1), while APO-C3 O-glycopeptide had not yet started to precipitate (only ~2% of precipitation for H1N1S1 and H1N1S2 glycoforms, see OG6 and OG7 glycopeptides in Table S-1). As an example, Figure 2 shows the extracted ion electropherograms (EIEs) of rhEPO O<sub>126</sub>-glycopeptide glycoforms in the supernatant and in the precipitate with a (A) five-fold and (B) eight-fold volume of acetone. The EIEs of the low abundant H1N1S1S1 glycoform with one NeuGc and one NeuAc (OG4) are also shown. As can be observed in Table S-1, this glycoform presented similar percentages of precipitation to the H1N1S2 with two NeuAc (OG3) using these acetone proportions. With regard to the precipitation of peptides, similar observations to O-glycopeptides could be made and the extent of precipitation varied with the amount of acetone added (see Table S-1). Hence, the amount of acetone could be finely tuned to achieve the selective enrichment of O- and N-glycopeptides from peptides.

A PCA model and, afterward, a PLS-DA model, including the percentages of precipitation with a five-fold and eight-fold volume of acetone and the structural features, were built for peptides and O-glycopeptides to find the variables that explain their precipitation with acetone. hAGP and bAGP peptides were also included in the model to increase the data set of peptides. Figure 3 shows the results obtained for the PLS-DA model. As can be observed in the scores plot (Figure 3A), LV1 enabled the differentiation between peptides/O-glycopeptides that were partially precipitated and those that did not precipitate with both amounts of acetone, while LV2 separated peptides from O-glycopeptides. The loadings plot showed that the variables related to the glycan moiety were responsible of the differentiation between O-glycopeptides and peptides (see Figure 3B). Moreover, the loadings and VIP scores plots (Figures 3B,C, respectively) revealed that the hydrophobicity parameters (i.e., GRAVY and aliphatic index) were the most important variables to explain the different trend of peptides and O-glycopeptides to precipitate. Therefore, peptides and O-glycopeptides with a high negative GRAVY value and low aliphatic index were partially enriched in the precipitate depending on the amount of acetone added because they were more hydrophilic (see Table S-1).

To study in more detail the variables that made rhEPO O<sub>126</sub>-glycopeptide glycoforms (OG1-OG4) precipitate more easily than APO-C3 O<sub>94</sub>-glycopeptide glycoforms (OG5-OG7), a PLS-DA model was built for O-glycopeptides separately. Even though the data set was reduced, the scores plot (Figure 4A) showed two groups with rhEPO O<sub>126</sub>-glycopeptide and APO-C3 O<sub>94</sub>-glycopeptide clearly separated using two LV (88.5% of the variance explained). The loadings and VIP scores plots (Figure 4B,C, respectively) indicated that apart from hydrophobicity other variables related to the peptide chain size (especially  $M_{\text{peptide}}$  and number of AAs, which are related to  $M_{\text{theo}}$ ) were essential to explain precipitation of O-glycopeptides. Therefore, rhEPO O<sub>126</sub>-glycopeptide was easier to precipitate, although the hydrophobicity of the peptide chain was similar to APO-C3 O<sub>94</sub>-glycopeptide (GRAVY and aliphatic indexes were  $-0.040/85.3$  versus  $-0.089/66.8$ , respectively, Table S-1) because the larger mass and length of the peptide chain in the case of APO-C3 O<sub>94</sub>-glycopeptide prevented its precipitation (see Table S-1).

### 3.3. Optimization of rhEPO O<sub>126</sub>-Glycopeptide Enrichment by Precipitation with Acetone

The previous results showed that the amount of acetone could be finely tuned for targeting the glycoforms of a specific O-glycopeptide biomarker. rhEPO O<sub>126</sub>-glycopeptide was chosen as a model to optimize the amount of acetone necessary to achieve the maximum precipitation of O<sub>126</sub> glycoforms with N-glycopeptides, but with the minimal coprecipitation of peptides. This glycopeptide was selected due to its importance in biopharmaceutical analysis and doping control because O<sub>126</sub>-glycopeptide glycoforms containing NeuGc are being investigated as biomarkers to differentiate endogenous from recombinant erythropoietins (OG4, H1N1S1S1, Table S-1).<sup>11-13</sup> Five different amounts of ice-cold acetone (4.5-fold, 5-fold, 6-fold, 8-fold, and 9-fold volumes) were added to rhEPO tryptic digests, and both supernatant and precipitate were analyzed by CE-MS. Figure 5 shows the bar plots with the percentages of precipitation obtained for (A-i) O<sub>126</sub>-glycopeptide glycoforms and (B-i) some representative rhEPO peptides. As can be seen and according to our previous results, O-glycopeptides started to precipitate using a five-fold volume of acetone (see Figure 5A-i). As they shared the same peptide chain, the explanation for the order of precipitation must be related to the carbohydrate moiety (Figure 5A-ii). Thus, the glycoforms with a higher percentage of carbohydrates and number of SIA (H1N1S1, H1N1S2, and H1N1S1S1) precipitated first. In contrast, the nonsialylated glycoform (H1N1) started to precipitate using a eight-fold volume of acetone. This indicated that although it does not present SIA, the medium became hydrophobic enough for a sudden and extensive precipitation. With respect to peptides, as was discussed previously, GRAVY and aliphatic index were the most meaningful variables related to their different trend to precipitate. Thus, the most hydrophilic rhEPO peptides (i.e., higher negative GRAVY and lower aliphatic index) began also to precipitate using a five-fold volume of acetone (Figure 5B-ii). In this way, P1, P5, P11, and P14 were precipitated to a considerable extent using eight-fold and nine-fold volume of acetone (Figure 5B-i). P4, P6, P8, P10, and P12 rhEPO peptides were not shown in the bar plot because their percentage of precipitation was less than 1% in all cases (see Table S-1 for the percentage of precipitation with five-fold and eight-fold volumes of acetone). As our aim was focused on maximizing precipitation of all O<sub>126</sub>-glycopeptide glycoforms with the minimal copresence of peptides, an eight-fold volume of ice-cold acetone was selected as the best compromise conditions to precipitate the O<sub>126</sub> glycoforms with N-glycopeptides.

## 4. CONCLUSIONS

The physicochemical parameters and properties related to the structure and composition of a large set of peptides, N- and O-glycopeptides were studied with multivariate data analysis methods to determine the variables responsible of their different trend to precipitate using a five-fold volume of acetone. In this regard, practically all N-glycopeptides were enriched in the precipitate as they present a large and structurally rich glycan moiety ( $\geq 45\%$  (m/m) in all cases), while peptides were mostly remained in the supernatant. With regard to O-glycopeptides, the peptide chain size and the hydrophobicity as well as the lower glycan content in comparison to those of N-glycopeptides explained why they

were only partially precipitated adding a five-fold volume of acetone. The study focused on peptides and O-glycopeptides precipitation using a five-fold and eight-fold volume of acetone revealed that hydrophobicity was critical to explain their different trend to precipitate. In addition, the size of the peptide chain was also essential to explain the different behavior of rhEPO O<sub>126</sub>-glycopeptide and APO-C3 O<sub>94</sub>-glycopeptide, which have similar hydrophobicity. APO-C3 O<sub>94</sub>-glycopeptides did not precipitate because of the larger peptide chain size. Finally, it was demonstrated for the rhEPO O<sub>126</sub>-glycopeptide, which is of interest in biopharmaceutical quality control and doping analysis, that the amount of acetone can be finely tuned for targeting the glycoforms of a certain O-glycopeptide. An eight-fold volume of acetone was selected as the best compromise conditions to precipitate the O<sub>126</sub>-glycopeptide with N-glycopeptides and the minimum coprecipitation of peptides. A similar strategy could be followed for the selective enrichment of many other N- and O-glycopeptides of interest, which may not be easily purified using other methodologies.

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jproteome.7b00524.

Scores and loadings plots of PCA model applied to peptides, N- and O-glycopeptides; scores and loadings plots of PCA model applied to N-glycopeptides; physicochemical parameters and properties of peptides, N- and O-glycopeptides (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Author

\*E-mail: fbenavente@ub.edu. Phone: (+34) 934039778. Fax: (+34) 934021233.

### ORCID

Fernando Benavente: 0000-0002-1688-1477

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

This work was supported by a grant from the Spanish Ministry of Economy and Competitiveness (CTQ2014-56777-R) and the Catedra UB Rector Francisco Buscarons Ubeda (Forensic Chemistry and Chemical Engineering). M.M.-A. acknowledges the University of Barcelona for an ADR fellowship.

## ■ REFERENCES

- (1) Varki, A.; Gagneux, P. Biological functions of glycans. In *Essentials of Glycobiology*; Varki, A., Cummings, R. D., Esko, J. D., Freeze, H. H., Stanley, P., Bertozzi, C. R., Hart, G. W., Etzler, M. E., Eds.; Cold Spring Harbor Laboratory Press: New York, 2009; pp 1–19.
- (2) Barroso, A.; Giménez, E.; Benavente, F.; Barbosa, J.; Sanz-Nebot, V. Classification of congenital disorders of glycosylation based on analysis of transferrin glycopeptides by capillary liquid chromatography-mass spectrometry. *Talanta* **2016**, *160*, 614–623.
- (3) Gornik, O.; Lauc, G. Glycosylation of serum proteins in inflammatory diseases. *Dis. Markers* **2008**, *25*, 267–278.
- (4) Reis, C.; Osorio, H.; Silva, L.; Gomes, C.; David, L. Alterations in glycosylation as biomarkers for cancer detection. *J. Clin. Pathol.* **2010**, *63*, 322–329.
- (5) Pinho, S. S.; Reis, C. A. Glycosylation in cancer: mechanisms and clinical implications. *Nat. Rev. Cancer* **2015**, *15*, 540–555.
- (6) Dube, D. H.; Bertozzi, C. R. Glycans in cancer and inflammation - potential for therapeutics and diagnostics. *Nat. Rev. Drug Discovery* **2005**, *4*, 477–488.
- (7) Porter, S. Human immune response to recombinant human proteins. *J. Pharm. Sci.* **2001**, *90*, 1–11.
- (8) Sauerborn, M.; Brinks, V.; Jiskoot, W.; Schellekens, H. Immunological mechanism underlying the immune response to recombinant human protein therapeutics. *Trends Pharmacol. Sci.* **2010**, *31*, 53–59.
- (9) Brooks, S. A. Appropriate glycosylation of recombinant proteins for human use: implications of choice of expression system. *Mol. Biotechnol.* **2004**, *28*, 241–255.
- (10) Jensen, P. H.; Kolarich, D.; Packer, N. H. Mucin-type O-glycosylation - Putting the pieces together. *FEBS J.* **2010**, *277*, 81–94.
- (11) Pascual, J. A.; Belalcázar, V.; de Bolos, C.; Gutiérrez, R.; Llop, E.; Segura, J. Recombinant erythropoietin and analogues: a challenge for doping control. *Ther. Drug Monit.* **2004**, *26*, 175–179.
- (12) Segura, J.; Pascual, J. A.; Gutiérrez-Gallego, R. Procedures for monitoring recombinant erythropoietin and analogues in doping control. *Anal. Bioanal. Chem.* **2007**, *388*, 1521–1529.
- (13) Giménez, E.; Ramos-Hernán, R.; Benavente, F.; Barbosa, J.; Sanz-Nebot, V. Analysis of recombinant human erythropoietin glycopeptides by capillary electrophoresis electrospray-time of flight-mass spectrometry. *Anal. Chim. Acta* **2012**, *709*, 81–90.
- (14) Wohlgemuth, J.; Karas, M.; Eichhorn, T.; Hendriks, R.; Andrecht, S. Quantitative site-specific analysis of protein glycosylation by LC-MS using different glycopeptide-enrichment strategies. *Anal. Biochem.* **2009**, *395*, 178–188.
- (15) Kašička, V. Recent developments in capillary and microchip electroseparations of peptides (2013-middle 2015). *Electrophoresis* **2016**, *37*, 162–188.
- (16) Štěpánová, S.; Kašička, V. Recent developments and applications of capillary and microchip electrophoresis in proteomic and peptidomic analyses. *J. Sep. Sci.* **2016**, *39*, 198–211.
- (17) Herrero, M.; Ibañez, E.; Cifuentes, A. Capillary electrophoresis-electrospray-mass spectrometry in peptide analysis and peptidomics. *Electrophoresis* **2008**, *29*, 2148–2160.
- (18) Calvano, C. D.; Zamboni, C. G.; Jensen, O. N. Assessment of lectin and HILIC based enrichment protocols for characterization of serum glycoproteins by mass spectrometry. *J. Proteomics* **2008**, *71*, 304–317.
- (19) Nascimento, K. S.; Cunha, A. I.; Nascimento, K. S.; Cavada, B. S.; Azevedo, A. M.; Aires-Barros, M. R. An overview of lectins purification strategies. *J. Mol. Recognit.* **2012**, *25*, 527–541.
- (20) Drake, R. R.; Schwegler, E. E.; Malik, G.; Díaz, J.; Block, T.; Mehta, A.; Semmes, O. J. Lectin capture strategies combined with mass spectrometry for the discovery of serum glycoprotein biomarkers. *Mol. Cell. Proteomics* **2006**, *5*, 1957–1967.
- (21) Madera, M.; Mechref, Y.; Novotny, M. V. Combining lectin microcolumns with high-resolution separation techniques for enrichment of glycoproteins and glycopeptides. *Anal. Chem.* **2005**, *77*, 4081–4090.
- (22) Huang, B. Y.; Yang, C. K.; Liu, C. P.; Liu, C. Y. Stationary phases for the enrichment of glycoproteins and glycopeptides. *Electrophoresis* **2014**, *35*, 2091–2107.
- (23) Palmisano, G.; Lendal, S. E.; Engholm-Keller, K.; Leth-Larsen, R.; Parker, B. L.; Larsen, M. R. Selective enrichment of sialic acid-containing glycopeptides using titanium dioxide chromatography with analysis by HILIC and mass spectrometry. *Nat. Protoc.* **2010**, *5*, 1974–1982.
- (24) Wada, Y.; Tajiri, M.; Yoshida, S. Hydrophilic affinity isolation and MALDI multiple-stage tandem mass spectrometry of glycopeptides for glycoproteomics. *Anal. Chem.* **2004**, *76*, 6560–6565.
- (25) Zhang, B.; Sheng, Q.; Li, X.; Liang, Q.; Yan, J.; Liang, X. Selective enrichment of glycopeptides for mass spectrometry analysis using C18 fractionation and titanium dioxide chromatography. *J. Sep. Sci.* **2011**, *34*, 2745–2750.

- (26) Zhu, J.; Wang, F.; Chen, R.; Cheng, K.; Xu, B.; Guo, Z.; Liang, X.; Ye, M.; Zou, H. Centrifugation assisted microreactor enables facile integration of trypsin digestion, hydrophilic interaction chromatography enrichment, and on-column deglycosylation for rapid and sensitive N-glycoproteome analysis. *Anal. Chem.* **2012**, *84*, 5146–5153.
- (27) Cao, W.; Cao, J.; Huang, J.; Zhang, L.; Yao, J.; Xu, H.; Yang, P. Enhanced N-glycosylation site exploitation of sialoglycopeptides by peptide IPG-IEF assisted TiO<sub>2</sub> chromatography. *Glycoconjugate J.* **2012**, *29*, 433–443.
- (28) Takakura, D.; Harazono, A.; Hashii, N.; Kawasaki, N. Selective glycopeptide profiling by acetone enrichment and LC/MS. *J. Proteomics* **2014**, *101*, 17–30.
- (29) Barroso, A.; Giménez, E.; Benavente, F.; Barbosa, J.; Sanz-Nebot, V. Modelling the electrophoretic migration behaviour of peptides and glycopeptides from glycoprotein digests in capillary electrophoresis-mass spectrometry. *Anal. Chim. Acta* **2015**, *854*, 169–177.
- (30) Simó, C.; Cifuentes, A. Capillary electrophoresis-mass spectrometry of peptides from enzymatic protein hydrolysis: Simulation and optimization. *Electrophoresis* **2003**, *24*, 834–842.
- (31) Simó, C.; González, R.; Barbas, C.; Cifuentes, A. Combining peptide modeling and capillary electrophoresis-mass spectrometry for characterization of enzymes cleavage patterns: Recombinant versus natural bovine pepsin A. *Anal. Chem.* **2005**, *77*, 7709–7716.
- (32) Benavente, F.; Balaguer, E.; Barbosa, J.; Sanz-Nebot, V. Modelling migration behavior of peptide hormones in capillary electrophoresis-electrospray mass spectrometry. *J. Chromatogr. A* **2006**, *1117*, 94–102.
- (33) Sillero, A.; Ribeiro, J. M. Isoelectric points of proteins: Theoretical determination. *Anal. Biochem.* **1989**, *179*, 319–325.
- (34) Kyte, J.; Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **1982**, *157*, 105–132.
- (35) Ikai, A. Thermostability and Aliphatic Index of Globular Proteins. *Biochem. J.* **1980**, *88*, 1895–1898.
- (36) Bober, L.; Koba, M.; Judycka-Proma, U.; Baczek, T. Pharmacological classification of drugs by principal component analysis applying molecular modeling descriptors and HPLC retention data. *J. Chromatogr. Sci.* **2011**, *49*, 758–763.
- (37) Lauria, A.; Ippolito, M.; Almerico, A. M. Principal component analysis on molecular descriptors as an alternative point of view in the search of new Hsp90 inhibitors. *Comput. Biol. Chem.* **2009**, *33*, 386–390.
- (38) Barker, M.; Rayens, W. Partial least squares for discrimination. *J. Chemom.* **2003**, *17*, 166–173.
- (39) Wold, S.; Sjöström, M.; Eriksson, L. PLS-regression: A basic tool of chemometrics. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109–130.
- (40) Cawley, G. C.; Talbot, N. L. C. Efficient leave-one-out cross-validation of kernel fisher discriminant classifiers. *Pattern Recognit.* **2003**, *36*, 2585–2592.
- (41) Stone, M. Cross-Validatory Choice and Assessment of Statistical Predictions. *J. R. Stat. Soc.* **1974**, *36*, 111–147.

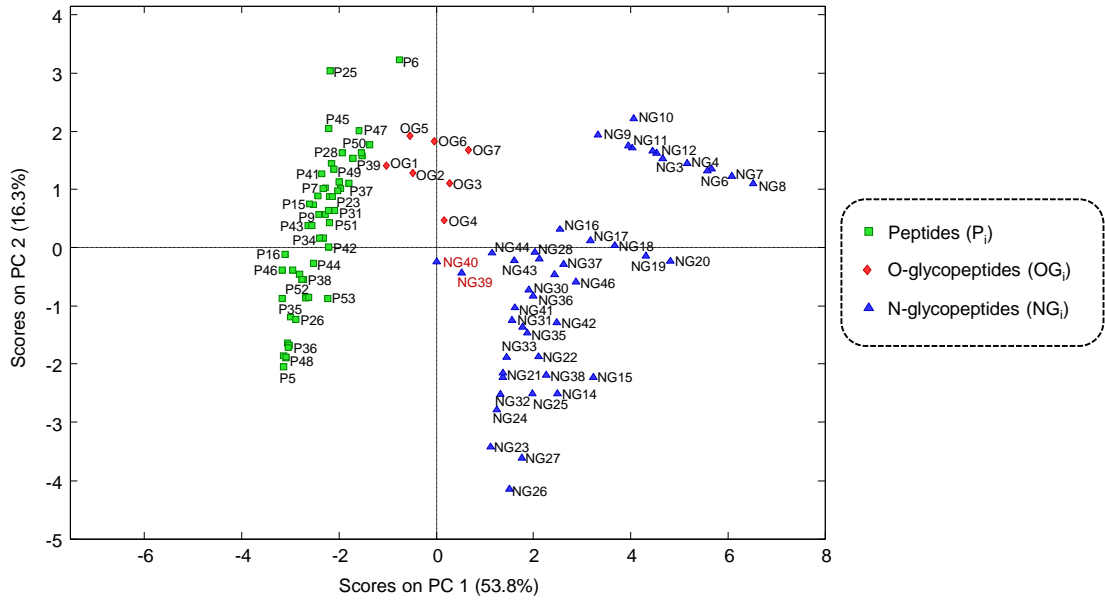


## **Supplementary Figures and Tables**

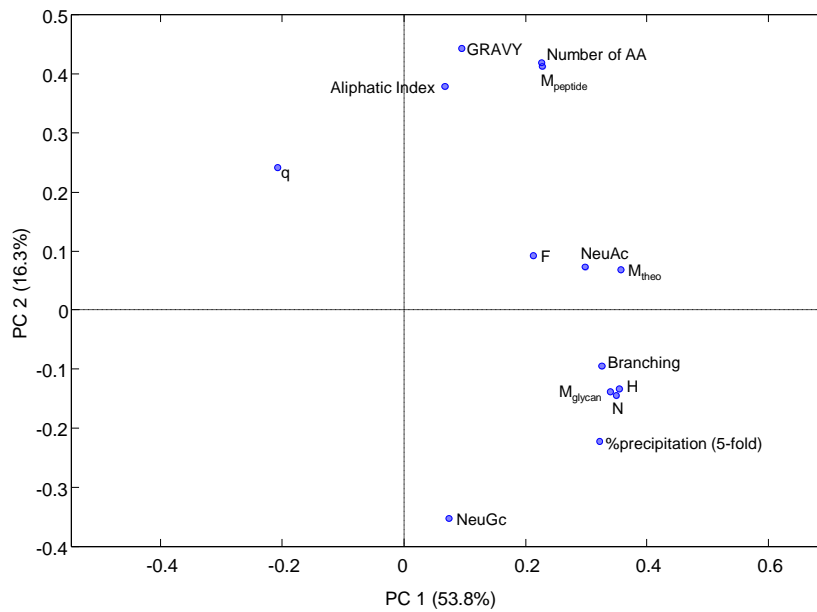
---



A) Scores plot



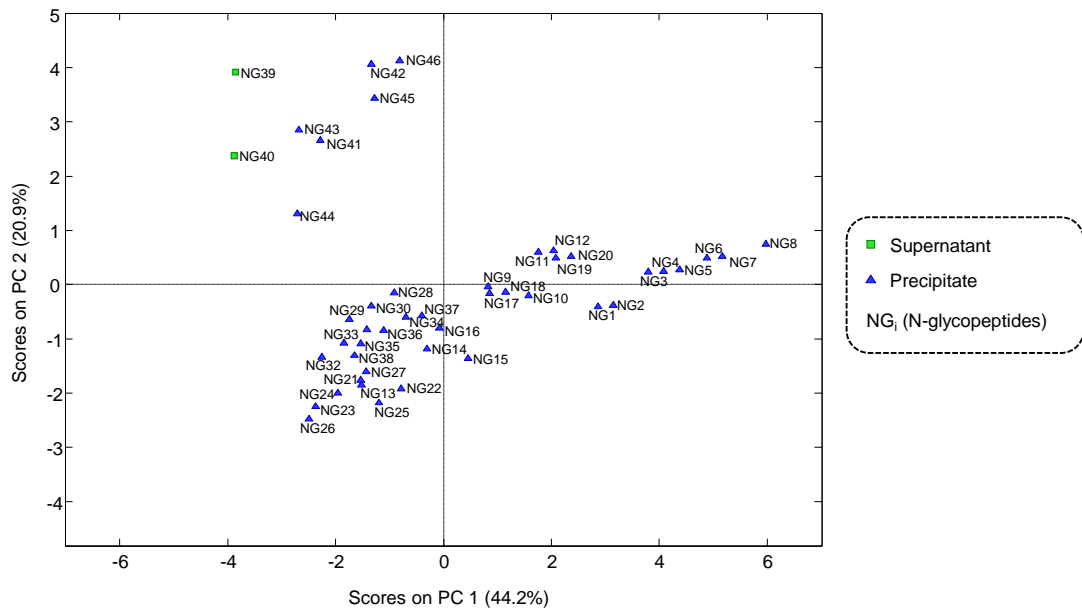
B) Loadings plot



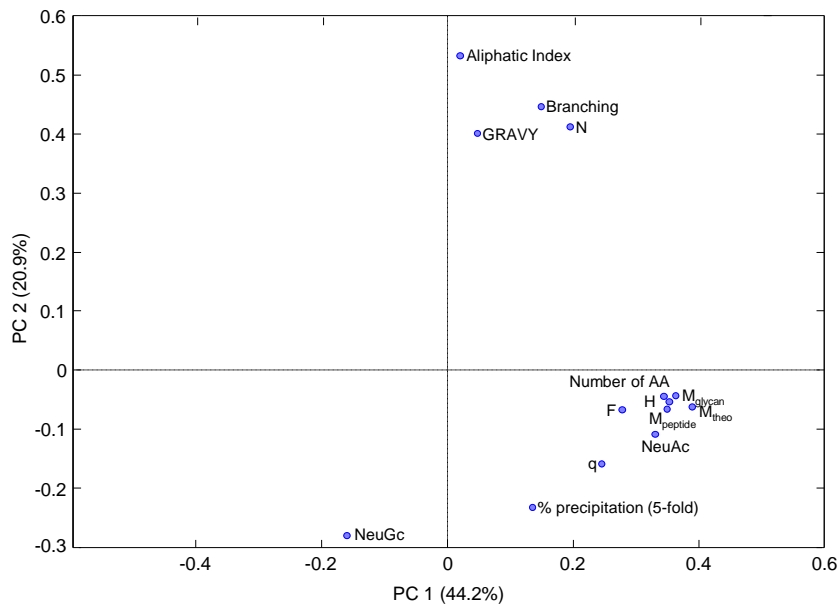
**Supplementary Figure S-1.** A) Scores plot and B) loadings plot of the PCA model applied to the percentage of precipitation adding 5-fold volume of acetone and the different physicochemical parameters and properties of peptides and glycopeptides of the enzymatic digests of ovalbumin (OVA), human and bovin alpha-1-acid glycoprotein (hAGP and bAGP), human apolipoprotein C-III (APO-C3) and rhEPO (see Table 1).



## A) Scores plot



## B) Loadings plot



**Supplementary Figure S-2.** A) Scores plot and B) loadings plot of the PCA model applied to the percentage of precipitation adding 5-fold volume of acetone and the different physicochemical parameters and properties of N-glycopeptides (see Table 1).

**Supplementary Table S-1.** Peptides, N- and O-glycopeptides detected by CE-MS in the tryptic digest of rhEPO, APO-C3, hAGP, BAGP and OVA glycoproteins and the physicochemical parameters and properties used to study the percentage of precipitation in the acetone enrichment method.

| Peptide sequence   | Glycan <sup>a</sup>   | Symbol <sup>b</sup> | M <sub>theo</sub> | M <sub>glycan</sub> | M <sub>peptide</sub> | q    | Number of AA | Hydrophobicity |                 | Monosaccharides |                |       |       |                | Branching | % precipitation |        |
|--|-----------------------|---------------------|-------------------|---------------------|----------------------|------|--------------|----------------|-----------------|-----------------|----------------|-------|-------|----------------|-----------|-----------------|--------|
|  |                       |                     |                   |                     |                      |      |              | GRAVY          | Aliphatic index | H <sup>a</sup>  | N <sup>a</sup> | NeuGc | NeuAc | F <sup>a</sup> |           | 5-fold          | 8-fold |
| <b>Erythropoietin (rhEPO)</b>                            |                       |                     |                   |                     |                      |      |              |                |                 |                 |                |       |       |                |           |                 |        |
| APPR   | -                     | P1                  | 439.2543          | -                   | 439.2543             | 1.90 | 4            | -1.475         | 25.0            | 0               | 0              | 0     | 0     | 0              | 0         | 3.6             | 6.6    |
| VLER   | -                     | P2                  | 515.3067          | -                   | 515.3067             | 1.89 | 4            | 0              | 170.0           | 0               | 0              | 0     | 0     | 0              | 0         | 2.0             | 1.8    |
| YLLEAK   | -                     | P3                  | 735.4167          | -                   | 735.4167             | 1.89 | 6            | 0.117          | 146.7           | 0               | 0              | 0     | 0     | 0              | 0         | 1.6             | 1.0    |
| VNFYAWK  | -                     | P4                  | 926.4650          | -                   | 926.4650             | 1.90 | 7            | -0.114         | 55.7            | 0               | 0              | 0     | 0     | 0              | 0         | 0               | 0.8    |
| R  | -                     | P5                  | 174.1117          | -                   | 174.1117             | 1.90 | 1            | -4.500         | 0               | 0               | 0              | 0     | 0     | 0              | 0         | 13.5            | 48.9   |
| MEVGQQAWE<br>VWQGLALLSE<br>AVLR                          | -                     | P6                  | 2525.3311         | -                   | 2525.3311            | 1.88 | 23           | 0.491          | 131.3           | 0               | 0              | 0     | 0     | 0              | 0         | 0               | 0      |
| GQALLVNSSQ<br>PWEPLQLHV<br>DK<br>(N <sub>83</sub> )      | H6N5F1S2              | NG1                 | 5074.1962         | 2715.9630           | 2358.2332            | 2.22 | 21           | -0.419         | 106.7           | 6               | 5              | 0     | 2     | 1              | 3         | 100             | 100    |
|  | H6N5F1S3              | NG2                 | 5365.2919         | 3007.0587           | 2358.2332            | 1.91 | 21           | -0.419         | 106.7           | 6               | 5              | 0     | 3     | 1              | 3         | 100             | 100    |
|  | H7N6F1S2              | NG3                 | 5439.3279         | 3081.0947           | 2358.2332            | 2.22 | 21           | -0.419         | 106.7           | 7               | 6              | 0     | 2     | 1              | 4         | 100             | 100    |
|  | H7N6F1S3              | NG4                 | 5730.4234         | 3372.1902           | 2358.2332            | 1.91 | 21           | -0.419         | 106.7           | 7               | 6              | 0     | 3     | 1              | 4         | 100             | 100    |
|  | H7N6F1S4              | NG5                 | 6021.5188         | 3663.2856           | 2358.2332            | 1.61 | 21           | -0.419         | 106.7           | 7               | 6              | 0     | 4     | 1              | 4         | 100             | 100    |
|  | H8N7F1S3              | NG6                 | 6095.5563         | 3737.3231           | 2358.2332            | 1.91 | 21           | -0.419         | 106.7           | 8               | 7              | 0     | 3     | 1              | 4         | 100             | 100    |
|  | H8N7F1S4              | NG7                 | 6386.6520         | 4028.4188           | 2358.2332            | 1.61 | 21           | -0.419         | 106.7           | 8               | 7              | 0     | 4     | 1              | 4         | 100             | 100    |
|  | H9N8F1S4              | NG8                 | 6751.7842         | 4393.5510           | 2358.2332            | 1.61 | 21           | -0.419         | 106.7           | 9               | 8              | 0     | 4     | 1              | 4         | 100             | 100    |
| AVSGLR   | -                     | P7                  | 601.3547          | -                   | 601.3547             | 1.90 | 6            | 0.683          | 130.0           | 0               | 0              | 0     | 0     | 0              | 0         | 1.4             | 1.1    |
| SLTTLLR  | -                     | P8                  | 802.4912          | -                   | 802.4912             | 1.90 | 7            | 0.671          | 167.1           | 0               | 0              | 0     | 0     | 0              | 0         | 0               | 0.9    |
| ALGAQK   | -                     | P9                  | 586.3438          | -                   | 586.3438             | 1.90 | 6            | -0.067         | 98.3            | 0               | 0              | 0     | 0     | 0              | 0         | 2.2             | 1.7    |
| EAI <sup>1</sup> SPPDAASA<br>APLR<br>(O <sub>126</sub> ) | H1N1                  | OG1                 | 1829.8895         | 365.1322            | 1464.7573            | 1.84 | 15           | -0.040         | 85.3            | 1               | 1              | 0     | 0     | 0              | 1         | 0               | 71.5   |
|  | H1N1S1                | OG2                 | 2120.9849         | 656.2276            | 1464.7573            | 1.53 | 15           | -0.040         | 85.3            | 1               | 1              | 0     | 1     | 0              | 1         | 7.5             | 59.8   |
|  | H1N1S2                | OG3                 | 2412.0803         | 947.3230            | 1464.7573            | 1.22 | 15           | -0.040         | 85.3            | 1               | 1              | 0     | 2     | 0              | 2         | 13.4            | 77.2   |
|  | H1N1S1S1 <sup>c</sup> | OG4                 | 2428.0752         | 963.3180            | 1464.7573            | 1.22 | 15           | -0.040         | 85.3            | 1               | 1              | 1     | 1     | 0              | 2         | 14.2            | 75.3   |
| TITADTFR   | -                     | P10                 | 923.4712          | -                   | 923.4712             | 1.85 | 8            | -0.125         | 61.3            | 0               | 0              | 0     | 0     | 0              | 0         | 0               | 0.9    |
| K  | -                     | P11                 | 146.1055          | -                   | 146.1055             | 1.90 | 1            | -3.900         | 0               | 0               | 0              | 0     | 0     | 0              | 0         | 9.8             | 22.2   |
| LFR  | -                     | P12                 | 434.2641          | -                   | 434.2641             | 1.90 | 3            | 0.700          | 130.0           | 0               | 0              | 0     | 0     | 0              | 0         | 0               | 0      |
| VYSNFLR  | -                     | P13                 | 897.4708          | -                   | 897.4708             | 1.90 | 7            | 0.100          | 97.1            | 0               | 0              | 0     | 0     | 0              | 0         | 2.7             | 1.2    |
| GK   | -                     | P14                 | 203.1270          | -                   | 203.1270             | 1.90 | 2            | -2.150         | 0               | 0               | 0              | 0     | 0     | 0              | 0         | 3.2             | 9.6    |

|  |          |      |           |           |           |      |    |        |       |   |   |   |   |   |   |     |      |
|--|----------|------|-----------|-----------|-----------|------|----|--------|-------|---|---|---|---|---|---|-----|------|
| LK   | -        | P15  | 259.1896  | -         | 259.1896  | 1.90 | 2  | -0.050 | 195.0 | 0 | 0 | 0 | 0 | 0 | 0 | 2.0 | 1.7  |
| <b>Apolipoprotein C-III (APOC3)</b>                    |          |      |           |           |           |      |    |        |       |   |   |   |   |   |   |     |      |
| HATK   | -        | P16  | 455.2492  | -         | 455.2492  | 2.90 | 4  | -1.500 | 25.0  | 0 | 0 | 0 | 0 | 0 | 0 | 0.6 | 1.2  |
| TAK  | -        | P17  | 318.1903  | -         | 318.1903  | 1.90 | 3  | -0.933 | 33.3  | 0 | 0 | 0 | 0 | 0 | 0 | 0   | 0    |
| DALSSVQESQ<br>VAQQAR                                   | -        | P18  | 1715.8438 | -         | 1715.8438 | 1.84 | 16 | -0.644 | 79.4  | 0 | 0 | 0 | 0 | 0 | 0 | 2.5 | 37.5 |
| GWVTDGFSSL<br>K  | -        | P19  | 1195.5873 | -         | 1195.5873 | 1.85 | 11 | -0.055 | 61.8  | 0 | 0 | 0 | 0 | 0 | 0 | 0.3 | 0.5  |
| DYWSTVK  | -        | P20  | 897.4232  | -         | 897.4232  | 1.85 | 7  | -0.986 | 41.4  | 0 | 0 | 0 | 0 | 0 | 0 | 1.6 | 0.7  |
| DK   | -        | P21  | 261.1325  | -         | 261.1325  | 1.85 | 2  | -3.700 | 0     | 0 | 0 | 0 | 0 | 0 | 0 | 1.7 | 4.9  |
| FSEFWDLDPE<br>VRPTSAVAA<br>(O <sub>94</sub> )          | H1N1     | OG5  | 2501.1486 | 365.1322  | 2136.0164 | 1.78 | 19 | -0.089 | 66.8  | 1 | 1 | 0 | 0 | 0 | 1 | 0   | 0    |
|  | H1N1S1   | OG6  | 2792.2440 | 656.2276  | 2136.0164 | 1.47 | 19 | -0.089 | 66.8  | 1 | 1 | 0 | 1 | 0 | 1 | 0   | 1.9  |
|  | H1N1S2   | OG7  | 3083.3394 | 947.3230  | 2136.0164 | 1.16 | 19 | -0.089 | 66.8  | 1 | 1 | 0 | 2 | 0 | 2 | 0   | 2.1  |
| <b>Human alpha-1-acid glycoprotein (hAGP)</b>          |          |      |           |           |           |      |    |        |       |   |   |   |   |   |   |     |      |
| QIPLCANLVP<br>VPITNATLDQI<br>TGK<br>(N <sub>15</sub> ) | H5N4S2   | NG9  | 4780.1768 | 2204.7725 | 2575.4044 | 1.23 | 24 | 0.446  | 130.0 | 5 | 4 | 0 | 2 | 0 | 2 | 100 | 100  |
|  | H5N4F1S2 | NG10 | 4926.2347 | 2350.8304 | 2575.4044 | 1.23 | 24 | 0.446  | 130.0 | 5 | 4 | 0 | 2 | 1 | 2 | 100 | 100  |
|  | H6N5S2   | NG11 | 5145.3090 | 2569.9047 | 2575.4044 | 1.23 | 24 | 0.446  | 130.0 | 6 | 5 | 0 | 2 | 0 | 3 | 100 | 100  |
|  | H6N5S3   | NG12 | 5436.4044 | 2861.0001 | 2575.4044 | 0.92 | 24 | 0.446  | 130.0 | 6 | 5 | 0 | 3 | 0 | 3 | 100 | 100  |
| WFYIASAFR  | -        | P22  | 1159.5814 | -         | 1159.5814 | 1.90 | 9  | 0.689  | 65.6  | 0 | 0 | 0 | 0 | 0 | 0 | 2.7 | 3.9  |
| NEEYNK<br>(N <sub>38</sub> )                           | H5N4S2   | NG13 | 3000.1124 | 2204.7725 | 795.3399  | 1.27 | 6  | -3.200 | 0     | 5 | 4 | 0 | 2 | 0 | 2 | 100 | 100  |
|  | H6N5S3   | NG14 | 3656.3400 | 2861.0001 | 795.3399  | 0.96 | 6  | -3.200 | 0     | 6 | 5 | 0 | 3 | 0 | 3 | 100 | 100  |
|  | H6N5F1S3 | NG15 | 3802.3979 | 3007.0580 | 795.3399  | 0.96 | 6  | -3.200 | 0     | 6 | 5 | 0 | 3 | 1 | 3 | 100 | 100  |
| SVQEIQATFFY<br>FTPNK<br>(N <sub>54</sub> )             | H5N4S2   | NG16 | 4123.7190 | 2204.7725 | 1918.9465 | 1.28 | 16 | -0.256 | 48.8  | 5 | 4 | 0 | 2 | 0 | 2 | 100 | 100  |
|  | H6N5S2   | NG17 | 4488.8512 | 2569.9047 | 1918.9465 | 1.28 | 16 | -0.256 | 48.8  | 6 | 5 | 0 | 2 | 0 | 3 | 100 | 100  |
|  | H6N5S3   | NG18 | 4779.9466 | 2861.0001 | 1918.9465 | 0.97 | 16 | -0.256 | 48.8  | 6 | 5 | 0 | 3 | 0 | 3 | 100 | 100  |
|  | H7N6S3   | NG19 | 5145.0788 | 3226.1323 | 1918.9465 | 0.97 | 16 | -0.256 | 48.8  | 7 | 6 | 0 | 3 | 0 | 4 | 100 | 100  |
|  | H7N6S4   | NG20 | 5436.1742 | 3517.2277 | 1918.9465 | 0.66 | 16 | -0.256 | 48.8  | 7 | 6 | 0 | 4 | 0 | 4 | 100 | 100  |
| TEDTIFLR   | -        | P23  | 993.5131  | -         | 993.5131  | 1.84 | 8  | -0.225 | 97.5  | 0 | 0 | 0 | 0 | 0 | 0 | 1.1 | 2.3  |
| EYQTR  | -        | P24  | 695.3238  | -         | 695.3238  | 1.89 | 5  | -2.700 | 0     | 0 | 0 | 0 | 0 | 0 | 0 | 1.7 | 16.5 |
| YVGGQEHFA<br>HLILR                                     | -        | P25  | 1751.9471 | -         | 1751.9471 | 3.89 | 15 | 0.313  | 130.0 | 0 | 0 | 0 | 0 | 0 | 0 | 0   | 3.3  |
| DTK  | -        | P26  | 362.1801  | -         | 362.1801  | 1.85 | 3  | -2.700 | 0     | 0 | 0 | 0 | 0 | 0 | 0 | 4.5 | 39.2 |

|  |                         |           |           |           |           |      |        |        |       |   |   |   |   |   |     |      |      |     |
|--|-------------------------|-----------|-----------|-----------|-----------|------|--------|--------|-------|---|---|---|---|---|-----|------|------|-----|
| TYMLAFDVN<br>DEK                               | -                       | P27       | 1444.6544 | -         | 1444.6544 | 1.79 | 12     | -0.450 | 65.0  | 0 | 0 | 0 | 0 | 0 | 0   | 1.3  | 30.4 |     |
| NWGLSVYAD<br>KPETTK                            | -                       | P28       | 1707.8468 | -         | 1707.8468 | 2.84 | 15     | -0.993 | 52.0  | 0 | 0 | 0 | 0 | 0 | 0   | 8.3  | 66.7 |     |
| EQLGEFYEAL<br>DCLR                             | -                       | P29       | 1741.7981 | -         | 1741.7981 | 1.83 | 14     | -0.371 | 90.7  | 0 | 0 | 0 | 0 | 0 | 0   | 1.2  | 23.2 |     |
| IPK  | -                       | P30       | 356.2423  | -         | 356.2423  | 1.90 | 3      | -0.333 | 130.0 | 0 | 0 | 0 | 0 | 0 | 0   | 0    | 0    |     |
| SDVVYTDWK                                      | -                       | P31       | 1111.5186 | -         | 1111.5186 | 1.79 | 9      | -0.689 | 64.4  | 0 | 0 | 0 | 0 | 0 | 0   | 1.1  | 3.8  |     |
| K  | -                       | P32       | 146.1055  | -         | 146.1055  | 1.90 | 1      | -3.900 | 0     | 0 | 0 | 0 | 0 | 0 | 0   | 16.7 | 38.9 |     |
| DK   | -                       | P33       | 261.1325  | -         | 261.1325  | 1.85 | 2      | -3.700 | 0     | 0 | 0 | 0 | 0 | 0 | 0   | 7.5  | 48.1 |     |
| CEPLEK   | -                       | P34       | 774.3582  | -         | 774.3582  | 1.89 | 6      | -1.033 | 65.0  | 0 | 0 | 0 | 0 | 0 | 0   | 4.2  | 46.4 |     |
| QHEK   | -                       | P35       | 540.2656  | -         | 540.2656  | 2.89 | 4      | -3.525 | 0     | 0 | 0 | 0 | 0 | 0 | 0   | 8.1  | 61.8 |     |
| ER   | -                       | P36       | 303.1543  | -         | 303.1543  | 1.89 | 2      | -4.000 | 0     | 0 | 0 | 0 | 0 | 0 | 0   | 6.0  | 44.2 |     |
| <b>Bovine alpha-1-acid glycoprotein (bAGP)</b> |                         |           |           |           |           |      |        |        |       |   |   |   |   |   |     |      |      |     |
| WFYIGSAFR                                      | -                       | P37       | 1145.5658 | -         | 1145.5658 | 1.90 | 9      | 0.444  | 54.4  | 0 | 0 | 0 | 0 | 0 | 0   | 3.3  | 2.7  |     |
| NPEY <sup>NK</sup><br>(N <sub>39</sub> )       | H5N4S2                  | NG21      | 2968.1225 | 2204.7724 | 763.3501  | 1.28 | 6      | -2.883 | 0     | 5 | 4 | 0 | 2 | 0 | 2   | 100  | 100  |     |
|  | H5N4F1S2                | NG22      | 3114.1804 | 2350.8303 | 763.3501  | 1.28 | 6      | -2.883 | 0     | 5 | 4 | 0 | 2 | 1 | 2   | 100  | 100  |     |
|  | H5N4S2 <sup>c</sup>     | NG23      | 3000.1123 | 2236.7622 | 763.3501  | 1.28 | 6      | -2.883 | 0     | 5 | 4 | 2 | 0 | 0 | 2   | 100  | 100  |     |
|  | H5N4S1S1 <sup>c</sup>   | NG24      | 2984.1174 | 2220.7673 | 763.3501  | 1.28 | 6      | -2.883 | 0     | 5 | 4 | 1 | 1 | 1 | 0   | 2    | 100  | 100 |
|  | H5N4F1S1S1 <sup>c</sup> | NG25      | 3130.1753 | 2366.8252 | 763.3501  | 1.28 | 6      | -2.883 | 0     | 5 | 4 | 1 | 1 | 1 | 2   | 100  | 100  |     |
|  | H5N4S3 <sup>c</sup>     | NG26      | 3307.2026 | 2543.8525 | 763.3501  | 0.97 | 6      | -2.883 | 0     | 5 | 4 | 3 | 0 | 0 | 2   | 100  | 100  |     |
| H6N5S2 <sup>c</sup>                            | NG27                    | 3365.2445 | 2601.8944 | 763.3501  | 1.28      | 6    | -2.883 | 0      | 6     | 5 | 2 | 0 | 0 | 3 | 100 | 100  |      |     |
| SAR  | -                       | P38       | 332.1808  | -         | 332.1808  | 1.90 | 3      | -1.167 | 33.3  | 0 | 0 | 0 | 0 | 0 | 0   | 6.4  | 8.2  |     |
| AIQAFFYLEP<br>R                                | -                       | P39       | 1424.7452 | -         | 1424.7452 | 1.89 | 12     | 0.408  | 90.0  | 0 | 0 | 0 | 0 | 0 | 0   | 7.6  | 22.7 |     |
| HAEDK  | -                       | P40       | 598.2711  | -         | 598.2711  | 2.84 | 5      | -2.460 | 20.0  | 0 | 0 | 0 | 0 | 0 | 0   | 15.4 | 72.6 |     |
| LITR   | -                       | P41       | 501.3275  | -         | 501.3275  | 1.90 | 4      | 0.775  | 195.0 | 0 | 0 | 0 | 0 | 0 | 0   | 2.1  | 1.5  |     |
| EYQTIEDK                                       | -                       | P42       | 1024.4713 | -         | 1024.4713 | 1.83 | 8      | -1.925 | 48.8  | 0 | 0 | 0 | 0 | 0 | 0   | 10.3 | 74.3 |     |
| CVY <sup>NCS</sup> FIK<br>(N <sub>76</sub> )   | H5N4S2                  | NG28      | 3394.2984 | 2204.7724 | 1189.5260 | 1.28 | 9      | 0.778  | 75.6  | 5 | 4 | 0 | 2 | 0 | 2   | 100  | 100  |     |
|  | H5N4S2 <sup>c</sup>     | NG29      | 3426.2882 | 2236.7622 | 1189.5260 | 1.28 | 9      | 0.778  | 75.6  | 5 | 4 | 2 | 0 | 0 | 2   | 100  | 100  |     |
|  | H5N4S1S1 <sup>c</sup>   | NG30      | 3410.2933 | 2220.7673 | 1189.5260 | 1.28 | 9      | 0.778  | 75.6  | 5 | 4 | 1 | 1 | 0 | 2   | 100  | 100  |     |
| IYR  | -                       | P43       | 450.2591  | -         | 450.2591  | 1.90 | 3      | -0.433 | 130.0 | 0 | 0 | 0 | 0 | 0 | 0   | 4.0  | 2.7  |     |

|   |                       |           |           |           |           |      |       |        |       |   |   |   |   |   |     |      |      |
|---|-----------------------|-----------|-----------|-----------|-----------|------|-------|--------|-------|---|---|---|---|---|-----|------|------|
| QNGTSLK<br>(N <sub>86</sub> )           | H5N4S2                | NG31      | 2951.1647 | 2204.7724 | 746.3923  | 1.28 | 7     | -1.286 | 55.7  | 5 | 4 | 0 | 2 | 0 | 2   | 100  | 100  |
|   | H5N4S2 <sup>c</sup>   | NG32      | 2983.1545 | 2236.7622 | 746.3923  | 1.28 | 7     | -1.286 | 55.7  | 5 | 4 | 2 | 0 | 0 | 2   | 100  | 100  |
|   | H5N4S1S1 <sup>c</sup> | NG33      | 2967.1596 | 2220.7673 | 746.3923  | 1.28 | 7     | -1.286 | 55.7  | 5 | 4 | 1 | 1 | 0 | 2   | 100  | 100  |
| VESDR                                   | -                     | P44       | 604.2816  | -         | 604.2816  | 1.84 | 5     | -1.620 | 58.0  | 0 | 0 | 0 | 0 | 0 | 0   | 8.3  | 55.5 |
| EHFVDLLSK                               | -                     | P45       | 1199.6550 | -         | 1199.6550 | 2.84 | 10    | 0.350  | 146.0 | 0 | 0 | 0 | 0 | 0 | 0   | 2.1  | 6.4  |
| HFR                                     | -                     | P46       | 458.2390  | -         | 458.2390  | 2.90 | 3     | -1.633 | 0     | 0 | 0 | 0 | 0 | 0 | 0   | 2.0  | 2.4  |
| TFMLAASWN<br>GTK<br>(N <sub>118</sub> ) | H5N4S2                | NG34      | 3530.4162 | 2204.7724 | 1325.6438 | 1.28 | 12    | 0.100  | 49.2  | 5 | 4 | 0 | 2 | 0 | 2   | 100  | 100  |
|   | H5N4S2 <sup>c</sup>   | NG35      | 3562.4060 | 2236.7622 | 1325.6438 | 1.28 | 12    | 0.100  | 49.2  | 5 | 4 | 2 | 0 | 0 | 2   | 100  | 100  |
|   | H5N4S1S1 <sup>c</sup> | NG36      | 3546.4111 | 2220.7673 | 1325.6438 | 1.28 | 12    | 0.100  | 49.2  | 5 | 4 | 1 | 1 | 0 | 2   | 100  | 100  |
|   | H5N4S3                | NG37      | 3821.5116 | 2495.8678 | 1325.6438 | 0.97 | 12    | 0.100  | 49.2  | 5 | 4 | 0 | 3 | 0 | 2   | 100  | 100  |
| H5N4S3 <sup>c</sup>                     | NG38                  | 3869.4963 | 2543.8525 | 1325.6438 | 0.97      | 12   | 0.100 | 49.2   | 5     | 4 | 3 | 0 | 0 | 2 | 100 | 100  |      |
| NVGVSFYADK<br>PEVTQEQQK                 | -                     | P47       | 2038.0007 | -         | 2038.0007 | 2.83 | 18    | -0.911 | 53.9  | 0 | 0 | 0 | 0 | 0 | 0   | 18.7 | 83.5 |
| K                                       | -                     | P48       | 146.1055  | -         | 146.1055  | 1.90 | 1     | -3.900 | 0     | 0 | 0 | 0 | 0 | 0 | 0   | 15.5 | 50.1 |
| EFLDVIK                                 | -                     | P49       | 862.4800  | -         | 862.4800  | 1.84 | 7     | 0.629  | 152.9 | 0 | 0 | 0 | 0 | 0 | 0   | 1.9  | 4.0  |
| CIGIQESEIYT<br>DEK                      | -                     | P50       | 1796.8502 | -         | 1796.8502 | 1.83 | 15    | -0.273 | 104.0 | 0 | 0 | 0 | 0 | 0 | 0   | 15.7 | 81.7 |
| DACGPLEK                                | -                     | P51       | 888.4011  | -         | 888.4011  | 1.84 | 8     | -0.600 | 61.3  | 0 | 0 | 0 | 0 | 0 | 0   | 9.2  | 54.6 |
| QHEEER                                  | -                     | P52       | 826.3569  | -         | 826.3569  | 2.88 | 6     | -3.617 | 0     | 0 | 0 | 0 | 0 | 0 | 0   | 58.7 | 87.6 |
| ETEAS                                   | -                     | P53       | 535.2126  | -         | 535.2126  | 0.89 | 5     | -1.340 | 20.0  | 0 | 0 | 0 | 0 | 0 | 0   | 8.6  | 50.2 |
| <b>Ovalbumin (OVA)</b>                  |                       |           |           |           |           |      |       |        |       |   |   |   |   |   |     |      |      |
| <b>NL</b><br>(N <sub>293</sub> )        | H3N6                  | NG39      | 1949.7722 | 1704.6347 | 245.1375  | 0.90 | 2     | 0.150  | 195.0 | 3 | 6 | 0 | 0 | 0 | 4   | 0    | -    |
|   | H5N4                  | NG40      | 1867.7191 | 1622.5816 | 245.1375  | 0.90 | 2     | 0.150  | 195.0 | 5 | 4 | 0 | 0 | 0 | 2   | 0    | -    |
|   | H5N6                  | NG41      | 2273.8779 | 2028.7404 | 245.1375  | 0.90 | 2     | 0.150  | 195.0 | 5 | 6 | 0 | 0 | 0 | 4   | 100  | -    |
|   | H5N8                  | NG42      | 2680.0366 | 2434.8991 | 245.1375  | 0.90 | 2     | 0.150  | 195.0 | 5 | 8 | 0 | 0 | 0 | 6   | 100  | -    |
| <b>NLTSVL</b><br>(N <sub>293</sub> )    | H3N6                  | NG43      | 2350.0044 | 1704.6347 | 645.3697  | 0.90 | 6     | 1.133  | 178.3 | 3 | 6 | 0 | 0 | 0 | 4   | 100  | -    |
|   | H5N4                  | NG44      | 2267.9514 | 1622.5817 | 645.3697  | 0.90 | 6     | 1.133  | 178.3 | 5 | 4 | 0 | 0 | 0 | 2   | 100  | -    |
|   | H5N7                  | NG45      | 2877.1895 | 2231.8198 | 645.3697  | 0.90 | 6     | 1.133  | 178.3 | 5 | 7 | 0 | 0 | 0 | 5   | 100  | -    |
|   | H5N8                  | NG46      | 3080.2688 | 2434.8991 | 645.3697  | 0.90 | 6     | 1.133  | 178.3 | 5 | 8 | 0 | 0 | 0 | 6   | 100  | -    |

Glycosylation site is marked in red.

<sup>a</sup> H, N, F, S correspond to hexose, N-acetylhexose, fucose, and sialic acid monosaccharides, respectively.

<sup>b</sup> P: peptides, NG: N-glycopeptides and OG: O-glycopeptides.

<sup>c</sup> One sialic acid (H<sub>5</sub>N<sub>4</sub>S1S1) or all sialic acids (H<sub>5</sub>N<sub>4</sub>S<sub>n</sub>) are N-glycolylneuraminic acids instead of N-acetylneuraminic acids.



Contents lists available at ScienceDirect

Talanta

journal homepage: [www.elsevier.com/locate/talanta](http://www.elsevier.com/locate/talanta)

## Analysis of glycopeptide biomarkers by on-line TiO<sub>2</sub> solid-phase extraction capillary electrophoresis-mass spectrometry



Montserrat Mancera-Arteu, Nejsi Lleshi, Victoria Sanz-Nebot<sup>\*</sup>, Estela Giménez, Fernando Benavente

Department of Chemical Engineering and Analytical Chemistry, Institute for Research on Nutrition and Food Safety (INSA-UB), University of Barcelona, Martí I Franquès 1-11, 08028, Barcelona, Spain

### ARTICLE INFO

#### Keywords:

Capillary electrophoresis  
Glycopeptides  
Mass spectrometry  
In-line solid-phase extraction  
On-line solid-phase extraction  
Titanium dioxide

### ABSTRACT

In this study is described an on-line titanium dioxide solid-phase extraction capillary electrophoresis-mass spectrometry (TiO<sub>2</sub>-SPE-CE-MS) method for the analysis of the glycopeptide glycoforms obtained from the tryptic digests of recombinant human erythropoietin (rhEPO). The O<sub>126</sub>-glycopeptide of rhEPO was used to optimize the methodology given its importance in quality control of biopharmaceuticals and doping analysis. Several aspects that affect the selective retention and elution, peak efficiency and electrophoretic separation of the O<sub>126</sub> glycoforms were investigated to maximize detection sensitivity while minimizing non-specific retention of peptides. Under the optimized conditions, the microcartridge lifetime was around 10 analyses and repeatability was acceptable (%RSD values of 9–11% and 6–11% for migration times and peak areas, respectively). The method was linear between 0.5 and 50 mg L<sup>-1</sup> and 10–50 mg L<sup>-1</sup> for O<sub>126</sub> glycoforms containing NeuAc and NeuGc, respectively, and limits of detection (LODs) were up to 100 times lower than by CE-MS. Although optimized for O-glycopeptides, the method proved also successful for pre-concentration of N<sub>83</sub>-glycopeptides, without compromising the separation between glycopeptide glycoforms with different number of sialic acids. Tryptic digests of other glycoproteins (i.e. human apolipoprotein CIII (APO-C3) and bovine alpha-1-acid glycoprotein (bAGP)) were also analyzed, demonstrating the applicability to glycopeptides with different glycan composition and nature.

### 1. Introduction

In proteins, glycosylation is one of the most usual post-translational modifications, with more than half of the proteins in mammalian proteomes being glycosylated. The carbohydrate chains (glycans) are also known to play a crucial role in many processes such as recognition, signaling and adhesion on the cell surfaces [1,2]. Furthermore, N- and O-glycan structures are modified in recombinant biopharmaceuticals and altered in a wide range of disorders including cancer, inflammatory processes and alcoholism, among others [3–8]. Therefore, in recent years, characterization of aberrant glycosylation has aroused biomedical and biotechnological interest in order to detect and/or monitor several pathologies as well as to characterize recombinant biopharmaceuticals or discriminate between endogenous and recombinant variants of glycoproteins.

Human erythropoietin (hEPO) is a glycoprotein hormone produced mainly in the kidneys and responsible of maintaining red blood cell number and tissue oxygen supply at adequate levels. It shows four

glycosylation sites, of which three are occupied by complex type N-glycans (at Asn24, Asn38 and Asn83) and one by an O-glycan (at Ser126). Recombinant variants of this glycoprotein (rhEPO) are widely used in the treatment of anemia in chronic kidney disease and cancer [9,10]. Nevertheless, rhEPO has become particularly popular in the last decades due to its misuse in endurance sport disciplines as a doping agent, hence forbidden by sport authorities since 1989 [11,12]. The presence of about 2% of non-human N-glycolylneuraminic acid (NeuGc) in the glycans of rhEPO O<sub>126</sub>-glycosylation site enables distinction between recombinant and endogenous forms of hEPO in doping control. Furthermore, this content needs to be controlled to ensure the quality and safety of biopharmaceutical products due to adverse effects of NeuGc-contaminated pharmaceuticals [13]. Analysis of glycopeptide and glycan biomarkers could be explored as an alternative to intact protein analysis, which presents certain limitations due to the high molecular mass and the detrimental effect of sugars in ionization [14,15]. In contrast to glycans, glycopeptide biomarkers offer information about the structure and composition of the carbohydrate

<sup>\*</sup> Corresponding author.

E-mail address: [vsanz@ub.edu](mailto:vsanz@ub.edu) (V. Sanz-Nebot).

<https://doi.org/10.1016/j.talanta.2019.120563>

Received 1 August 2019; Received in revised form 4 November 2019; Accepted 12 November 2019

Available online 15 November 2019

0039-9140/ © 2019 Elsevier B.V. All rights reserved.

moiety, as well as about glycosylation sites and their degree of occupation.

Capillary electrophoresis-mass spectrometry (CE-MS) has become a powerful hyphenated technique in glycoproteomics, tackling the analysis of intact glycoproteins, glycopeptides and glycans [7,16,17]. In addition to the CE-MS methods for the analysis of intact rhEPO [14,15], CE-MS methods for the analysis of rhEPO glycopeptides have previously been established for a more detailed protein characterization [18]. However, CE, as other microscale separation techniques, has low concentration sensitivity mainly due to the reduced sample volume injected to obtain optimum separations [19,20]. Over the years, different strategies have been described to decrease the limits of detection in CE [19–23]. Among these strategies, on-line solid-phase extraction capillary electrophoresis mass spectrometry (SPE-CE-MS) is one of the most versatile and successful alternatives [19–21]. In the typical SPE-CE configuration, a microcartridge, filled with a sorbent that shows affinity for the compounds of interest, is integrated in-line near the inlet of the separation capillary in order to clean-up and preconcentrate the target analytes from a large volume of sample (typically ~50–100 µL). Analysis of a broad variety of small molecules, peptides and proteins using different sorbents have been demonstrated by SPE-CE-MS [19–21], but not yet with titanium dioxide (TiO<sub>2</sub>) sorbents or for the selective analysis of glycopeptides. In this regard, TiO<sub>2</sub> beads have been used for a long time for the off-line purification of phosphopeptides, or other phosphorylated analytes [24–26], and more recently for sialylated glycopeptides [27–29]. The mechanism of retention of both phosphopeptides and glycopeptides is mainly attributed to interaction between the negatively charged phosphate or carboxyl groups, respectively, with the positively charged TiO<sub>2</sub> surface [26,28].

In this paper, for the first time to the best of our knowledge, TiO<sub>2</sub> beads were used as a sorbent in TiO<sub>2</sub>-SPE-CE-MS to selectively analyze glycopeptides from the typical enzymatic digests of glycoproteins prepared in bottom-up proteomics approaches. The O<sub>126</sub>-glycopeptide glycoforms of rhEPO were used to optimize the method, due to their relevance in quality control of biopharmaceuticals and doping analysis. Afterwards, the method was also evaluated for the analysis of rhEPO N-glycopeptides. Finally, it was applied to the tryptic digests of human apolipoprotein C-III (APO-C3), which contains only O-glycopeptides, and bovine alpha-1-acid glycoprotein (bAGP), which contains only N-glycopeptides.

## 2. Experimental section

### 2.1. Chemicals

All chemicals used in the preparation of buffers and solutions were of analytical reagent grade. Acetic acid (HAc, glacial), formic acid (HFor 98–100%), sodium hydroxide and ammonia (25%) were supplied by Merck (Darmstadt, Germany). DL-Dithiothreitol (DTT, ≥99%), iodoacetamide (IAA, ≥98%), ammonium hydrogen carbonate (≥99.9%) and lactic acid (≥99.9%) were purchased from Sigma-Aldrich (St. Louis, MO, USA). Isopropanol was provided by Scharlab (Barcelona, Spain) while acetonitrile and water by Sigma-Aldrich (all of them of LC-MS quality grade). Trypsin (sequencing grade modified, 16000 U mg<sup>-1</sup>) was purchased from Promega (Madison, WI, USA). ESI low concentration (ESI-L) tuning mix was supplied by Agilent Technologies (Waldbronn, Germany) for tuning and calibration of the mass spectrometer.

### 2.2. Glycoprotein samples

rhEPO produced in a chinese hamster ovary (CHO) cell line was provided by the European Pharmacopeia as a Biological Reference Product (BRP-lot4). Each sample vial contained 100 µg of rhEPO (a mixture of alpha and beta rhEPO), 24 mg of D-(+)-trehalose, 2.4 mg of arginine hydrochloride, 0.08 mg of Tween-20, 3.6 mg of NaCl and

2.5 mg of NaH<sub>2</sub>PO<sub>4</sub>·2H<sub>2</sub>O. The content of each vial was dissolved in water to obtain a 1000 mg L<sup>-1</sup> protein solution. Excipients of low-molecular mass were removed by ultracentrifugation using Microcon YM-10 centrifugal filters from Millipore (M<sub>r</sub> cut-off 10,000, Bedford, MA, USA) as described elsewhere [18]. Centrifugations were carried out in a Mikro 20 centrifuge (Hettich, Tuttlingen, Germany) at room temperature. bAGP (99%) and apoC-III (APO-C3) were obtained from Sigma-Aldrich. Stock solutions of 1000 mg L<sup>-1</sup> were prepared in water and aliquoted. Aliquots were evaporated to dryness using a Savant SPD-111 V SpeedVac concentrator (Thermo-Fisher Scientific, Waltham, MA, USA) and stored at -20 °C until enzymatic digestion.

rhEPO and bAGP, which contain cysteine residues that form disulfide bonds, were firstly reduced and alkylated to facilitate digestion. Briefly, an aliquot of 50 µg of dried glycoprotein was dissolved in 50 µL of 50 mM NH<sub>4</sub>HCO<sub>3</sub> (pH 7.9) and 2.5 µL of 0.5 M DTT in the same buffer was added. The mixture was incubated in a thermoshaker at 56 °C for 30 min and then alkylated by adding 7 µL of 50 mM IAA in 50 mM NH<sub>4</sub>HCO<sub>3</sub> (pH 7.9) and shaking for 30 min at room temperature in the dark. Excess of low-molecular mass reagents was removed with Microcon YM-10 centrifugal filters as described elsewhere [18]. The final protein residue was dissolved in 50 mM NH<sub>4</sub>HCO<sub>3</sub> (pH 7.9) to obtain a final concentration of 1000 mg L<sup>-1</sup>. *Trypsin digestion:* a 50 µL aliquot of 1000 mg L<sup>-1</sup> protein solution in 50 mM NH<sub>4</sub>HCO<sub>3</sub> (pH 7.9) was digested. APO-C3 digest was directly prepared from aliquots of 10 µg of dried glycoprotein, while rhEPO and bAGP digests were prepared from aliquots of 50 µg of reduced and alkylated dried glycoprotein. Trypsin was added in an enzyme to protein ratio of 1:40 m/m. The mixture was vortexed and subsequently incubated at 37 °C for 18 h. Digestion was stopped by heating at 100 °C for 10 min, and the digest was stored at -20 °C until analysis [18]. All digestions were performed in triplicate. Incubations were performed in a TS-100 thermoshaker (Biosan, Riga, Latvian Republic).

### 2.3. CE-MS

CE-MS experiments were performed in a HP<sup>300</sup>CE system coupled to a 6220 oa-TOF LC/MS mass spectrometer with an orthogonal sheath-flow interface (Agilent Technologies). The sheath liquid (50:50 (v/v) iPrOH/H<sub>2</sub>O with 0.05% (v/v) of HFor) was delivered at a flow rate of 3.3 µL min<sup>-1</sup> by a KD Scientific 100 series infusion pump (Holliston, MS, USA) and degassed for 10 min by sonication before use. CE control and separation data acquisition (e.g. voltage, temperature and current) were performed using Chemstation software (Agilent Technologies) that was running in combination with the MassHunter workstation software (Agilent Technologies) for control, data acquisition and processing of the mass spectrometer. The mass spectrometer was tuned and calibrated following the manufacturer's instructions. A "check tune" of the instrument was performed every day in positive mode to ensure accurate mass assignments. Instrumental parameters were optimized for the analysis of rhEPO O<sub>126</sub>- and N<sub>83</sub>-glycopeptides in a previous study [18]. The optimized operational conditions in positive electrospray ionization (ESI) mode were: capillary voltage 4000 V, drying gas (N<sub>2</sub>) temperature 200 °C, drying gas flow rate 4 L min<sup>-1</sup>, nebulizer gas (N<sub>2</sub>) 10 psig, fragmentor voltage 190 V, skimmer voltage 60 V and OCT 1 RF Vpp voltage 300 V. Data were collected in profile (continuum) at 1 spectrum s<sup>-1</sup> (approx. 10,000 transients spectrum<sup>-1</sup>) between m/z 100 and 3200 working in the highest resolution mode (4 GHz).

A bare fused-silica capillary of 70 cm total length (L<sub>T</sub>) x 75 µm internal diameter (I.D.) x 360 µm outer diameter (O.D.) (Polymicro Technologies, Phoenix, AZ, USA) was used in CE-MS. Activation and conditioning procedures were carried out off-line in order to avoid contamination with NaOH of the mass spectrometer. New capillaries were activated by flushing (930 mbar) sequentially for 30 min each with 1 M NaOH, water and BGE (50 mM HAc and 50 mM HFor, pH 2.2). Capillaries were conditioned every day by flushing with NaOH (5 min), water (7 min) and BGE (10 min). Samples were reconstituted with BGE

and injected for 15 s at 50 mbar. Electrophoretic separations were performed at 25 °C and 25 kV under normal polarity (cathode in the outlet). Between runs, capillaries were flushed with water (1 min), 1 M HAc (3 min), water (1 min) and BGE (5 min). Capillaries were stored overnight filled with water. Before CE-MS, all solutions were passed through a 0.22- $\mu$ m nylon filter (MSI, Westboro, MS, USA).

pH measurements were carried out with a Crison 2002 potentiometer and a Crison electrode 52–03 (Crison instruments, Barcelona, Spain).

#### 2.4. SPE-CE-MS

The TiO<sub>2</sub>-coated magnetic beads ( $\leq 25 \mu\text{m}$  spherical) were obtained from the Pierce Magnetic TiO<sub>2</sub> Phosphopeptide Enrichment Kit (Thermo Scientific, Massachusetts, US). Bare fused silica capillaries (Polymicro Technologies) were used for all the procedures.

Construction of the single-frit particle-packed microcartridge for TiO<sub>2</sub>-SPE-CE-MS was carried out as described elsewhere with little modifications [30]. The inlet end was prepared by connecting the microcartridge (0.7 cm L<sub>T</sub>  $\times$  250  $\mu\text{m}$  i.d.  $\times$  365  $\mu\text{m}$  o.d. capillary) with a plastic sleeve to a previously conditioned inlet capillary (7.5 cm L<sub>T</sub>  $\times$  75  $\mu\text{m}$  i.d.  $\times$  365  $\mu\text{m}$  o.d.). The microcartridge was completely filled by vacuum with TiO<sub>2</sub> sorbent beads. Another plastic sleeve was connected to the microcartridge, and a small piece of cotton (approximately 1 mm) was placed in the plastic tube before connecting the separation capillary (64.5 cm L<sub>T</sub>  $\times$  75  $\mu\text{m}$  i.d.  $\times$  365  $\mu\text{m}$  o.d.). This cotton frit in the microcartridge outlet end prevented the TiO<sub>2</sub> particles from leaking. Particle leaking would promote current instability or breakdown and poor reproducibility. Under the optimized conditions, the TiO<sub>2</sub> sorbent was first conditioned by flushing (930 mbar) for 2 min with binding buffer (80% v/v acetonitrile, 10% v/v HFor and 0.1 M lactic acid). Afterwards, protein digests were reconstituted in loading buffer (80% v/v acetonitrile, 2% v/v HFor and 0.1 M lactic acid) to the desired concentration and were loaded by flushing for 10 min (60  $\mu\text{L}$ , estimated with the Hagen-Poiseuille equation [31]). A final flush for 1 min with binding buffer and 1 min with washing buffer (80% v/v acetonitrile, 2% v/v HFor) eliminated non-specifically retained molecules. All these steps were performed by switching off the nebulizer gas and the ESI capillary voltage to prevent the entrance of contaminants into the MS. Both were switched on and the capillary was filled by flushing for 2 min with BGE (the same as in CE-MS) to equilibrate the capillary before the electrophoretic separation. Then, a small volume of eluent (1 M NH<sub>4</sub>OH) was injected at 50 mbar for 20 s (110 nL) [31] and, in order to prevent the eluent plug from traveling backwards due to the microcartridge backpressure, BGE was also injected at 25 mbar for 120 s. Separation was conducted at +20 kV for 35 min (cathode in the outlet) applying also a 50 mbar pressure to counteract microcartridge backpressure. Postconditioning to avoid carryover was performed by flushing for 1 min with water, followed by injection of eluent (50 mbar, 40 s) and application of a 100 mbar pressure during 10 min from the water vial.

#### 2.5. Quality parameters

All quality parameters were calculated from data obtained by measuring peak area and migration time ( $t_m$ ) from the extracted ion electropherogram (EIE) of rhEPO glycopeptide glycoforms (considering the  $m/z$  of the most abundant molecular ions, i.e. ions with charges +2, +3 and +4). Repeatability was evaluated as the percent relative standard deviation (%RSD) of peak areas and migration times obtained in consecutive analysis of digested rhEPO at 10 mg L<sup>-1</sup> ( $n = 3$ ). Linearity was studied by analyzing digested rhEPO at concentrations of protein between 25 and 1000 mg L<sup>-1</sup> for CE-MS and between 0.5 and 100 mg L<sup>-1</sup> for SPE-CE-MS. An estimation of the limits of detection (LODs) was obtained by analyzing digested rhEPO at low-concentrations (close to the LOD level, as determined from the approach based on

S/N = 3). The LODs of the different glycopeptide types (i.e. O<sub>126</sub>-glycopeptide glycoforms containing NeuAc, O<sub>126</sub>-glycopeptide glycoforms containing NeuGc and the most relevant N<sub>83</sub>-glycopeptide glycoforms) were established at the lowest concentration in which all the expected glycoforms in each case were detected. The lifetime of the microcartridges was evaluated by repeatedly analyzing digested rhEPO at a concentration of 10 mg L<sup>-1</sup>.

### 3. Results and discussion

#### 3.1. CE-MS

With the aim of establishing a robust and reliable method to analyze sialylated glycopeptides by TiO<sub>2</sub>-SPE-CE-MS, glycopeptides from the tryptic digest of rhEPO were first analyzed by CE-MS. rhEPO was chosen as a model glycoprotein for its importance in quality control of biopharmaceutical and doping analysis, but also because of the broad microheterogeneity, as it presents both N- and O- glycosylation with varying degrees of glycan branching and sialylation. The coverage of peptides and glycopeptides in the tryptic digest of rhEPO was mapped by CE-MS using a method previously developed in our group with a minor modification [18,32]. That methodology used a bare fused capillary of 50  $\mu\text{m}$  i.d. for the electrophoretic separation of peptides and glycopeptides with an acidic BGE (50 mM HAc:50 mM HFor, pH 2.2) and a sheath liquid of iPrOH:H<sub>2</sub>O (60:40, v/v; 0.05% HFor). By contrast, in this study, a 75  $\mu\text{m}$  i.d. capillary was preferred, as a wider capillary provided better performance later by TiO<sub>2</sub>-SPE-CE-MS. In the typical unidirectional SPE-CE configuration, the microcartridge is mounted in series to the separation capillary, and TiO<sub>2</sub> microcartridges were introducing a certain backpressure that resulted in poor performance with 50  $\mu\text{m}$  i.d. capillaries [20].

Samples of rhEPO at 1000 mg L<sup>-1</sup> were digested with trypsin and analyzed by CE-MS in positive mode. Table 1 shows the average peak area and migration time of all the detected peptides and glycoforms of O<sub>126</sub>- and N<sub>83</sub>-glycopeptides and the mass error calculated from three replicate measurements. Full peptide sequence coverage and the detected glycoforms of O<sub>126</sub>- and N<sub>83</sub>-glycopeptides agreed with our previous studies [18,32]. It is worth mentioning, that digestion with trypsin also results in a 2N-glycopeptide (N<sub>2,4</sub>-N<sub>38</sub>) that is poorly detected by CE-MS because its size and total charge promote very poor ionization efficiency. In general, the peak area values obtained for peptides were significantly large, which could therefore hinder the ionization yield of both O<sub>126</sub>- and N<sub>83</sub>-glycopeptides. Furthermore, as expected, the peak area values of the O<sub>126</sub>-glycopeptide glycoforms containing N-acetylneuraminic acid (NeuAc) were substantially higher than those containing N-glycolylneuraminic acid (NeuGc), which were found at very low concentrations. Regarding the N<sub>83</sub>-glycopeptide, some of the peak area values of the glycoforms were very small due to the low abundance and the high sialylated carbohydrate content, which promotes lower ionization yields. Fig. 1 shows the extracted ion electropherograms (EIEs) of the most abundant O<sub>126</sub> (A-i) and N<sub>83</sub> (B-i) glycopeptide glycoforms. As can be observed, baseline separation of O<sub>126</sub> and N<sub>83</sub>-glycopeptide glycoforms containing different number of sialic acids were achieved with a 75  $\mu\text{m}$  i.d. capillary, as in our previous studies with a 50  $\mu\text{m}$  i.d. capillary [18,32]. Linearity of the method was investigated at concentrations of digested rhEPO varying from 25 to 1000 mg L<sup>-1</sup> for the most abundant O<sub>126</sub> glycopeptide glycoforms, which contained NeuAc, and from 100 to 1000 mg L<sup>-1</sup> for the most relevant N<sub>83</sub>-glycopeptide glycoforms, which presented 4Ant, 2, 3 or 4 NeuAc and 1 Fuc (see Fig. 1A-ii and 1B-ii, respectively), with coefficients of determination of R<sup>2</sup> > 0.99. With respect to the O<sub>126</sub>-glycopeptide, the LOD was established at 25 mg L<sup>-1</sup> of digested rhEPO to detect all glycoforms containing NeuAc and at 100 mg L<sup>-1</sup> of digested rhEPO for the less abundant glycoforms with NeuGc. In the case of N<sub>83</sub> glycoforms, they were not detected at concentrations lower than 100 mg L<sup>-1</sup> of digested rhEPO.



**Table 1**

Summary of peptides, O<sub>126</sub><sup>a</sup> and N<sub>83</sub><sup>b</sup> glycopeptide glycoforms detected by CE-MS and TiO<sub>2</sub>-SPE-CE-MS in the tryptic digest of rhEPO at a concentration of 1000 mg L<sup>-1</sup> and 100 mg L<sup>-1</sup> of digested glycoprotein, respectively (n = 3).

|                            | Sequence                                 | M <sub>theo</sub> | CE-MS                    |                              |       |                      | SPE-CE-MS |                              |      |
|----------------------------|--|-------------------|--------------------------|------------------------------|-------|----------------------|-----------|------------------------------|------|
|                            |  |                   | Error <sup>c</sup> (ppm) | Area (x10 <sup>6</sup> a.u.) | %RSD  | t <sub>M</sub> (min) | %RSD      | Area (x10 <sup>6</sup> a.u.) |      |
| Peptides                   | APPR                                     | 439.2543          | 3.2                      | 1.0                          | 4.3   | 5.0                  | 1.4       | 0.01                         |      |
|                            | LICDSR-(Cys-IAA <sup>d</sup> )           | 762.3694          | 1.1                      | 2.8                          | 2.7   | 5.6                  | 1.3       | 0.3                          |      |
|                            | VLER                                     | 515.3067          | 3.4                      | 8.9                          | 4.9   | 5.2                  | 2.7       | 0.05                         |      |
|                            | YLLEAK                                   | 735.4167          | 4.2                      | 8.3                          | 5.6   | 5.5                  | 2.6       | 0.07                         |      |
|                            | VNFYAWK                                  | 926.4650          | 1.2                      | 4.1                          | 1.5   | 5.6                  | 1.3       | 0.009                        |      |
|                            | MEVGQQAQVEVWQGLALLSEAVLR                 | 2525.3312         | 5.3                      | 0.3                          | 3.1   | 6.8                  | 3.1       | 0.02                         |      |
|                            | AVSGLR                                   | 601.3548          | 1.9                      | 4.6                          | 0.002 | 5.4                  | 1.3       | 0.01                         |      |
|                            | SLTTLLR                                  | 802.4913          | 7.3                      | 19.2                         | 7.7   | 5.6                  | 1.3       | 0.01                         |      |
|                            | ALGAQK                                   | 586.3439          | 1.7                      | 3.6                          | 0.3   | 5.3                  | 2.7       | 0.1                          |      |
|                            | TITADTFR                                 | 923.4712          | 2.4                      | 5.4                          | 4.0   | 5.8                  | 2.4       | 0.03                         |      |
|                            | VYSNFLR                                  | 897.4709          | 1.4                      | 3.8                          | 0.7   | 5.6                  | 2.5       | 0.2                          |      |
|                            | LYTGEACR-(Cys-IAA <sup>d</sup> )         | 968.4386          | 1.1                      | 3.7                          | 4.9   | 5.8                  | 2.4       | 0.2                          |      |
|                            | TGDR                                     | 447.2077          | 5.8                      | 0.05                         | 2.5   | 5.2                  | 2.7       | 0.05                         |      |
|                            | O <sub>126</sub> glycoforms <sup>e</sup> | /0NeuAc           | 1829.8895                | 5.4                          | 0.5   | 0.9                  | 6.6       | 1.1                          | 0.5  |
|                            |  | /1NeuAc           | 2120.9849                | 4.5                          | 13.7  | 2.3                  | 7.5       | 2.8                          | 10.7 |
| /2NeuAc                    |  | 2412.0803         | 4.9                      | 4.8                          | 2.6   | 8.4                  | 2.5       | 6.2                          |      |
| /1NeuGc                    |  | 2136.9798         | 5.7                      | 0.2                          | 1.8   | 7.5                  | 2.8       | 0.1                          |      |
| /1NeuGc1NeuAc              |  | 2428.0752         | 3.0                      | 0.1                          | 8.6   | 8.4                  | 2.5       | 0.1                          |      |
| N <sub>83</sub> glycoforms | /3Ant2NeuAc1Fuc                          | 5074.1962         | 9.1                      | 0.06                         | 11.6  | 7.9                  | 1.8       | 0.1                          |      |
|                            | /3Ant3NeuAc1Fuc                          | 5365.2919         | 5.7                      | 0.2                          | 4.9   | 8.5                  | 3.3       | 0.3                          |      |
|                            | /4Ant2NeuAc1Fuc                          | 5439.3279         | 5.1                      | 0.08                         | 12.3  | 7.9                  | 2.7       | 0.09                         |      |
|                            | /4Ant3NeuAc1Fuc                          | 5730.4234         | 3.7                      | 0.6                          | 4.2   | 8.5                  | 2.5       | 0.5                          |      |
|                            | /4Ant4NeuAc1Fuc                          | 6021.5188         | 1.9                      | 1.5                          | 6.6   | 9.1                  | 3.9       | 1.1                          |      |
|                            | /4Ant1LacNAc2NeuAc1Fuc                   | 5804.4602         | 1.1                      | 0.09                         | 9.7   | 8.0                  | 2.7       | 0.09                         |      |
|                            | /4Ant1LacNAc3NeuAc1Fuc                   | 6095.5556         | 2.2                      | 0.8                          | 0.2   | 8.5                  | 2.5       | 0.5                          |      |
|                            | /4Ant1LacNAc4NeuAc1Fuc                   | 6386.6510         | 3.5                      | 1.9                          | 1.5   | 9.1                  | 4.7       | 1.4                          |      |
|                            | /4Ant2LacNAc3NeuAc1Fuc                   | 6460.6878         | 6.5                      | 0.3                          | 8.5   | 8.5                  | 3.3       | 0.3                          |      |
|                            | /4Ant2LacNAc4NeuAc1Fuc                   | 6751.7832         | 4.6                      | 0.7                          | 12.1  | 9.1                  | 3.1       | 0.5                          |      |

Detected peptides with less than 4 amino acids were not considered (Je, R, K, LFR, GK, LK).

<sup>a</sup> Error was calculated in ppm as:  $(|M_{exp} - M_{theo}|/M_{theo}) \times 10^6$  (exp = experimental and theo = theoretical).

<sup>b</sup> IAA stands for iodoacetamide.

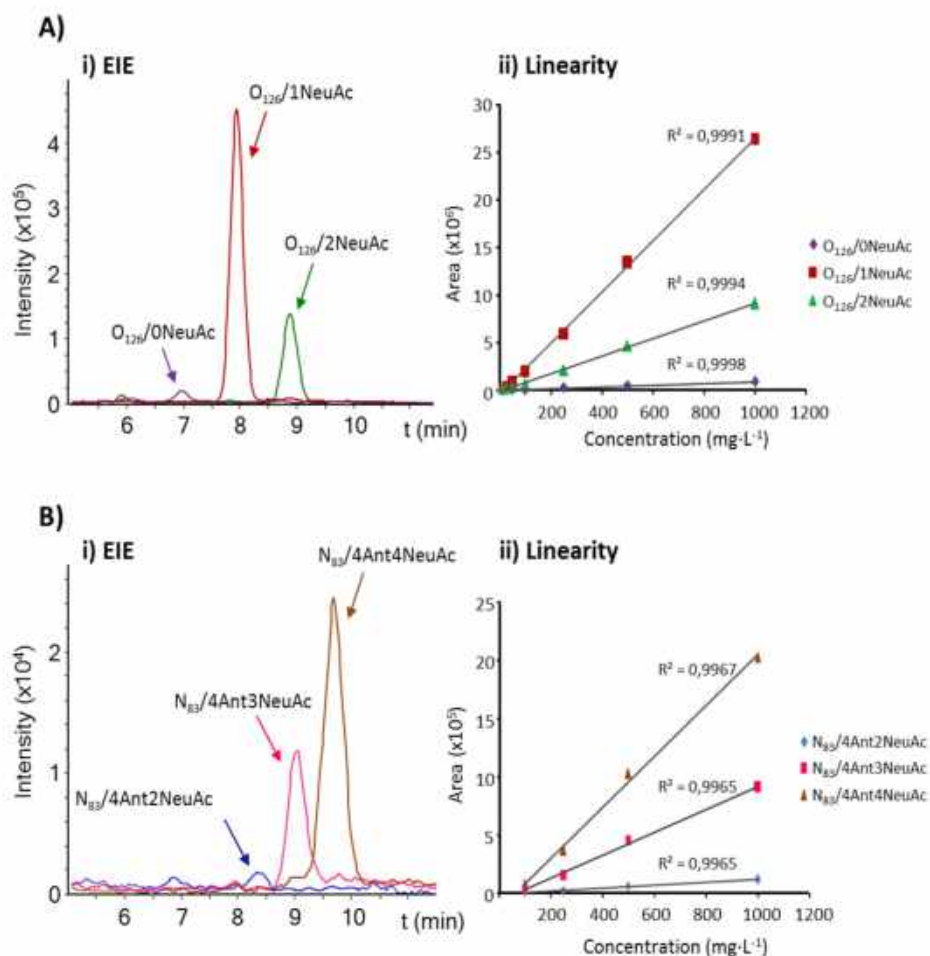
<sup>c</sup> O-glycopeptide glycoforms are composed by one N-acetylglucosamine and one galactose monosaccharides plus the indicated number of sialic acids in each case.

### 3.2. TiO<sub>2</sub>-SPE-CE-MS optimization

The TiO<sub>2</sub> beads obtained from the commercial kit designed for off-line purification of phosphopeptides and sialylated glycopeptides were used to prepare the microcartridges for TiO<sub>2</sub>-SPE-CE-MS. The composition of the different buffers supplied in the kit was unknown and non-compatible with mass spectrometry, as we confirmed with some preliminary experiments. Typically, phosphopeptides and glycopeptides have been selectively enriched off-line with TiO<sub>2</sub> using hydroorganic mixtures with high contents of acetonitrile and HFor or trifluoroacetic acid for conditioning, loading and washing, and NH<sub>4</sub>OH solutions for the elution [27–29,33,34]. In this study, the TiO<sub>2</sub> microcartridges were conditioned using a binding buffer containing 80% v/v acetonitrile and 10% v/v HFor. In the preliminary experiments, 50 mg L<sup>-1</sup> rhEPO tryptic digests were loaded at 930 mbar for 5 min in a loading buffer with 80% v/v of acetonitrile and 2% v/v of HFor. After washing with binding buffer (1 min) and loading buffer (1 min) and filling with BGE (50 mM HAc and 50 mM HFor, pH 2.2) (2 min), the retained glycopeptides were eluted with 1 M NH<sub>4</sub>OH. Some experiments were also performed washing only with loading buffer (2 min), but the short washing step with binding buffer was necessary to reduce non-specific retention of peptides, which were in part eluted during this washing step (see Supplementary Fig. 1). The graph bar of Fig. 2A shows the total peak area of peptides and the different O<sub>126</sub> glycoforms detected by TiO<sub>2</sub>-SPE-CE-MS under different conditions. As can be observed (see the bars labelled as “preliminary conditions”), a small amount of peptides were still retained in the TiO<sub>2</sub> sorbent, but non-specific retention was lower compared to selective retention of the O<sub>126</sub> glycoforms. Anyway, these preliminary results needed to be improved.

Most authors using TiO<sub>2</sub> for off-line purification of phosphopeptides and glycopeptides use lactic or glycolic acid in the loading and binding

buffers to improve the selectivity by reducing unspecific binding of peptides [24,27]. According to them, 0.1 M of lactic acid was added to the binding and loading buffers and a 50 mg L<sup>-1</sup> rhEPO digest was analyzed by TiO<sub>2</sub>-SPE-CE-MS. Under these conditions, retention of peptides decreased while the intensity of O<sub>126</sub> glycoforms significantly increased, as can be seen in Fig. 2A. Higher concentrations than 0.1 M of lactic acid produced higher noise and current instabilities during separation. Using 0.1 M of lactic acid in the loading and binding buffer, it was also tested the effect of adding an extra washing step with a lower percentage of acetonitrile (20% v/v of acetonitrile with 2% v/v of HFor, 1 min) in order to remove the most hydrophilic peptides non-specifically retained in the sorbent, as proposed by other authors in off-line protocols [27,29]. However, while the total peptide area was scarcely reduced (Fig. 2A), the area of O<sub>126</sub> glycoforms drastically decreased. This demonstrated that glycopeptides were partially lost when the TiO<sub>2</sub> sorbent was washed with high water contents. Therefore, this extra washing step was discarded, and binding and loading buffers with 80% v/v of acetonitrile, 0.1 M lactic acid and 10% or 2% v/v of HFor, respectively, were used for the rest of experiments. At this point, it is also important to note that, despite the TiO<sub>2</sub> sorbent has typically been described to retain sialylated glycopeptides, the O<sub>126</sub>/0NeuAc glycoform was also detected using these conditions (see Fig. 2), as also reported for other non-sialylated glycopeptides [28]. This was probably due to non-specific interactions with the peptide moiety and hydrophilic interactions with the attached glycans. Another remarkable fact is that the separation between glycoforms with different number of sialic acids was maintained, despite 50 mbar of pressure were necessary to be applied during separation to counteract the microcartridge backpressure (see in Fig. 2Bi the separation for the O<sub>126</sub> NeuAc glycoforms). Even applying this positive pressure, migration times increased around 10 min compared to CE-MS (Fig. 1Ai).



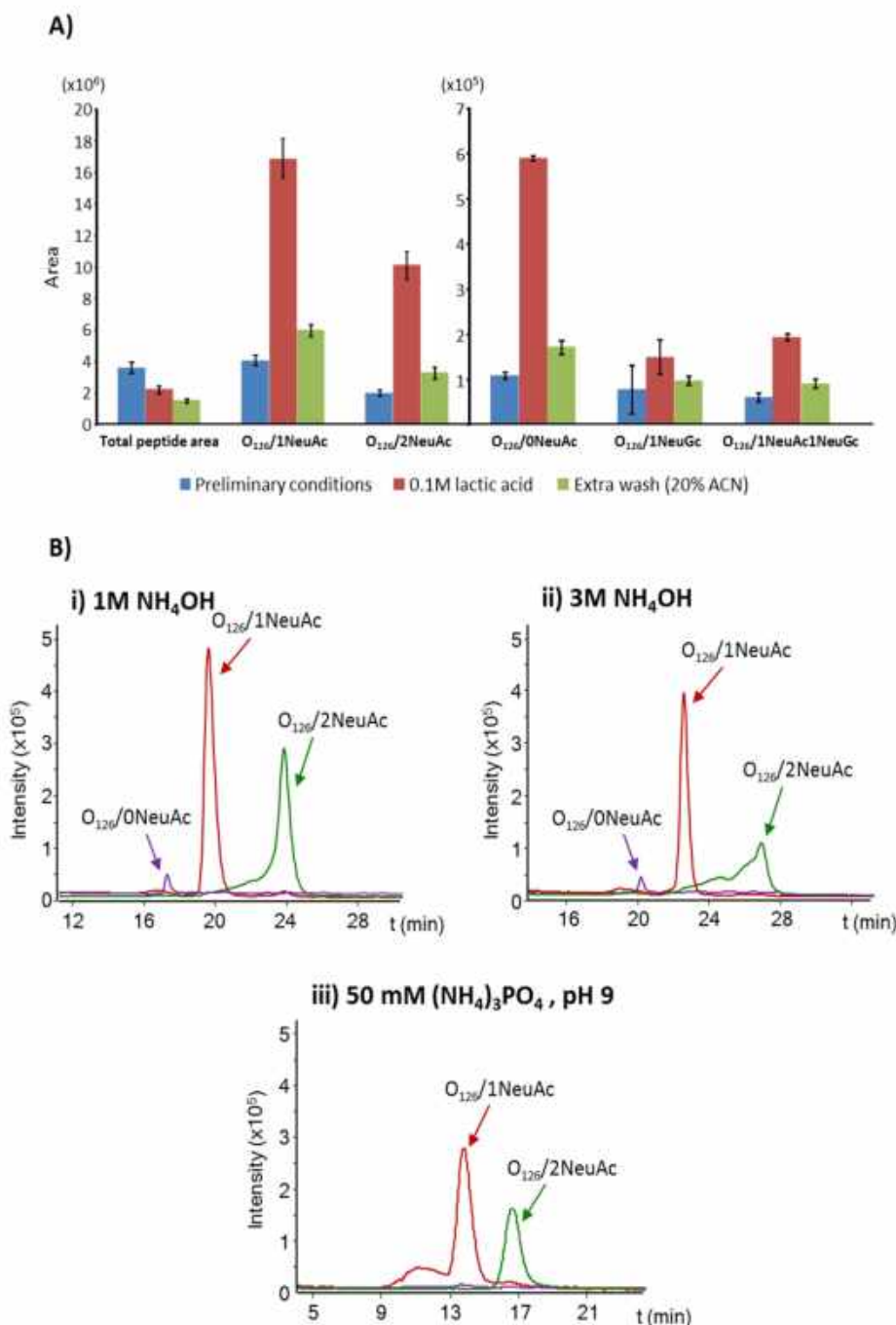
**Fig. 1.** CE-MS analysis of rhEPO tryptic digest. A) Extracted ion electropherograms (EIEs) of the most abundant  $O_{126}$ -glycopeptide glycoforms ( $1000 \text{ mg L}^{-1}$  rhEPO) (i) and linearity ( $25\text{--}1000 \text{ mg L}^{-1}$  rhEPO) (ii). B) EIEs of the most relevant  $N_{83}$ -glycopeptide glycoforms ( $1000 \text{ mg L}^{-1}$  rhEPO) (i) and linearity ( $100\text{--}1000 \text{ mg L}^{-1}$  rhEPO) (ii).

The eluent composition was also investigated trying to improve the results obtained with the  $1 \text{ M NH}_4\text{OH}$  solution. Some authors have suggested the use of higher  $\text{NH}_4\text{OH}$  concentrations than  $1 \text{ M}$  to elute highly sialylated glycopeptides in off-line procedures [29,35]. To test this hypothesis, a  $10 \text{ mg L}^{-1}$  rhEPO digest was analyzed using as eluents  $1 \text{ M}$ ,  $3 \text{ M}$  and  $6 \text{ M NH}_4\text{OH}$  solutions. As can be observed comparing Fig. 2B-i and 2B-ii, the peak area of  $O_{126}$  glycoforms decreased when  $3 \text{ M NH}_4\text{OH}$  was used, especially the one with higher sialic acid content ( $O_{126}/2\text{NeuAc}$ ). Results were even worse when using  $6 \text{ M NH}_4\text{OH}$ , obtaining data with increased noise and observing current instabilities. The use of ammonium phosphate ( $\text{pH } 9$ ) as eluent was also evaluated. Due to the high affinity of the phosphate groups for the  $\text{TiO}_2$  sorbent, it was expected to improve glycopeptide recoveries. Nonetheless, the peak area of  $O_{126}$  glycoforms decreased, being unable to detect  $O_{126}/0\text{NeuAc}$  glycoform (see Fig. 2B-iii). Finally, an acidic eluent ( $0.1\% \text{ v/v HFor}$ ) was also tested [28], but the obtained results were very poor. Therefore, the  $1 \text{ M NH}_4\text{OH}$  solution was confirmed as the eluent for the analysis of glycopeptides by  $\text{TiO}_2\text{-SPE-CE-MS}$ . Under these conditions, sample loading time was investigated loading a  $10 \text{ mg L}^{-1}$  rhEPO digest for 5, 10, 20 and 30 min at 930 mbar. The peak area of  $O_{126}$ -glycopeptide glycoforms increased progressively from 5 to 20 min and then started decreasing due to analyte breakthrough (see Supplementary Fig. 2). Despite the increased sensitivity, loading the sample for 20 min caused current instabilities that affected the

separation repeatability. Therefore, a loading time of 10 min was selected for the optimized method.

### 3.3. Analysis of rhEPO glycopeptides by $\text{TiO}_2\text{-SPE-CE-MS}$

The optimized method was repeatable in terms of migration times and peak areas. The %RSDs ( $n = 3$ ) were between 9 and 11% and between 6 and 11% for  $O_{126}$  glycoforms, respectively, in consecutive analysis of a  $10 \text{ mg L}^{-1}$  rhEPO digest. With respect to the micro-cartridge lifetime, it worked properly for around 10 analyses. The method was linear ( $R^2 > 0.99$ ) between  $0.5$  and  $50 \text{ mg L}^{-1}$  for  $O_{126}$  glycoforms containing NeuAc and between  $10$  and  $50 \text{ mg L}^{-1}$  for  $O_{126}$  glycoforms with NeuGc (see Fig. 3A-i and 3A-ii, respectively). When loading higher concentrations, the sorbent was saturated, and the expected increase in the peak areas was not observed. The LOD was established at  $0.25 \text{ mg L}^{-1}$  to detect all the  $O_{126}$  glycoforms with NeuAc and at  $10 \text{ mg L}^{-1}$  for those with NeuGc, which corresponds to an improvement of up to 100 times with respect to CE-MS method. The developed method lowers significantly the working concentrations for  $O_{126}$  glycoforms from rhEPO containing both NeuAc and NeuGc. This is an attractive trait, since it enables detection of glycoforms with NeuGc at low concentrations, which is particularly important in quality control of biopharmaceuticals and doping analysis. Fig. 3B shows the EIEs of the  $O_{126}$  glycoforms with NeuGc at  $100 \text{ mg L}^{-1}$  of digested rhEPO.



**Fig. 2.** A) Bar graph showing the effect on the peak areas of rEPO  $O_{126}$ -glycopeptide glycoforms and the total peptides detected by  $TiO_2$ -SPE-CE-MS of the presence in the binding and loading buffers of 0.1 M of lactic acid and of an extra washing step B) EIEs of the most abundant  $O_{126}$ -glycopeptide glycoforms using the established binding and loading buffers with 0.1 M of lactic acid and different eluents: (i) 1 M  $NH_4OH$ , (ii) 3 M  $NH_4OH$  and (iii) 50 mM  $(NH_4)_3PO_4$ , pH 9) ( $10\text{ mg L}^{-1}$  rEPO).

The average peak areas of all detected  $O_{126}$ -glycopeptide glycoforms by  $TiO_2$ -SPE-CE-MS are presented in Table 1, as well as the values for the peptides and the  $N_{83}$  glycopeptide glycoforms. The  $TiO_2$ -SPE-CE-MS method, although developed and optimized for the  $O_{126}$  glycopeptide glycoforms of rEPO enabled also detection of all  $N_{83}$  glycopeptide glycoforms reported by CE-MS (see Table 1). It was found that allowed a sensitivity enhancement of 10 times, detecting at  $100\text{ mg L}^{-1}$  of

digested rEPO all the  $N_{83}$  glycoforms identified at  $1000\text{ mg L}^{-1}$  by CE-MS, without any reoptimization for  $N$ -glycopeptides. The LOD for the  $N_{83}$  glycoforms was higher than for the  $O_{126}$  glycoforms, as expected because these larger and more negatively-charged glycoforms are harder to ionize in positive ionization mode. By way of an example, Fig. 3C shows the EIEs of the most relevant  $N_{83}$  glycoforms obtained by  $TiO_2$ -SPE-CE-MS and previously presented for CE-MS in Fig. 1B-i. As in

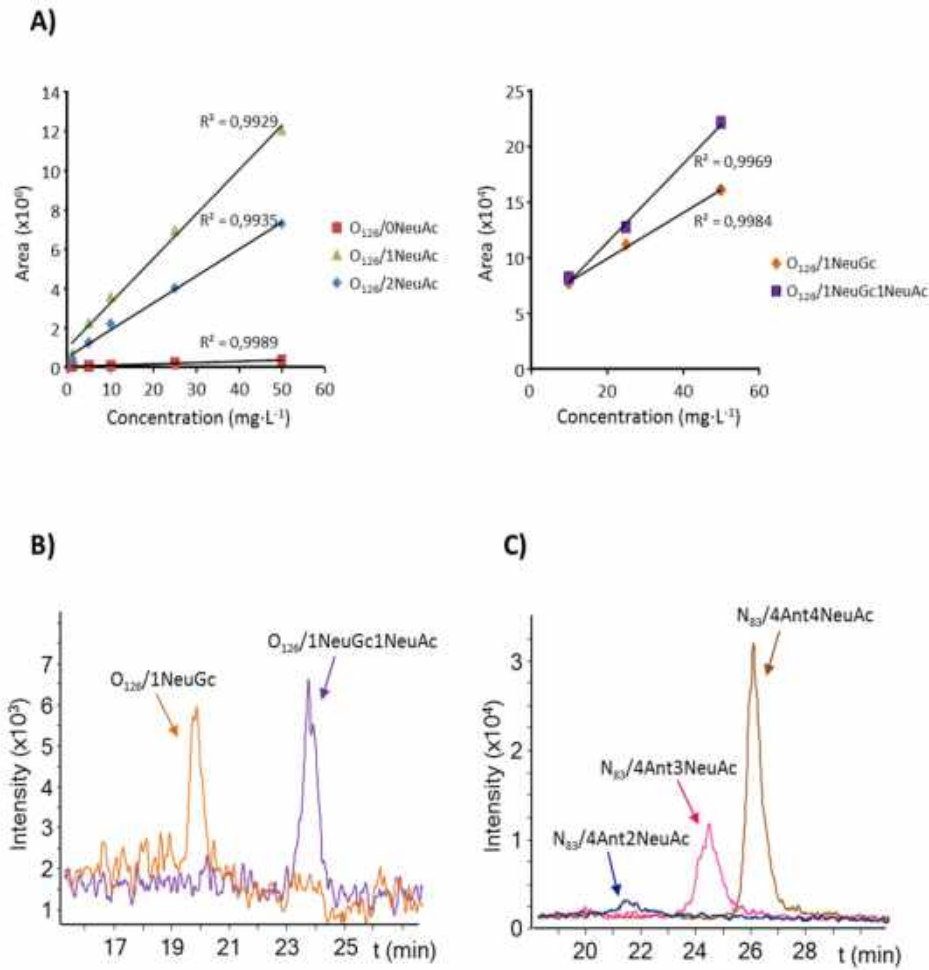


Fig. 3. A) Linearity of the optimized  $\text{TiO}_2$ -SPE-CE-MS method for the rhEPO  $O_{126}$ -glycopeptide glycoforms. B) EIEs of rhEPO  $O_{126}$  glycoforms with NeuGc and C) the most relevant  $N_{83}$  glycoforms by  $\text{TiO}_2$ -SPE-CE-MS under the optimized conditions ( $100 \text{ mg L}^{-1}$  rhEPO).

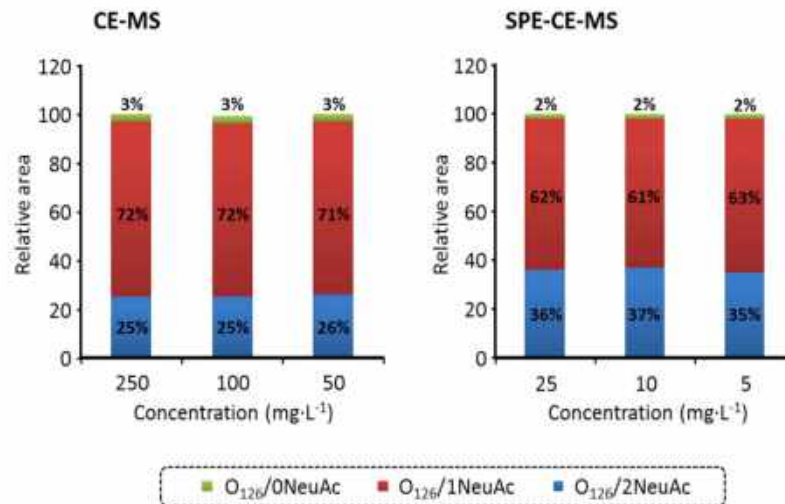


Fig. 4. Bar graph showing the relative peak areas of the most abundant  $O_{126}$ -glycopeptide glycoforms at different concentrations of digested glycoprotein by CE-MS and  $\text{TiO}_2$ -SPE-CE-MS. (Relative peak area was calculated as the peak area of each glycoform divided by the sum of the peak areas of all glycoforms).

**Table 2**

Summary of peptides and O<sub>74</sub> glycopeptide glycoforms detected by CE-MS and TiO<sub>2</sub>-SPE-CE-MS in the tryptic digest of APO-C3 at a concentration of 600 mg L<sup>-1</sup> and 10 mg L<sup>-1</sup> of digested glycoprotein, respectively (n = 3).

|   | Sequence          | M <sub>theo</sub> | CE-MS                    |                             | SPE-CE-MS                   |
|---|-------------------|-------------------|--------------------------|-----------------------------|-----------------------------|
|   |                   |                   | Error <sup>a</sup> (ppm) | Area(x10 <sup>7</sup> a.u.) | Area(x10 <sup>7</sup> a.u.) |
| Peptides                                | SEAEDASLLSFMQGYMK | 1905.8488         | 9.3                      | 0.03                        | 0.004                       |
|   | HATK              | 455.2492          | 7.4                      | 1                           | 0.5                         |
|   | DALSSVQESQVAQQAR  | 1715.8438         | 2.9                      | 0.8                         | 1                           |
|   | GWVTDGFSSLK       | 1195.5873         | 2.1                      | 4                           | 0.01                        |
|   | DYWSTVK           | 897.4232          | 2.4                      | 2                           | 0.05                        |
| O <sub>74</sub> glycoforms <sup>c</sup> | /0NeuAc           | 2501.1486         | 3.7                      | 0.003 (0.2%) <sup>b</sup>   | 0.005 (0.1%) <sup>b</sup>   |
|   | /1NeuAc           | 2792.2440         | 0.4                      | 0.3 (24.9%) <sup>b</sup>    | 0.7 (12.3%) <sup>b</sup>    |
|   | /2NeuAc           | 3083.3394         | 1.7                      | 0.9 (74.8%) <sup>b</sup>    | 5 (87.6%) <sup>b</sup>      |

Detected peptides with less than 4 amino acids were not considered (i.e. TAK, DK).

<sup>a</sup> Error was calculated in ppm as:  $(|M_{exp} - M_{theo}|/M_{theo}) \times 10^6$  (exp = experimental and theo = theoretical).

<sup>b</sup> Relative peak area was calculated as the peak area of each glycoform divided by the sum of the peak areas of all glycoforms.

<sup>c</sup> O-glycopeptide glycoforms are composed by one N-acetylglucosamine and one galactose monosaccharides plus the indicated number of sialic acids in each case.

**Table 3**

Summary of peptides and N-glycopeptide glycoforms detected by CE-MS and TiO<sub>2</sub>-SPE-CE-MS in the tryptic digest of bAGP at a concentration of 1000 mg L<sup>-1</sup> and 100 mg L<sup>-1</sup> of digested glycoprotein, respectively (n = 3).

|                             | Sequence                                | M <sub>theo</sub> | CE-MS                    |                             | SPE-CE-MS                   |            |
|-----------------------------|---|-------------------|--------------------------|-----------------------------|-----------------------------|------------|
|                             |   |                   | Error <sup>a</sup> (ppm) | Area(x10 <sup>6</sup> a.u.) | Area(x10 <sup>6</sup> a.u.) |            |
| Peptides                    | WFYIGSAFR                               | 1145.5658         | 2.9                      | 3                           | 0.05                        |            |
|                             | AIQAFFYLEPR                             | 1424.7452         | 1.7                      | 3                           | 0.01                        |            |
|                             | HAEDK                                   | 598.2711          | 3.4                      | 2                           | 3                           |            |
|                             | LITR                                    | 501.3275          | 2.6                      | 4                           | 0.2                         |            |
|                             | EYQTIEDK                                | 1024.4713         | 4.1                      | 2                           | 0.2                         |            |
|                             | VESDR                                   | 604.2816          | 2.8                      | 3                           | 0.1                         |            |
|                             | EHPVDLLSK                               | 1199.655          | 5.3                      | 6                           | 0.4                         |            |
|                             | NVGVSYADKPEVTQEQL                       | 2038.0007         | 3.3                      | 5                           | 1                           |            |
|                             | EFLDVIK                                 | 862.48            | 1.7                      | 11                          | 0.02                        |            |
|                             | CIGIQESEIHYTDER-(Cys-IAA <sup>b</sup> ) | 1796.8502         | 6.2                      | 1                           | 0.04                        |            |
|                             | DACGPLEK-(Cys-IAA <sup>b</sup> )        | 888.4011          | 4.6                      | 5                           | 0.3                         |            |
|                             | QHEEER                                  | 826.3569          | 6.5                      | 0.9                         | 0.8                         |            |
|                             | ETEAS                                   | 535.2126          | 7.2                      | 0.8                         | 0.03                        |            |
|                             | N <sub>83</sub> glycoforms              | /2Ant2NeuGc       | 3000.1123                | 2.6                         | 0.28 (70%)                  | 0.27 (53%) |
|                             |   | /2Ant3NeuGc       | 3307.2026                | 4.3                         | 0.12 (30%)                  | 0.24 (47%) |
| N <sub>70</sub> glycoforms  | /2Ant2NeuAc                             | 3394.2984         | 8.4                      | 0.14 (28%)                  | 0.12 (31%)                  |            |
|                             | /2Ant2NeuGc                             | 3426.2882         | 4.0                      | 0.15 (30%)                  | 0.14 (36%)                  |            |
| N <sub>86</sub> glycoforms  | /2Ant1NeuAc1NeuGc                       | 3410.2933         | 1.1                      | 0.21 (42%)                  | 0.13 (33%)                  |            |
|                             | /2Ant2NeuAc                             | 2951.1647         | 8.8                      | 0.14 (21%)                  | 0.20 (27%)                  |            |
| N <sub>111</sub> glycoforms | /2Ant2NeuGc                             | 2983.1545         | 4.7                      | 0.25 (37%)                  | 0.26 (36%)                  |            |
|                             | /2Ant1NeuAc1NeuGc                       | 2967.1596         | 4.6                      | 0.28 (42%)                  | 0.27 (37%)                  |            |
|                             | /2Ant2NeuAc                             | 3530.4162         | 2.8                      | 0.32 (25%)                  | 0.10 (15%)                  |            |
|                             | /2Ant2NeuGc                             | 3562.4060         | 1.1                      | 0.39 (31%)                  | 0.16 (24%)                  |            |
|                             | /2Ant1NeuAc1NeuGc                       | 3546.4111         | 1.8                      | 0.46 (36%)                  | 0.18 (27%)                  |            |
|                             | /2Ant3NeuGc                             | 3869.4963         | 8.2                      | 0.10 (8%)                   | 0.23 (34%)                  |            |

Detected peptides with less than 4 amino acids were not considered (i.e. SAR, IVR, HFR, K).

<sup>a</sup> Error was calculated in ppm as:  $(|M_{exp} - M_{theo}|/M_{theo}) \times 10^6$  (exp = experimental and theo = theoretical).

<sup>b</sup> IAA stands for iodoacetamide.

<sup>c</sup> Relative peak area was calculated as the peak area of each glycoform divided by the sum of the peak areas of all the glycoforms of a certain glycosylation site.

CE-MS, N<sub>83</sub> glycoforms migrated close and after the O<sub>126</sub> glycoforms (compare Fig. 2B-i and 3C). Moreover, separation between glycoforms containing different number of sialic acids was not compromised with or without on-line preconcentration.

Another important concern was to determine if the optimized method was providing an accurate and reliable glycopeptide glycoform fingerprint of rHEPO. In order to study if the TiO<sub>2</sub> sorbent preferentially retained the glycoforms containing more sialic acids, rHEPO digests were analyzed at different concentrations by CE-MS and TiO<sub>2</sub>-SPE-CE-MS. Fig. 4 shows the bar graphs for the peak areas of each O<sub>126</sub> glycoform (i.e. O<sub>126</sub>/0NeuAc, O<sub>126</sub>/1NeuAc or O<sub>126</sub>/2NeuAc) relative to the sum of all O<sub>126</sub> glycoforms. As can be observed, in all the studied concentrations, the relative peak area of O<sub>126</sub>/2NeuAc using TiO<sub>2</sub>-SPE-CE-MS was slightly higher (around 36%) than by CE-MS (around 25%).

Consequently, the relative peak areas of the glycoforms containing less sialic acids slightly decreased compared to CE-MS. This proved that the TiO<sub>2</sub> sorbent shows certain preference for more sialylated glycoforms. In the case of N<sub>83</sub> glycopeptide, using as a case study the N<sub>83</sub> tetra-antennary glycoforms with two, three and four sialic acids, no significant differences were found between CE-MS and TiO<sub>2</sub>-SPE-CE-MS (relative peak areas were around 5%, 30% and 65% for N<sub>83</sub>/4Ant2NeuAc, N<sub>83</sub>/4Ant3NeuAc and N<sub>83</sub>/4Ant4NeuAc, respectively, in both cases). Therefore, it can be tentatively assumed that the slightly preference of the TiO<sub>2</sub> sorbent for the more sialylated structures in rHEPO was observed because O<sub>126</sub> glycoforms have a small carbohydrate moiety, which is basically composed by sialic acid residues.

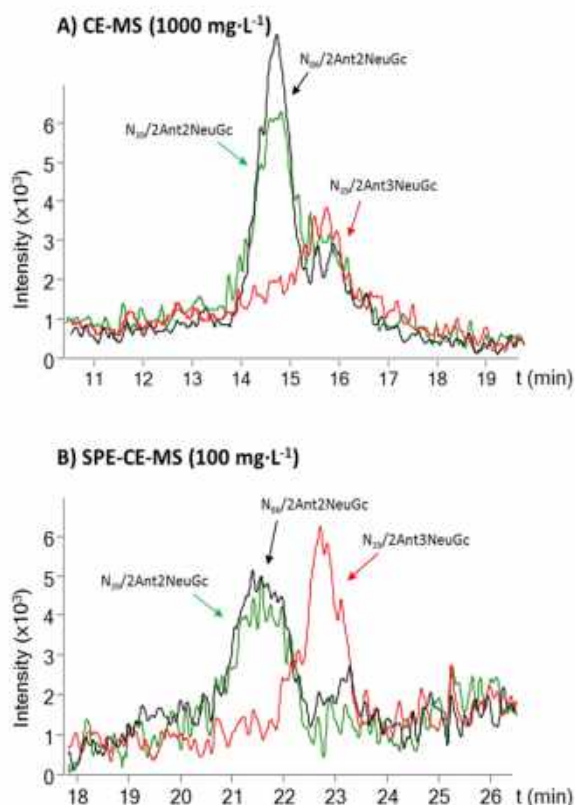


Fig. 5. EIEs of bAGP  $N_{39}/2\text{Ant}2\text{NeuGc}$ ,  $N_{39}/2\text{Ant}3\text{NeuGc}$  and  $N_{86}/2\text{Ant}2\text{NeuGc}$  glycopeptide glycoforms by (A) CE-MS ( $1000 \text{ mg L}^{-1}$  bAGP) and (B)  $\text{TiO}_2$ -SPE-CE-MS ( $100 \text{ mg L}^{-1}$  bAGP).

### 3.4. Analysis of other model glycoproteins

In order to test the potential of the  $\text{TiO}_2$ -SPE-CE-MS method developed for rhEPO with substantially different glycopeptides, the tryptic digests of APO-C3 and bAGP were also analyzed by CE-MS and  $\text{TiO}_2$ -SPE-CE-MS. APO-C3 is a small glycoprotein containing 70 amino acids with an O-glycosite at position 74, while bAGP is a 42 kDa glycoprotein containing 5 N-glycosylation sites at positions 16, 39, 76, 86 and 118.

Regarding APO-C3, the tryptic digests were analyzed at  $600 \text{ mg L}^{-1}$  and  $10 \text{ mg L}^{-1}$  of digested glycoprotein by CE-MS and  $\text{TiO}_2$ -SPE-CE-MS, respectively. Results of detected peptides and  $O_{74}$  glycoforms in both cases, including the mass error and peak areas, are listed in Table 2. At these concentrations, the same peptides and glycopeptides were detected using both methods. Even though the analyzed concentration by  $\text{TiO}_2$ -SPE-CE-MS was 60 times lower than by CE-MS, the peak areas obtained for  $O_{74}$  glycoforms were higher compared to those obtained by CE-MS. At the same time, the opposite was in general observed for peptides (Table 2), proving that the  $\text{TiO}_2$  sorbent selectively retained the glycopeptides. As happened with rhEPO digests, the glycoform without sialic acids ( $O_{74}/0\text{NeuAc}$ ) was also detected by  $\text{TiO}_2$ -SPE-CE-MS. Similarly, comparison between the relative peak areas obtained for  $O_{74}$  glycoforms by CE-MS and SPE-CE-MS (see Table 2) revealed again certain preference of the  $\text{TiO}_2$  sorbent for glycoforms containing more sialic acids. Thus, the relative peak area of the  $O_{74}/2\text{NeuAc}$  glycoform by  $\text{TiO}_2$ -SPE-CE-MS was higher (around 88%) than by CE-MS (around 75%), whereas the relative peak areas of the glycoforms with less sialic acids were similar ( $O_{74}/0\text{NeuAc}$ ) or lower ( $O_{74}/1\text{NeuAc}$ ).

With regard to bAGP, the tryptic digests were analyzed at

$1000 \text{ mg L}^{-1}$  and  $100 \text{ mg L}^{-1}$  of digested glycoprotein by CE-MS and  $\text{TiO}_2$ -SPE-CE-MS, respectively. Table 3 lists all the detected peptides and glycopeptide glycoforms with both techniques, including the mass error and peak areas. Although at ten times lower concentration of digested bAGP by  $\text{TiO}_2$ -SPE-CE-MS, the same peptides and glycopeptides were detected as by CE-MS. In addition, selective retention of glycopeptides was obtained by  $\text{TiO}_2$ -SPE-CE-MS. No glycoforms from the  $N_{116}$ -glycopeptide could be detected probably because the peptide size of this glycopeptide is about two to three times larger than for the rest of glycopeptides, making it harder to ionize. Fig. 5 shows as an example the EIEs of some  $N_{39}$  and  $N_{86}$  biantennary glycoforms obtained by CE-MS and  $\text{TiO}_2$ -SPE-CE-MS. Regarding migration time, no separation between  $N_{39}$ ,  $N_{76}$ ,  $N_{86}$  and  $N_{118}$  glycopeptide glycoforms with two sialic acids was obtained (see Fig. 5 for some  $N_{39}$  and  $N_{86}$  glycoforms). This is to be expected since all these glycoforms contain two negative charges and share very similar molecular masses. On the other hand, separation between  $N_{39}$  glycoforms with two and three sialic acids was obtained using both CE-MS and  $\text{TiO}_2$ -SPE-CE-MS methods, as shown in Fig. 5. Furthermore, a higher affinity of the  $\text{TiO}_2$  sorbent for glycans with more sialic acids was also revealed in this case. As can be observed in Fig. 5 and Table 3, the relative peak area of the  $N_{39}/2\text{Ant}3\text{NeuGc}$  and  $N_{118}/2\text{Ant}3\text{NeuGc}$  was higher by  $\text{TiO}_2$ -SPE-CE-MS than by CE-MS. These results indicate that, as with the O-glycopeptides, the  $\text{TiO}_2$  sorbent presents certain preference for more sialylated structures in N-glycopeptides probably because the carbohydrate moiety is smaller (biantennary N-glycopeptide glycoforms) than in rhEPO (tri and tetra-antennary N-glycopeptide glycoforms).

## 4. Conclusions

An on-line  $\text{TiO}_2$ -SPE-CE-MS method was optimized to selectively retain and enrich glycopeptides obtained from glycoprotein digests. Conditions for the  $\text{TiO}_2$ -SPE-CE-MS analysis of rhEPO  $O_{126}$ -glycopeptide glycoforms were carefully fine-tuned to maximize detection sensitivity and reduce non-specific retention of peptides, without compromising separation between glycoforms. Under the optimized conditions, repeatability was good (9–11 and 6–11% RSD for migration times and peak areas) and the microcartridge lifetime was around 10 analyses. The method was linear and the LODs were up to 100 times lower than by CE-MS. It was also demonstrated that the optimized method allowed preconcentrating the rhEPO  $N_{83}$ -glycopeptide glycoforms, as well as APO-C3 O-glycopeptides and bAGP N-glycopeptides. Therefore, the established  $\text{TiO}_2$ -SPE-CE-MS method can be used to analyze glycopeptides with different compositions and regardless of the protein derived from. It has to be taken into account the higher affinity of the  $\text{TiO}_2$  sorbent for more sialylated glycoforms, which would decrease with the size of the carbohydrate moiety. Results are promising with immediate application in the analysis of diluted protein digests of biopharmaceuticals or in comparative glycoproteomic studies.

## Declaration of competing interest

The authors have declared no conflict of interest.

## Acknowledgments

This work was supported by a grant from the Spanish Ministry of Economy and Competitiveness (RTI2018-097411-B-I00) and the Cathedra UB Rector Francisco Buscarons Ubeda (Forensic Chemistry and Chemical Engineering). Montserrat Mancera-Arteu acknowledges the University of Barcelona for an ADR fellowship.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.talanta.2019.120563>.

## References

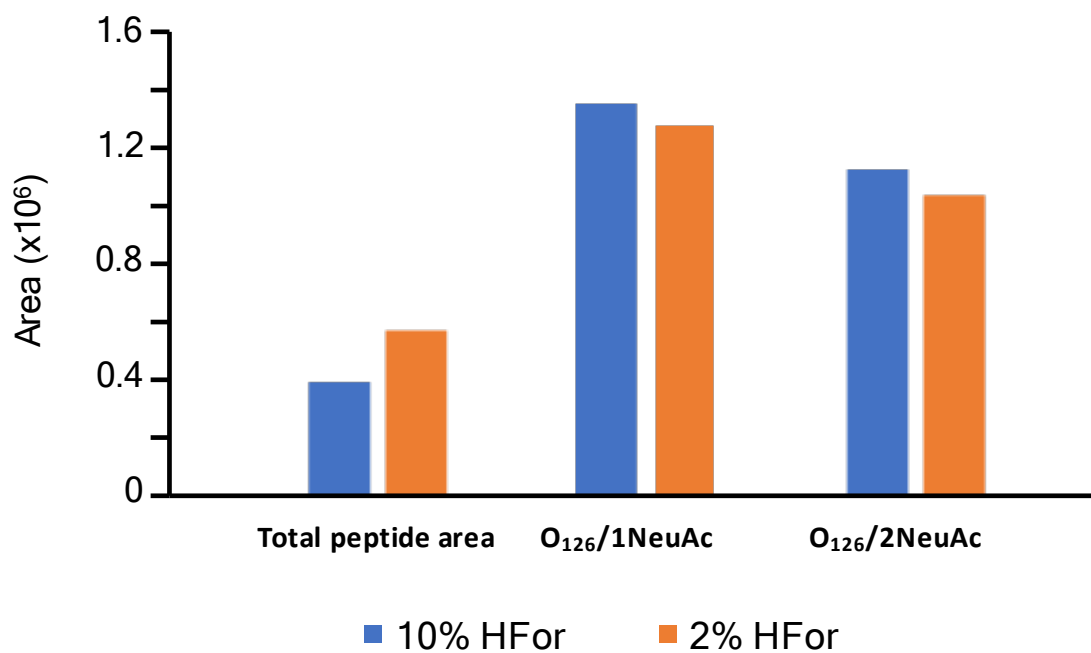
- [1] A. Varki, S. Kornfeld, Chapter 1. Historical background and overview, in: A. Varki, R.D. Cummings, J.D. Esko, H.H. Freeze, P. Stanley, C.R. Bertozzi, G.W. Hart, M.E. Etzler (Eds.), *Essentials Glycobiol.* third ed., Cold Spring Harbor Laboratory Press, New York, NY, 2017 Cold Spring Harbor (NY).
- [2] R.D. Cummings, Stuck on sugars – how carbohydrates regulate cell adhesion, recognition, and signaling, *Glycoconj. J* (2019), <https://doi.org/10.1007/s10719-019-09876-0>.
- [3] S. Carillo, S. Mittermayr, A. Farrell, S. Albrecht, J. Bones, Glycosylation analysis of therapeutic glycoproteins produced in CHO cells, in: P. Meleady (Ed.), *Heterologous Protein Prod. CHO Cells*. Methods Mol. Biol. first ed., Humana Press, New York, NY, 2017.
- [4] A. Planinc, J. Bones, B. Dejaeger, P. Van Antwerpen, C. Delporte, Glycan characterization of biopharmaceuticals: updates and perspectives, *Anal. Chim. Acta* 921 (2016) 13–27, <https://doi.org/10.1016/j.aca.2016.03.049>.
- [5] M.J. Kallemla, D. Park, B.L. Carlito, Glycans and glycoproteins as specific biomarkers for cancer, *Anal. Bioanal. Chem.* 409 (2017) 395–410, <https://doi.org/10.1007/s00216-016-9880-6>.
- [6] C. Reily, T.J. Stewart, M.B. Renfrow, J. Novak, Glycosylation in health and disease, *Nat. Rev. Nephrol.* 15 (2019) 346–366, <https://doi.org/10.1038/s41581-019-0129-4>.
- [7] A. Barroso, E. Giménez, F. Benavente, J. Barbosa, V. Sanz-Nebot, Analysis of human transferrin glycopeptides by capillary electrophoresis and capillary liquid chromatography-mass spectrometry. Application to diagnosis of alcohol dependence, *Anal. Chim. Acta* 804 (2013) 167–175, <https://doi.org/10.1016/j.aca.2013.09.044>.
- [8] 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, <https://doi.org/10.1016/j.talanta.2016.07.055>.
- [9] J. Cody, E. Hodson, Recombinant human erythropoietin versus placebo or no treatment for the anaemia of chronic kidney disease in people not requiring dialysis, *Cochrane Database Syst. Rev* (2016) 1465–1858, <https://doi.org/10.1002/14651858.CD003266.pub3>.
- [10] J. Glaspy, Current status of use of erythropoietic agents in cancer patients, *Semin. Thromb. Hemost.* 40 (2014) 306–312, <https://doi.org/10.1055/s-0034-1370768>.
- [11] F. Laine, J. de Caesteir, Recombinant erythropoietin in urine, *Nature* 405 (2000) 635–637.
- [12] O. Salamin, T. Kuuranne, M. Saugy, N. Leuenberger, Erythropoietin as a performance-enhancing drug: its mechanistic basis, detection, and potential adverse effects, *Mol. Cell. Endocrinol.* 464 (2018) 75–87, <https://doi.org/10.1016/j.mce.2017.03.033>.
- [13] D. Ghaderi, R.E. Taylor, V. Padler-karavani, S. Diaz, A. Varki, Implications of the presence of N-glycolylneuraminic acid in recombinant therapeutic glycoproteins, *Nat. Biotechnol.* 28 (2011) 863–867, <https://doi.org/10.1038/nbt.1651>.
- [14] C. Neusiß, U. Demelbauer, M. Pelzing, Glycoform characterization of intact erythropoietin by capillary electrophoresis-electrospray-time of flight-mass spectrometry, *Electrophoresis* 26 (2005) 1442–1450, <https://doi.org/10.1002/elps.200410269>.
- [15] E. Giménez, F. Benavente, J. Barbosa, V. Sanz-Nebot, Analysis of intact erythropoietin and novel erythropoiesis-stimulating protein by capillary electrophoresis-electrospray- ion trap mass spectrometry, *Electrophoresis* 29 (2008) 2161–2170, <https://doi.org/10.1002/elps.200700788>.
- [16] Z. Szabo, A. Guttman, T. Rejtár, B.L. Karger, Improved sample preparation method for glycan analysis of glycoproteins by CE-LIF and CE-MS, *Electrophoresis* 31 (2010) 1389–1395, <https://doi.org/10.1002/elps.201000037>.
- [17] A.D. Zamfir, C. Flangea, A. Serb, A.-M. Zagrean, A.M. Rizzi, E. Sisu, Separation and identification of glycoforms by capillary electrophoresis with electrospray ionization mass spectrometric detection, in: J.J. Kohler, S.M. Patrie (Eds.), *Mass Spectrom. Glycoproteins*. Methods Mol. Biol. Methods Protoc. Humana Press, Totowa, NJ, 2013, [https://doi.org/10.1007/978-1-62703-146-2\\_11](https://doi.org/10.1007/978-1-62703-146-2_11).
- [18] E. Giménez, R. Ramos-Hernán, 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, <https://doi.org/10.1016/j.aca.2011.10.028>.
- [19] F. Benavente, S. Medina-Casunellas, E. Giménez, V. Sanz-Nebot, On-line solid-phase extraction capillary electrophoresis mass spectrometry for preconcentration and clean-up of peptides and proteins, in: N. Tran, M. Taverna (Eds.), *Capill. Electrophor. Proteins Pept. Methods Mol. Biol.* Humana Press, New York, NY, 2016, [https://doi.org/10.1007/978-1-4939-4014-1\\_6](https://doi.org/10.1007/978-1-4939-4014-1_6).
- [20] L. Pont, R. Pero-Gascon, E. Gimenez, V. Sanz-Nebot, F. Benavente, A critical retrospective and prospective review of designs and materials in on-line solid-phase extraction capillary electrophoresis, *Anal. Chim. Acta* (2019), <https://doi.org/10.1016/j.aca.2019.05.022>.
- [21] R. Ramautar, G.W. Somsen, G.J. de Jong, Developments in coupled solid-phase extraction-capillary electrophoresis 2013–2015, *Electrophoresis* 37 (2016) 35–44, <https://doi.org/10.1002/elps.201500401>.
- [22] A. Šlampová, Z. Malá, P. Gebauer, Recent progress of sample stacking in capillary electrophoresis (2016–2018), *Electrophoresis* 40 (2019) 40–54, <https://doi.org/10.1002/elps.201800261>.
- [23] M.C. Breadmore, W. Grochocli, U. Kalsoom, M.N. Alves, S.C. Phung, M.T. Rokh, J.M. Gabot, A. Ghiasvand, F. Li, A.I. Shalun, A.S.A. Keyon, A.A. Alhusban, H.H. See, A. Wuehrlich, M. Dawod, J.P. Quirino, Recent advances in enhancing the sensitivity of electrophoresis and electrochromatography in capillaries in microchips (2016–2018), *Electrophoresis* 40 (2019) 17–39, <https://doi.org/10.1002/elps.201800384>.
- [24] T.E. Thingholm, M.R. Larsen, Chapter 9. The use of titanium dioxide for selective enrichment of phosphorylated peptides, in: L. von Stechow (Ed.), *Phospho-Proteomics. Methods Mol. Biol.* first ed., Springer, New York, NY, 2016, <https://doi.org/10.1007/978-1-4939-3049-4>.
- [25] M. Mazanek, G. Mitaloviae, F. Herzog, C. Stingl, J.R. Hutchins, J.M. Peters, K. Mechtler, Titanium dioxide as a chemo-affinity solid phase in offline phosphopeptide chromatography prior to HPLC-MS/MS analysis, *Nat. Protoc.* 2 (2007) 1059–1069, <https://doi.org/10.1038/nprot.2006.280>.
- [26] A.L.K. Eriksson, K. Edwards, A. Hagfeldt, V.A. Hernández, Physicochemical characterization of phosphopeptide/titanium dioxide interactions employing the quartz crystal microbalance technique, *J. Phys. Chem. B* 117 (2013) 2019–2025, <https://doi.org/10.1021/jp310161m>.
- [27] G. Palmisano, S.E. Lendal, K. Engholm-Keller, R. Leth-Larsen, B.L. Parker, M.R. Larsen, Selective enrichment of sialic acid-containing glycopeptides using titanium dioxide chromatography with analysis by HILIC and mass spectrometry, *Nat. Protoc.* 5 (2010) 1974–1982, <https://doi.org/10.1038/nprot.2010.167>.
- [28] Q. Sheng, X. Li, W. Yin, L. Yu, Y. Ke, X. Liang, Retention mechanism and enrichment of glycopeptides on titanium dioxide, *Anal. Methods* 5 (2013) 7072–7080, <https://doi.org/10.1039/c3ay41294f>.
- [29] C. Zhang, Z. Ye, P. Xue, Q. Shu, Y. Zhou, Y. Ji, Y. Fu, J. Wang, F. Yang, Evaluation of different N-glycopeptide enrichment methods for N-glycosylation sites mapping in mouse brain, *J. Proteome Res.* 15 (2016) 2960–2968, <https://doi.org/10.1021/acs.jproteome.5b00098>.
- [30] R. Pero-Gascon, V. Sanz-Nebot, M. V. Berezovski, F. Benavente, Analysis of circulating microRNAs and their post-transcriptional modifications in cancer serum by on-line solid-phase extraction-capillary electrophoresis-mass spectrometry, *Anal. Chem.* 90 (2018) 6618–6625, <https://doi.org/10.1021/acs.analchem.8b00405>.
- [31] H.H. Lauer, G.P. Roring, *High Performance Capillary Electrophoresis*, second ed., Agilent Technologies, Waldbronn, Germany, 2014, pp. 60–61.
- [32] E. Giménez, R. Ramos-Hernán, F. Benavente, J. Barbosa, V. Sanz-Nebot, Capillary electrophoresis time of flight mass spectrometry for a confident elucidation of a glycopeptide map of recombinant human erythropoietin, *Rapid Commun. Mass Spectrom.* 25 (2011) 2307–2316, <https://doi.org/10.1002/rcm.5114>.
- [33] T.E. Thingholm, M.R. Larsen, The use of Titanium Dioxide for selective enrichment of phosphorylated peptides, *Methods Mol. Biol.* 1355 (2016) 135–146, [https://doi.org/10.1007/978-1-4939-3049-4\\_9](https://doi.org/10.1007/978-1-4939-3049-4_9).
- [34] L. Yu, Z. Zhu, K.C. Chan, H.J. Issaq, D.S. Dimitrov, T.D. Veenstra, Improved titanium dioxide enrichment of phosphopeptides from HeLa cells and high confident phosphopeptide identification by cross-validation of MS/MS and MS/MS/MS spectra, *J. Proteome Res.* 6 (2007) 4150–4162, <https://doi.org/10.1021/pr070152u>.
- [35] W. Cao, J. Cao, J. Huang, L. Zhang, J. Yao, H. Xu, P. Yang, Enhanced N-glycosylation site exploitation of sialoglycopeptides by peptide IPG-IEF assisted TiO<sub>2</sub> chromatography, *Glycoconj. J.* 29 (2012) 433–443, <https://doi.org/10.1007/s10719-012-9404-3>.

## Supplementary Figures

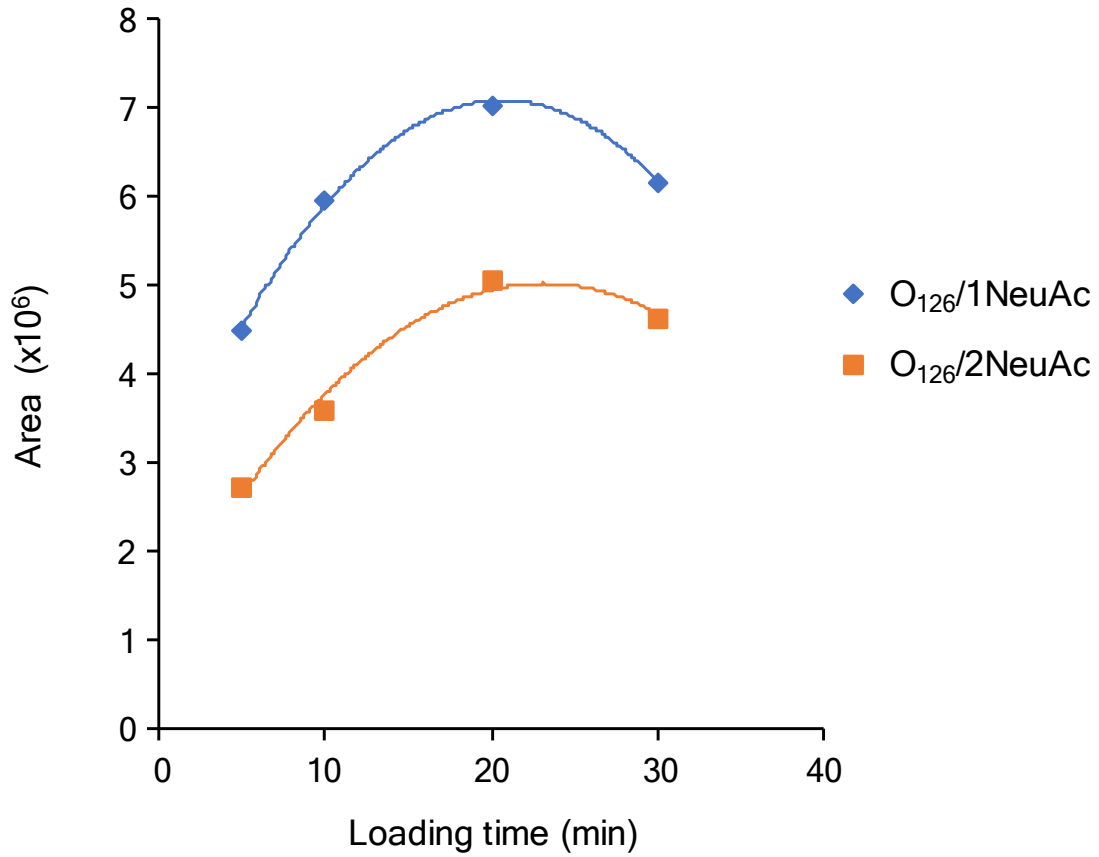
---







**Supplementary Figure S-1.** Bar graph showing the effect on the peak areas of rhEPO O<sub>126</sub>/1NeuAc and O<sub>126</sub>/2NeuAc glycopeptide glycoforms and the total peptides detected by TiO<sub>2</sub>-SPE-CE-MS washing with binding and loading buffers (which contain 10% and 2% of HFor, respectively) or only with loading buffer (50 mg·L<sup>-1</sup> rhEPO. The rest of conditions were the preliminary ones).



**Supplementary Figure S-2.** Evaluation of sample loading time for O<sub>126</sub>/1NeuAc and O<sub>126</sub>/2NeuAc glycopeptide glycoforms of rhEPO by TiO<sub>2</sub>-SPE-CE-MS (50 mg·L<sup>-1</sup> rhEPO. The rest of conditions are the optimized ones).

## **Chapter 3. Glycan isomer characterization**

---



Alterations in the glycoprofile of several proteins have been described in several diseases. Typical glycan modifications are associated with changes in sialylation, fucosylation and/or branching. In the particular case of cancer, increased levels of  $\alpha$ 2-6 sialic acids and overexpression of certain Sialyl Lewis epitopes were mostly described. As sialic acids and fucoses can show different linkage-types in glycans, which lead to the presence of several isomers, the separation and in-depth characterization of these glycan isomers has raised great importance in the last years. The possibility of identifying alterations in specific sialic acid or fucose linkage isomers could facilitate the discovery of diagnostic biomarkers as certain isomers could be associated with a specific pathology.

LC-MS in combination with exoglycosidase digestions has been used to elucidate and structurally characterize isomeric glycans. However, the reported strategies do not present enough retention time reproducibility and the difference in the ionization yield between samples can lead to errors in the assignation of the isomers. Additionally, commercially available enzymes can only cleave certain linkage-types and, in some cases, their specificity might not be absolutely guaranteed. For this reason, methods using tandem mass spectrometry (MS/MS) has been also developed to obtain structural information about the sequence and linkage-type of the glycans. Nevertheless, although several ions have been described as diagnostic of specific structural features, such as sialic acids or fucoses linkage-types, scarce studies previously characterize the linkage-types present in each isomer with a complementary technique to confirm or discard the diagnostic value of such ions.

In this chapter, a method for the separation and identification of glycans and their corresponding isomers by capillary LC-MS using a zwitterionic hydrophilic interaction column (CapZIC-HILIC-MS) is optimized. In addition, two methods for the characterization of glycan isomer structures as well as their sialic acid and fucose linkage-types by CapZIC-HILIC-MS are also presented. The first one uses exoglycosidase digestion in combination with glycan relative

isotope labeling (GRIL) strategy using [ $^{12}\text{C}_6$ ]/[ $^{13}\text{C}_6$ ]-aniline, while the other is focused on the development of a CapZIC-HILIC-MS/MS approach. Human alpha-1-acid glycoprotein (hAGP) is used in both studies as model glycoprotein due to its wide range of sialofucosylated glycan structures and relation with certain inflammatory processes and types of cancer.

The work performed in the above mentioned topics has resulted in the publications listed below:

- **Publication 3.1.-** Identification and characterization of isomeric N-glycans of human alfa-acid-glycoprotein by stable isotope labelling and ZIC-HILIC-MS in combination with exoglycosidase digestion. Mancera-Arteu, M., Giménez, E., Barbosa, J., Sanz-Nebot, V. *Analytica Chimica Acta* (2016), 940, 92-103.
- **Publication 3.2.-** Zwitterionic-hydrophilic interaction capillary liquid chromatography coupled to tandem mass spectrometry for the characterization of human alpha-acid-glycoprotein N-glycan isomers. Mancera-Arteu, M., Giménez, E., Barbosa, J., Peracaula, R., Sanz-Nebot, V. *Analytica Chimica Acta* (2017), 991, 76-88.



Contents lists available at ScienceDirect

Analytica Chimica Acta

journal homepage: [www.elsevier.com/locate/aca](http://www.elsevier.com/locate/aca)

# Identification and characterization of isomeric *N*-glycans of human alfa-acid-glycoprotein by stable isotope labelling and ZIC-HILIC-MS in combination with exoglycosidase digestion



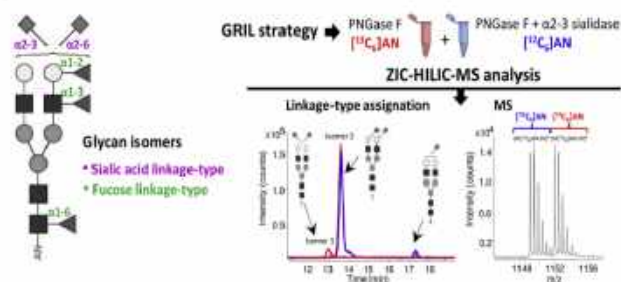
Montserrat Mancera-Arteu, Estela Giménez\*, José Barbosa, Victòria Sanz-Nebot

Department of Analytical Chemistry, University of Barcelona, Diagonal 645, E-08028 Barcelona, Spain

## HIGHLIGHTS

- Enhanced sensitivity is obtained in the detection of glycan isomers.
- GRIL strategy with exoglycosidase digestion reliably characterize glycan isomers.
- Sialic acids and fucoses linkage-type were assigned in hAGP glycan isomers.

## GRAPHICAL ABSTRACT



## ARTICLE INFO

### Article history:

Received 8 June 2016  
 Received in revised form  
 27 July 2016  
 Accepted 29 July 2016  
 Available online 9 August 2016

### Keywords:

Glycan  
 Isomer  
 Aniline  
 hAGP  
 ZIC-HILIC  
 Mass spectrometry  
 Enzymes

## ABSTRACT

In this study, a ZIC-HILIC-MS methodology for the analysis of *N*-glycan isomers was optimized to obtain greater detection sensitivity and thus identify more glycan structures in hAGP. In a second step, this method was combined with glycan reductive isotope labelling (GRIL) through [<sup>12</sup>C<sub>6</sub>]/[<sup>13</sup>C<sub>6</sub>]-aniline and exoglycosidase digestion to characterize the different glycan isomers. The GRIL method allows the peak areas resulting from two different labelled samples to be compared, since neither retention time shifts nor variations in the ionization of glycans between these samples are obtained. First, sialic acid linkage assignments were performed for most hAGP glycan isomers with α-2-3 sialidase digestion. Bi-, tri- and tetraantennary glycan isomers with different terminal sialic acid linkages to galactose (α-2-3 or α-2-6) were assigned, and the potential of this technique for the structural characterization of isobaric isomers was therefore demonstrated. Furthermore, fucose linkage isomers of hAGP glycans were also characterized using this isotope-labelling approach in combination with α-1-3,4 fucosidase and β-1-4 galactosidase digestion. α-1-3 antennary fucoses and α-1-6 core fucosylation were detected in hAGP fucosylated glycans. These established methodologies can be extremely useful for patho-glycomic studies to characterize glycoproteins of biomedical interest and find novel glycan isomers that could be used as biomarkers in cancer research.

© 2016 Elsevier B.V. All rights reserved.

\* Corresponding author.  
 E-mail address: [estelagimenez@ub.edu](mailto:estelagimenez@ub.edu) (E. Giménez).

## 1. Introduction

Glycoproteins play an important role in major biological processes, which are modulated and controlled by the oligosaccharides attached to these proteins [1–4]. These oligosaccharide structures



(also called glycans) are involved in many diseases such as hereditary disorders, immune deficiencies, chronic inflammation and cancer [5,6]. Hence, a wide range of research activities to find glycan-based biomarkers for the detection of different pathologies have been conducted [7–10]. In the case of cancer, typical *N*-glycan modifications in tumours are associated with changes in sialylation, fucosylation or branching. Increased levels of  $\alpha$ 2-6 sialic acids and overexpression of Sialyl Lewis<sup>X</sup> (SLe<sup>X</sup>) and Sialyl Lewis<sup>A</sup> (SLe<sup>A</sup>) epitopes were reported in carcinomas [11–14].

The *N*-glycans of a glycoprotein may present several isomers that mainly differ in the linkage type of the terminal sialic acids or the fucoses. *N*-acetylneuraminic acids (NeuAc) can be  $\alpha$ 2-3 or  $\alpha$ 2-6 linked to galactose (Gal). Moreover, fucoses can be  $\alpha$ 1-3 and  $\alpha$ 1-4 linked to *N*-acetylglucosamine (GlcNAc), on an external branch (Fuc Antenna),  $\alpha$ 1-2 linked to Gal or  $\alpha$ 1-6 linked to the core GlcNAc (Fuc Core) [15–17]. Some studies in recent years have shown the importance of correctly identifying and characterizing these isomers, since one type of isomer can be associated with a specific pathology. The possibility of quantitatively analysing alterations in specific sialic acid or fucose linkage isomers could therefore facilitate the discovery of diagnostic biomarkers rather than simply analysing the total sialylation or fucosylation levels of the glycans [16–19]. Some authors have recently proposed the analysis of *N*-linked glycans by liquid chromatography coupled to mass spectrometry detection (LC-MS) in combination with exoglycosidase digestion to elucidate and structurally characterize isomeric glycans [6,16,20–27]. This characterization is carried out by observing retention time shifts and changes in the peak areas of the isomers after exoglycosidase treatment. However, the strategies reported present certain drawbacks. Some do not present sufficient retention time reproducibility to build a reliable structure/retention time library. Moreover, the differences in ionization yield between

glycans and between samples can lead to errors in the assignment of the isomers, especially when there is an increase in the complexity of the glycans and the number of isomers.

Glycan reductive isotope labelling (GRIL) makes it possible to compare two samples in the same run, correct the lack of retention time reproducibility between runs and avoid variations in the ionization of glycans due to ion-suppression effects or tuning of the mass spectrometer [28–30]. The GRIL methodology allows the glycan fingerprints to be precisely compared before and after digestion with a specific exoglycosidase, and therefore provides an accurate characterization of glycan isomers. In a previous work, a GRIL strategy using [<sup>12</sup>C<sub>6</sub>]/[<sup>13</sup>C<sub>6</sub>] coded aniline was developed for the relative quantification of *N*-glycans, and the labelled glycans were analysed by  $\mu$ ZIC-HILIC-MS [31]. This methodology rigorously separated and quantified the different isomers of a glycan and was recently evaluated for the characterization of human alpha-1-acid-glycoprotein (hAGP) glycans in serum samples from healthy volunteers and patients with pancreatic ductal adenocarcinoma (PDAC) and chronic pancreatitis (ChrP) [32]. In that study, we observed an increase in several fucosylated glycans in PDAC and demonstrated that some of them made it possible to differentiate between the two pathologies. Nevertheless, changes in the proportion of some glycan isomers were also observed in certain PDAC samples [32]. These results found in hAGP samples are a clear example of the need to establish new methodologies to thoroughly characterize isomeric *N*-glycans of glycoproteins.

To that end, an analytical method to improve the detection and characterization of isomeric *N*-glycans by  $\mu$ ZIC-HILIC-MS was developed in this work. Studies using different exoglycosidases in combination with relative quantification by isotope labelling with [<sup>12</sup>C<sub>6</sub>]/[<sup>13</sup>C<sub>6</sub>] aniline have been performed to reliably identify the linkage type of sialic acids and fucoses in the glycan isomers. hAGP

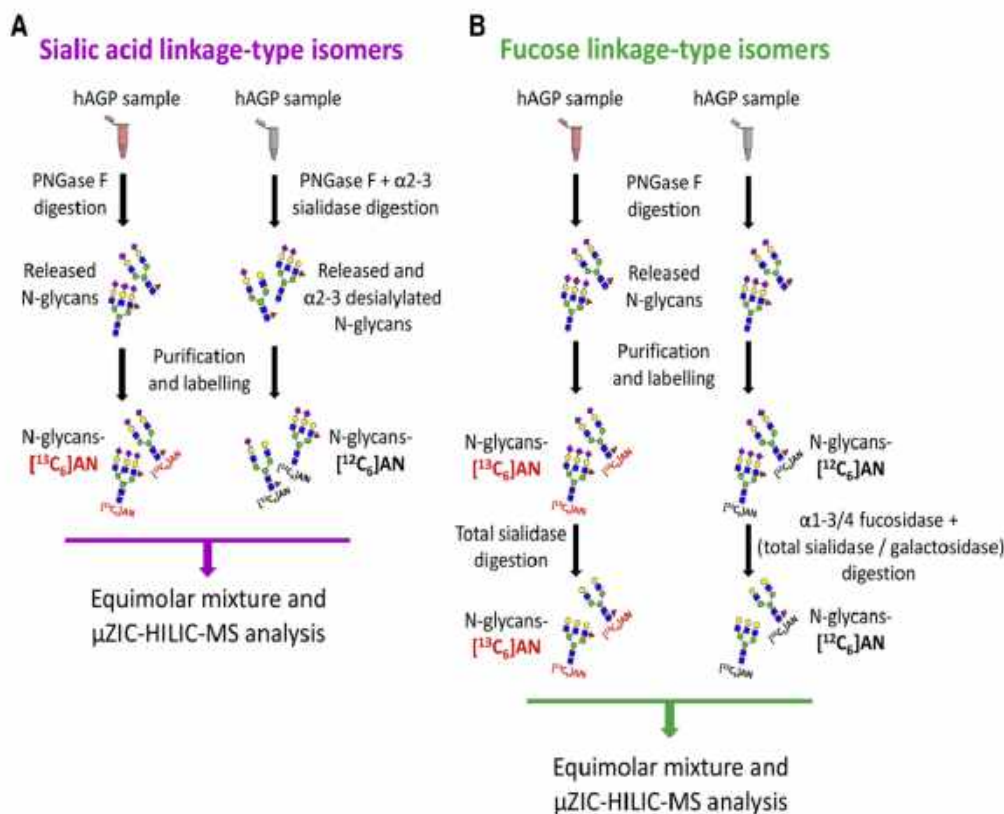


Fig. 1. Experimental workflow followed in this work to characterize sialic acid (A) and fucose linkage-type isomers (B) by  $\mu$ ZIC-HILIC-MS.

has been selected as a glycoprotein model because it appears to show great potential as a biomarker in pancreatic cancer and other diseases [12,32–35], and because it presents a wide range of complex-type *N*-glycans and isomers, thereby allowing the proposed method to be properly evaluated. The goal of these linkage assignments is to establish a robust and reliable methodology for the characterization of isomeric glycans in patho-glycomic studies in order to identify new glycoprotein biomarkers.

## 2. Materials and methods

### 2.1. Chemicals

All chemicals used to prepare buffers and solutions were of analytical reagent grade. Acetic acid (HAc, glacial), formic acid (FA 98%–100%), dimethyl sulfoxide (DMSO) and acetone were supplied by Merck (Darmstadt, Germany). Ammonium hydrogen carbonate ( $\text{NH}_4\text{HCO}_3$ ), sodium phosphate dodecahydrate ( $\text{Na}_3\text{PO}_4 \cdot 12\text{H}_2\text{O}$ ), sodium cyanoborohydride ( $\text{NaBH}_3\text{CN}$ ), [ $^{12}\text{C}_6$ ]-aniline ([ $^{12}\text{C}_6$ ]AN), [ $^{13}\text{C}_6$ ]-aniline ([ $^{13}\text{C}_6$ ]AN), 2-mercaptoethanol ( $\beta$ -ME) and sodium dodecyl sulfate (SDS) were purchased from Sigma-Aldrich (St. Louis, MO, USA). Hydrogen chloride (HCl, 37%) was supplied by Panreac (Barcelona, Spain) and “NP-40 alternative” by Calbiochem (Darmstadt, Germany). Ammonium acetate ( $\text{NH}_4\text{Ac}$ ) and acetonitrile (ACN), from Merck (Madrid, Spain), and LC-MS grade water from Fluka (Madrid, Spain), were used for  $\mu\text{ZIC-HILIC-TOF-MS}$  analysis. Peptide-*N*-glycosidase F (PNGase F) and neuraminidase ( $\alpha$ 2-3,6,8 sialidase) were obtained from Roche Diagnostics (Basel, Switzerland), and  $\alpha$ 2-3 neuraminidase S ( $\alpha$ 2-3 sialidase),  $\alpha$ 1-3,4 fucosidase and  $\beta$ 1-4 galactosidase came from New England Biolabs (Ipswich, MA, USA). ESI low concentration (ESI-L) tuning mix was supplied by Agilent Technologies (Waldbronn, Germany) for the tuning and calibration of the oa-TOF mass spectrometer.

### 2.2. Release, purification and labelling of hAGP *N*-glycans

Standard of human  $\alpha$ 1-acid-glycoprotein (hAGP, 99%) was purchased from Sigma-Aldrich. Twenty-five micrograms of hAGP was reduced with 0.5%  $\beta$ -ME in the presence of 0.5% SDS in 50 mM  $\text{Na}_3\text{PO}_4$  (pH 7.5) and boiled for 30 min in a water bath. When the sample was at room temperature, a volume of 50 mM  $\text{Na}_3\text{PO}_4$  (pH 7.5) with 1% (v/v) of NP-40 alternative was added to achieve a final concentration of 0.1% SDS and  $\beta$ -ME in the sample. To release the *N*-glycans, 1  $\mu\text{L}$  of PNGase F (1 U) solution was added and the mixture was carefully vortexed and incubated at 37 °C for 18 h. Digestion was stopped by heating the sample in a thermoblock at 100 °C for 15 min and the sample was stored at –20 °C until use. The *N*-glycans released were purified by solid phase extraction (SPE) using Hypercarb cartridges (25 mg, 1 mL volume, Thermo Fisher Scientific). First, SPE cartridges were conditioned and equilibrated with 1 mL of 60% ACN, 0.1% FA and 2 mL of water, respectively. After dissolving the digested glycoprotein sample in ~500  $\mu\text{L}$  of water, it was loaded and the SPE cartridge was rinsed with 1 mL of water. Retained *N*-glycans were eluted with 600  $\mu\text{L}$  of 60% ACN and 0.1% FA, and the eluate was evaporated to dryness in a SpeedVac, and the dried *N*-glycans were stored at –20 °C until use. The labelling was carried out using a reaction mixture of 0.35 M aniline and 1 M  $\text{NaCNBH}_3$  in DMSO with 30% HAc. The dried hAGP glycans were mixed with 10  $\mu\text{L}$  of the reaction mixture and incubated at 70 °C for 2 h in a water bath. Samples were cooled to room temperature and the labelled glycans were precipitated with acetone as described in Ref. [32]. Finally, samples were evaporated to dryness in a SpeedVac to remove the excess acetone, and the dried *N*-glycans were stored at –20 °C. Centrifugations were performed in a Mikro 220R centrifuge (Hettich-Zentrifugen, Tuttlingen, Germany).

### 2.3. Exoglycosidase digestion of hAGP *N*-glycans

hAGP *N*-glycans were treated with different exoglycosidases (see the experimental workflow of Fig. 1). **Sialidase digestion:** Sialic acids were released by enzymatic digestion with sialidase ( $\alpha$ 2-3,6,8 sialidase or  $\alpha$ 2-3 sialidase) using two different digestion methods. Method 1: Once the hAGP sample was subjected to PNGase F digestion, the sample was evaporated to dryness in a SpeedVac and reconstituted to the initial volume with 50 mM  $\text{NH}_4\text{Ac}$  (pH 5.0). Subsequently, 1  $\mu\text{L}$  of  $\alpha$ 2-3,6,8 sialidase (50 mU) or 6.25  $\mu\text{L}$  of  $\alpha$ 2-3 sialidase (50 U) was added, and the solution was incubated at 37 °C for 18 h ( $\alpha$ 2-3,6,8 sialidase) or 1 h ( $\alpha$ 2-3 sialidase). Method 2: In this case, sialidase digestion was performed after PNGase F treatment in the same buffer solution of 50 mM  $\text{Na}_3\text{PO}_4$  (pH 7.5) and incubated at the same conditions of temperature and digestion time as Method 1. In both procedures, digestion was stopped by heating the samples in a thermoblock for 10 min at 100 °C, and the samples were purified and labelled with aniline in accordance with the protocol described in Section 2.2. **Fucosidase digestion:**  $\alpha$ 1-3,4 fucosidase was used to release specific fucose residues of hAGP *N*-glycans. Three variants of the fucosidase digestion protocol that differed in the procedure or the number of enzymes used were tested (see supplementary figure). Method 1: After digestion of hAGP with PNGase F, the sample was evaporated to dryness in a SpeedVac and fucosidase digestion was carried out. Method 2: Before fucosidase digestion, the glycans released with PNGase F were purified and labelled as described in Section 2.2. Method 3:

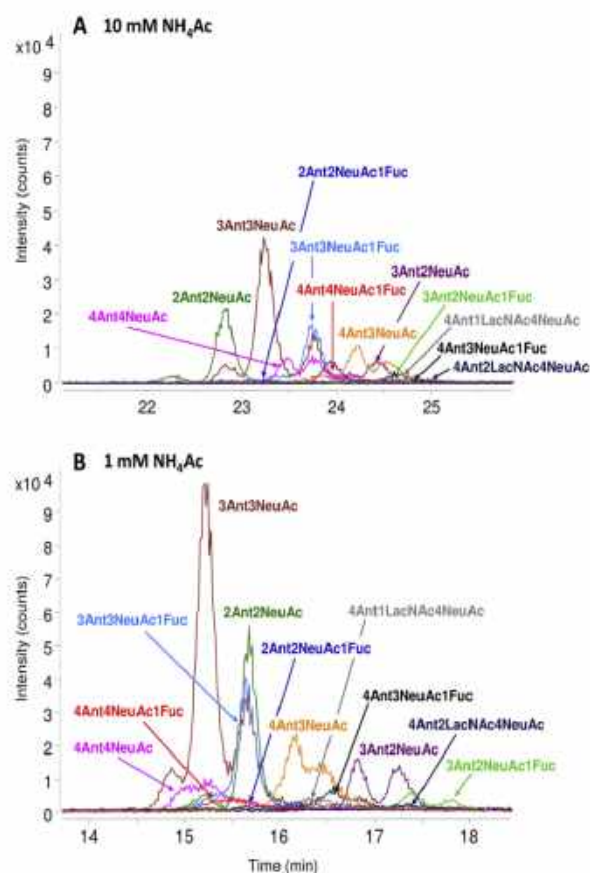


Fig. 2. Extracted ion chromatograms (EICs) of the major *N*-glycans released from hAGP standard sample and labelled with [ $^{12}\text{C}_6$ ]-aniline using the chromatographic gradient reported in Ref. [31], with an aqueous mobile phase of 10 mM  $\text{NH}_4\text{Ac}$  (A) and 1 mM  $\text{NH}_4\text{Ac}$  (B).

Once the hAGP *N*-glycans were purified and labelled, digestion with fucosidase was carried out in combination with other enzymes (1  $\mu$ L of  $\alpha$ 2-3,6,8 sialidase (50 mU) and/or 1  $\mu$ L of  $\beta$ 1-4 galactosidase (8 U)). In all these methods, the dried glycans were dissolved in 8  $\mu$ L of water in order to perform fucosidase digestion. Subsequently, 1  $\mu$ L of 50 mM NaAc:5 mM CaCl<sub>2</sub> (pH 5.5) and 1  $\mu$ L of BSA at a concentration of 1000  $\mu$ g/mL, provided by the manufacturer, were added [36]. Afterwards, 1  $\mu$ L of fucosidase (4 U) was added and the samples were incubated at 37 °C overnight. Digestion was stopped by heating the samples in a thermoblock at 100 °C for 10 min, and samples were stored at -20 °C until analysis.

#### 2.4. $\mu$ ZIC-HILIC-ESI-MS

The  $\mu$ LC-TOF-MS experiments were performed in a 1200 Series capillary liquid chromatography system coupled to a 6220 oa-TOF LC/MS mass spectrometer with an orthogonal G1385-44300 interface (Agilent Technologies). To perform the chromatographic separation, a ZIC-HILIC column packed with 3.5 mm particles, 150  $\times$  0.3 mm L<sub>T</sub>  $\times$  ID (SeQuant, Umeå, Sweden) was used; the stationary phase consisted of a surface with immobilized zwitterionic sulfobetaine moieties. Experiments were performed at room temperature with gradient elution at a flow rate of 4  $\mu$ L/min and injections of between 0.15 and 0.25  $\mu$ L of glycan sample labelled with aniline (glycan concentration: ~100-50 pmol/ $\mu$ L). Eluting solvents were A: 1 mM NH<sub>4</sub>Ac solution and B: acetonitrile. The following gradient conditions were used: solvent B from 90% to 80% (within 5 min) and from 80% to 65% (within 20 min) as the linear gradient, followed by cleaning and equilibration steps of B: 65%  $\rightarrow$  50% (within 5 min), 50%  $\rightarrow$  0% (within 5 min), 0% (over 15 min), 0%  $\rightarrow$  90% (within 5 min) and 90% (over 10 min). The mass spectrometer was equipped with a dual-nebulizer ESI source, and the orthogonal nebulizer was used for the  $\mu$ LC-TOF-MS experiments; the second nebulizer, which is generally used to introduce the internal reference mass standard solution in conventional LC-MS experiments, was disabled to avoid any interference with the  $\mu$ LC-TOF-MS experiments [37]. Tuning and calibration of the mass spectrometer were carried out in accordance with the manufacturer's instructions. The measurement parameters were subsequently fine-tuned by direct infusion of maltohexaose labelled with [<sup>12</sup>C<sub>6</sub>]-aniline to maximize the signal for the singly charged molecular ion [31]. The best ("optimum") operational conditions in negative mode were: capillary voltage -3500 V, drying gas (N<sub>2</sub>) temperature 200 °C, drying gas flow rate 4 L/min, nebulizer gas (N<sub>2</sub>) 15 psi, fragmentor voltage 190 V, skimmer voltage 70 V and OCT 1 RF Vpp voltage 300 V. Data were collected in profile (continuum) at 1 spectrum/s (approx. 10,000 transients/spectrum) between *m/z* 100 and 3200, at the highest resolution mode (4 GHz).  $\mu$ LC-TOF-MS control and data acquisition were performed using MassHunter Workstation software (Agilent Technologies). Qualitative Analysis (version B.04.00) from Mass Hunter Workstation Software was used for data analysis, using a mass tolerance window of 20 ppm to extract the ion chromatogram of each glycan.

### 3. Results and discussion

#### 3.1. Optimization of chromatographic conditions by $\mu$ ZIC-HILIC-TOF-MS

In a previous work, a  $\mu$ ZIC-HILIC-TOF-MS method was developed to separate hAGP *N*-glycan isomers labelled with aniline [32]. Although this method made it possible to separate the different glycan isomers, it was not sensitive enough to detect minor glycans when the glycoprotein of interest had a very low concentration in biological samples or when it presented a low percentage of

glycosylation. In order to improve sensitivity, the chromatographic conditions used for this work were optimized by evaluating the influence of the pH and ionic strength of the mobile phase. Different pH conditions (pH values of 3, 5, 7 and 8) were tested. Although changes in pH were expected to produce selectivity changes for glycans with the stationary phase due to the deprotonation degree of sialic acids, there was no improvement in separation and the MS signal of *N*-glycans decreased considerably compared to the initial conditions (10 mM NH<sub>4</sub>Ac, pH = 6.4). Afterwards, the effect of the salt content was also evaluated, with concentrations of 10, 5 and 1 mM of NH<sub>4</sub>Ac in the mobile phase. Fig. 2A and B shows the extracted ion chromatograms (EICs) of the major *N*-glycans released from hAGP and labelled with [<sup>12</sup>C<sub>6</sub>]-aniline ([<sup>12</sup>C<sub>6</sub>]AN), using 10 mM NH<sub>4</sub>Ac and 1 mM NH<sub>4</sub>Ac mobile phases, respectively. In both cases, the gradient elution used was that reported in Refs. [31,32]. The retention of glycans was affected by the ionic strength and the elution order of some was altered (e.g. 2Ant2NeuAc), but separation generally did not significantly improve with respect to 10 mM NH<sub>4</sub>Ac. By contrast, glycan signals considerably increased at a low salt concentration. Hence, 1 mM of NH<sub>4</sub>Ac at pH 6.4 was chosen as the optimum aqueous mobile phase. Once the mobile phase composition was established, the gradient elution was optimized to improve separation of glycan isomers and obtain an appropriate analysis time.

In line with our previous works [32,33], several bi-, tri- and tetraantennary fucosylated and non-fucosylated *N*-glycans were detected using this approach (see Table 1). Nevertheless, the improvement in sensitivity made it possible to detect more glycan structures of hAGP standard in this study. The biantennary glycan with one NeuAc and the tetraantennary mono- and disialylated glycans were also identified. In addition, it was possible to detect the glycans 4Ant3NeuAc and 4Ant4NeuAc with one and two extra units of *N*-acetylglucosamine (HexHexNAc or LacNAc), some of them also fucosylated. Isobaric isomers were detected for most hAGP glycans using the ZIC-HILIC method reported in Ref. [32]. Moreover, the gradient elution optimized in this work provided a better resolution of isomers, especially for less abundant glycans (data not shown).

#### 3.2. Characterization of sialic acid linkage isomers

Once the method for separating and detecting hAGP glycans and their corresponding isomers was established, these structural isomers were characterized using  $\alpha$ 2-3 sialidase (see Fig. 1A), since it was expected that the main differences between them would be the linkage type of the terminal NeuAc ( $\alpha$ 2-3 or  $\alpha$ 2-6). Prior to the characterization of glycan isomers,  $\alpha$ 2-3,6,8 sialidase (total sialidase) was used to establish the optimum digestion conditions. Two digestion procedures were tested to completely remove the sialic acids from the hAGP glycans. Method 1 used two different buffer solutions for digestion with PNGase F and sialidase, and maintained the optimum working pH for each enzyme (50 mM Na<sub>3</sub>PO<sub>4</sub>, pH 7.5 for PNGase F and 50 mM NH<sub>4</sub>Ac, pH 5.0 for sialidase). By contrast, digestion of the two enzymes in method 2 was carried out using the same buffer solution (50 mM Na<sub>3</sub>PO<sub>4</sub>, pH 7.5). Higher MS signals for the desialylated glycans were obtained by digesting both enzymes in the PNGase F optimum working pH (Method 2), even though the optimum pH for sialidase digestion is 5–6. This method was therefore selected as the most suitable for performing sialidase digestion, as sialic acids with the highest yield were removed and less time was required to carry out the enzymatic digestion.

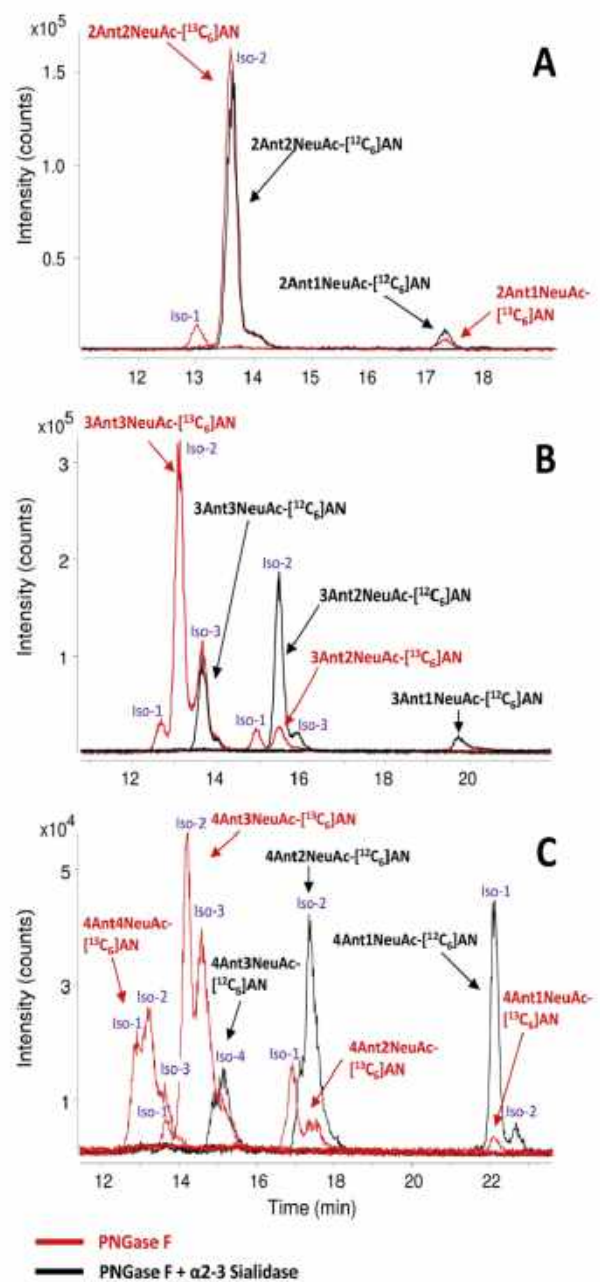
In this study, isotope-coded labelling with [<sup>12</sup>C<sub>6</sub>]-aniline ([<sup>12</sup>C<sub>6</sub>]AN) and [<sup>13</sup>C<sub>6</sub>]-aniline ([<sup>13</sup>C<sub>6</sub>]AN) was performed to reliably evaluate the changes in hAGP glycans caused by digestion of the protein with a certain exoglycosidase. This GRIL strategy made it possible to

**Table 1**hAGP *N*-glycans labelled with [<sup>12</sup>C<sub>6</sub>]AN detected by  $\mu$ ZIC-HILIC-MS in negative ion mode (*N*-glycan concentration: 50 pmol/ $\mu$ L).

| hAGP glycan-[ <sup>12</sup> C <sub>6</sub> ]AN |          | t <sub>R</sub> (min) | M <sub>theo</sub> | M <sub>exp</sub> | Error (ppm) |
|--|----------|----------------------|-------------------|------------------|-------------|
| 2Ant1NeuAc                                     | Isomer 1 | 17.3                 | 2008.7267         | 2008.7298        | 1.5         |
|  | Isomer 2 | 12.9                 | 2299.8232         | 2299.8288        | 2.4         |
| 2Ant2NeuAc                                     | Isomer 1 | 13.5                 | 2299.8232         | 2299.8218        | 0.6         |
|  | Isomer 2 | 14.0                 | 2445.8810         | 2445.8844        | 1.4         |
| 2Ant2NeuAc1Fuc                                 | Isomer 1 | 14.2                 | 2445.8810         | 2445.8884        | 3.0         |
|  | Isomer 2 | 14.8                 | 2664.9554         | 2664.9590        | 1.4         |
| 3Ant2NeuAc                                     | Isomer 1 | 15.4                 | 2664.9554         | 2664.9539        | 0.6         |
|  | Isomer 2 | 15.5                 | 2811.0132         | 2810.9946        | 6.6         |
| 3Ant2NeuAc1Fuc                                 | Isomer 1 | 16.0                 | 2811.0132         | 2810.9850        | 10.0        |
|  | Isomer 2 | 12.6                 | 2956.0513         | 2956.0484        | 1.0         |
| 3Ant3NeuAc                                     | Isomer 1 | 13.0                 | 2956.0513         | 2956.0463        | 1.7         |
|  | Isomer 2 | 13.5                 | 2956.0513         | 2956.0469        | 1.5         |
|  | Isomer 3 | 13.1                 | 3102.1093         | 3102.0830        | 8.5         |
| 3Ant3NeuAc1Fuc                                 | Isomer 1 | 13.5                 | 3102.1093         | 3102.1025        | 2.2         |
|  | Isomer 2 | 13.9                 | 3248.1656         | 3248.1578        | 2.4         |
| 4Ant1NeuAc                                     | Isomer 1 | 22.1                 | 2738.9911         | 2738.9606        | 11.1        |
| 4Ant2NeuAc                                     | Isomer 1 | 17.0                 | 3030.0865         | 3030.0824        | 1.3         |
|  | Isomer 2 | 17.5                 | 3030.0865         | 3030.0818        | 1.5         |
| 4Ant2NeuAc1Fuc                                 | Isomer 1 | 17.6                 | 3176.1444         | 3176.1245        | 6.3         |
|  | Isomer 2 | 18.1                 | 3176.1444         | 3176.1386        | 1.8         |
| 4Ant3NeuAc                                     | Isomer 1 | 13.5                 | 3321.1837         | 3321.1616        | 6.6         |
|  | Isomer 2 | 14.1                 | 3321.1837         | 3321.1778        | 1.8         |
|  | Isomer 3 | 14.4                 | 3321.1837         | 3321.1772        | 1.9         |
|  | Isomer 4 | 14.9                 | 3321.1837         | 3321.1664        | 5.2         |
|  | Isomer 5 | 13.9                 | 3467.2414         | 3467.2259        | 4.5         |
| 4Ant3NeuAc1Fuc                                 | Isomer 1 | 14.6                 | 3467.2414         | 3467.2289        | 3.6         |
|  | Isomer 2 | 14.9                 | 3467.2414         | 3467.2181        | 6.7         |
|  | Isomer 3 | 15.3                 | 3467.2414         | 3467.2226        | 5.4         |
|  | Isomer 4 | 15.0                 | 3613.2978         | 3613.2707        | 7.5         |
| 4Ant3NeuAc2Fuc                                 | Isomer 1 | 15.4                 | 3613.2978         | 3613.2509        | 13.0        |
|  | Isomer 2 | 12.8                 | 3612.2791         | 3612.2714        | 2.1         |
| 4Ant4NeuAc                                     | Isomer 1 | 13.1                 | 3612.2791         | 3612.2582        | 5.8         |
|  | Isomer 2 | 13.5                 | 3612.2791         | 3612.2681        | 3.0         |
|  | Isomer 3 | 13.1                 | 3758.3368         | 3758.3177        | 5.1         |
| 4Ant4NeuAc1Fuc                                 | Isomer 1 | 13.5                 | 3758.3368         | 3758.3147        | 5.9         |
|  | Isomer 2 | 13.7                 | 3904.3932         | 3904.3727        | 5.2         |
| 4Ant4NeuAc3Fuc                                 | Isomer 1 | 14.2                 | 4050.4511         | 4050.3887        | 15.4        |
| 4Ant1LacNAc3NeuAc                              | Isomer 1 | 16.2                 | 3686.3141         | 3686.3096        | 1.2         |
| 4Ant1LacNAc4NeuAc                              | Isomer 1 | 14.6                 | 3977.4096         | 3977.4023        | 1.8         |
| 4Ant1LacNAc3NeuAc1Fuc                          | Isomer 1 | 16.5                 | 3832.3720         | 3832.3859        | 3.6         |
| 4Ant1LacNAc4NeuAc1Fuc                          | Isomer 1 | 14.9                 | 4123.4675         | 4123.4318        | 8.7         |
| 4Ant1LacNAc4NeuAc2Fuc                          | Isomer 1 | 15.3                 | 4269.5254         | 4269.5423        | 4.0         |
| 4Ant1LacNAc4NeuAc3Fuc                          | Isomer 1 | 15.5                 | 4415.5833         | 4415.5567        | 6.0         |
| 4Ant2LacNAc3NeuAc                              | Isomer 1 | 17.3                 | 4051.4463         | 4051.4378        | 2.1         |
| 4Ant2LacNAc4NeuAc                              | Isomer 1 | 15.8                 | 4342.5417         | 4342.5251        | 3.8         |
| 4Ant2LacNAc4NeuAc1Fuc                          | Isomer 1 | 16.2                 | 4488.5997         | 4488.6103        | 2.4         |

unequivocally assign the sialic acid linkage types present in each isomer by direct comparison of the peak areas of the intact glycans (hAGP only digested with PNGase F) with those digested with a specific sialidase (hAGP digested with PNGase F and  $\alpha$ 2-3 sialidase), since it avoids experimental variability in retention times and peak areas between runs. For this purpose, three independent samples of 25  $\mu$ g of standard hAGP were digested with PNGase F and labelled with [ $^{13}\text{C}_6$ ]AN, while another three were digested with PNGase F and  $\alpha$ 2-3 sialidase in accordance with method 2, and labelled with [ $^{12}\text{C}_6$ ]AN. In this case, the time allowed for PNGase F digestion was 18 h to ensure the complete de-glycosylation of hAGP, but  $\alpha$ 2-3 sialidase digestion was limited to one hour, as recommended by the manufacturer, to avoid any possible loss of  $\alpha$ 2-6 sialic acids in overnight digestion. After derivatization, equimolar mixtures of hAGP glycan-[ $^{13}\text{C}_6$ ]AN and hAGP  $\alpha$ 2-3-desialylated glycan-[ $^{12}\text{C}_6$ ]AN were prepared and analysed by  $\mu$ ZIC-HILIC-TOF-MS. Fig. 3 illustrates the results obtained for the bi-, tri- and tetraantennary non-fucosylated glycans. The EICs of the glycans detected in the hAGP sample digested only with PNGase F are shown in red, and the EICs of the glycans detected in the hAGP sample additionally digested with  $\alpha$ 2-3 sialidase are shown in black. With regard to biantennary non-fucosylated glycans (Fig. 3A), hAGP digested with PNGase F showed the presence of 2Ant2NeuAc with two isomers and 2Ant1NeuAc. After  $\alpha$ 2-3 sialidase digestion, isomer 1 of 2Ant2NeuAc disappeared and the peak area of the glycan 2Ant1NeuAc increased, while isomer 2 of 2Ant2NeuAc was maintained at the same retention time and presented practically the same peak area (see Fig. 3A and Table 2). These facts reveal that isomer 2 of 2Ant2NeuAc only presents  $\alpha$ 2-6 linked sialic acids and isomer 1 contains one  $\alpha$ 2-6 linked sialic acid and one  $\alpha$ 2-3 linked sialic acid. Moreover, the sialic acid of 2Ant1NeuAc is only  $\alpha$ 2-6 linked, as the 2Ant0NeuAc was not detected after  $\alpha$ 2-3 sialidase digestion (see Table 2). It is worth mentioning that the peak area of 2Ant1NeuAc after  $\alpha$ 2-3 sialidase digestion was not exactly equal to the sum of the peak areas corresponding to isomer 1 of 2Ant2NeuAc and the glycan 2Ant1NeuAc. This is because a glycan with two sialic acids shows a higher ionization yield in negative mode compared to the same glycan with only one sialic acid residue.

With respect to the triantennary non-fucosylated glycans, isomers 1 and 2 of the glycan 3Ant3NeuAc completely disappeared after  $\alpha$ 2-3 sialidase treatment (EICs in black), and only the third isomer remained. This presented a similar peak area, which suggests that this isomer presents three  $\alpha$ 2-6 NeuAc linkages. To correctly assign the remaining isomers of glycan 3Ant2NeuAc, it was first necessary to study the behaviour of glycan 3Ant0NeuAc. In this case, after specific sialidase digestion, the first isomer disappeared, the peak area of the second increased considerably and a third peak appeared at a different retention time (Isomer 3). Additionally, it was assumed that hAGP does not present any triantennary glycan with all  $\alpha$ 2-3 NeuAc linkages, since the glycan 3Ant0NeuAc was not detected in the hAGP sample digested with  $\alpha$ 2-3 sialidase. Isomer 1 of the glycan 3Ant2NeuAc shows one  $\alpha$ 2-3 and one  $\alpha$ 2-6 linked sialic acid and, after sialidase digestion, it turned into the glycan 3Ant1NeuAc, which appeared at a  $t_R = 19.6$  min, with a similar peak area (292,825 versus 314,765, see Table 2). This fact revealed that no other isomer contributed to the peak area of the glycan 3Ant1NeuAc. Hence, isomers 1 and 2 of the glycan 3Ant3NeuAc contain two  $\alpha$ 2-6 linked sialic acids and one  $\alpha$ 2-3 linked sialic acid, and the second isomer of 3Ant2NeuAc also showed sialic acids with  $\alpha$ 2-6 linkages. As with the biantennary glycans, the differences in ionization between the intact and desialylated glycans made it difficult to accurately compare the areas obtained (see Table 2). After evaluation of the different possibilities, however, we propose that isomers 1 and 2 of the glycan 3Ant3NeuAc turn into isomers 3 and 2 of the glycan 3Ant2NeuAc,



**Fig. 3.** Equimolar mixture of aniline-labelled N-glycans released from 25  $\mu$ g of standard hAGP digested only with PNGase F and labelled with [ $^{13}\text{C}_6$ ]-aniline (EICs in red), and 25  $\mu$ g of standard hAGP digested with PNGase F and  $\alpha$ 2-3 sialidase and labelled with [ $^{12}\text{C}_6$ ]-aniline (EICs in black). EICs of the biantennary (A), triantennary (B) and tetraantennary (C) non-fucosylated glycans. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

respectively, after  $\alpha$ 2-3 sialidase digestion.

hAGP tetraantennary glycans were more difficult to characterize, since multiple combinations of the terminal sialic acid linkages were possible. Nevertheless, most linkages in each glycan isomer were characterized by evaluating all of the possible assignments and taking account of the peak areas obtained in the two samples (intact hAGP glycans and desialylated  $\alpha$ 2-3 hAGP glycans, Fig. 3C). The tentative assignment obtained for the tetraantennary glycans is also presented in Table 2. The sialic acid linkage assignment of the remaining glycans, including fucosylated glycans (illustrated in Table 3), was carried out with the same strategy by

direct comparison of the peak areas. Less abundant glycans could not be characterized, however, since the MS signal was too low to perform a reliable assignment (e.g. 4Ant4NeuAc3Fuc, 4Ant2LacNAc4NeuAc, 4Ant1LacNAc4NeuAc1Fuc and 4Ant2LacNAc4NeuAc1Fuc). It is important to note that, although certain ionization differences between the intact and desialylated glycans could not be avoided with the GRIL approach, this complex characterization of isomeric glycans would not have been possible without this strategy. To our knowledge, very few studies in the literature have characterized triantennary and tetraantennary glycans with such a large number of isomers for each glycan [27]. This is mainly due to








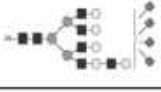
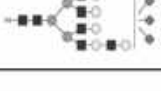
the limitations presented by other methodologies [6,16,20–26]. In our previous work [32], certain isobaric glycans (e.g. isomer 1 and 2 of 3Ant3NeuAc) were found to be underexpressed in pancreatic cancer with respect to control (ratio values were between 0.6 and 0.4, depending on the stage of the PDAC), while the abundance of isomer 3 of 3Ant3NeuAc was similar in the control and PDAC samples (ratio values close to 1). These findings could be related to our sialic acid linkage assignments, since isomer 3 only presented  $\alpha$ 2-6 linked sialic acids, while isomers 1 and 2 showed  $\alpha$ 2-3 linked-type sialic acids. These results demonstrate the importance of characterizing these isomers, because changes in the expression of

**Table 2**

Retention time and peak areas of native non-fucosylated hAGP-glycans (derivatized with [ $^{13}\text{C}_6$ ]AN) and non-fucosylated hAGP-glycans desialylated with  $\alpha$ 2-3 sialidase (derivatized with [ $^{12}\text{C}_6$ ]AN) by  $\mu$ ZIC-HILIC-TOF-MS. Sialic acid linkage assignment for each isomer.

|            | $t_R$ (min)             | Glycan-[ $^{13}\text{C}_6$ ]AN area (n = 3) | $\alpha$ 2-3 desialylated glycan-[ $^{12}\text{C}_6$ ]AN area (n = 3) | NeuAc linkage assignment |   |   |
|------------|-------------------------|---|---|--------------------------|---|---|
| 2Ant2NeuAc | Isomer 1                | 13.6  | 169,825   | –                        | 1 NeuAc $\alpha$ 2-6<br>1 NeuAc $\alpha$ 2-3                    |   |
|            | Isomer 2                | 14.0  | 2,010,769   | 1,783,084                | 2 NeuAc $\alpha$ 2-6  |   |
| 2Ant1NeuAc | Isomer 1                | 17.3  | 68,941  | 135,571                  | 1 NeuAc $\alpha$ 2-6  |   |
| 3Ant3NeuAc | Isomer 1                | 26.5  | –   | –                        | –   | – |
|            | Isomer 1                | 12.6  | 396,885   | –                        | 2 NeuAc $\alpha$ 2-6<br>1 NeuAc $\alpha$ 2-3                    |   |
|            | Isomer 2                | 13.1  | 4,156,023   | –                        | 2 NeuAc $\alpha$ 2-6<br>1 NeuAc $\alpha$ 2-3                    |   |
| 3Ant2NeuAc | Isomer 3                | 13.7  | 1,536,871   | 1,412,485                | 3 NeuAc $\alpha$ 2-6  |   |
|            | Isomer 1                | 14.9  | 292,825   | –                        | 1 NeuAc $\alpha$ 2-6<br>1 NeuAc $\alpha$ 2-3                    |   |
|            | Isomer 2                | 15.4  | 386,282   | 2,277,812                | 2 NeuAc $\alpha$ 2-6  |   |
| 3Ant1NeuAc | Isomer 3 <sup>b</sup>   | 15.9  | –   | 225,692                  | –   | – |
| 3Ant1NeuAc | Isomer 1 <sup>b</sup>   | 19.6  | –   | 314,765                  | –   | – |
| 4Ant4NeuAc | 3Ant0NeuAc <sup>c</sup> | Isomer 1                                    | 29.8  | –                        | –   | – |
|            | Isomer 1                | 12.9  | 252,257   | –                        | 1 NeuAc $\alpha$ 2-6<br>2 NeuAc $\alpha$ 2-3<br>1 NeuAc unknown |   |
|            | Isomer 2                | 13.3  | 317,537   | –                        | 1 NeuAc $\alpha$ 2-6<br>2 NeuAc $\alpha$ 2-3<br>1 NeuAc unknown |   |
| 4Ant4NeuAc | Isomer 3                | 13.5  | 126,903   | –                        | 2 NeuAc $\alpha$ 2-6<br>2 NeuAc $\alpha$ 2-3                    |   |

Table 2 (continued)

|                                | $t_R$ (min)           | Glycan- $^{13}C_6$ ]AN area (n = 3) | $\alpha$ 2-3 desialylated glycan- $^{13}C_6$ ]AN area (n = 3) | NeuAc linkage assignment |   |   |
|--------------------------------|-----------------------|-------------------------------------|---|--------------------------|---|---|
| 4Ant3NeuAc                     | Isomer 1              | 13.6                                | 101,293   | –                        | 1 NeuAc $\alpha$ 2-6<br>2 NeuAc $\alpha$ 2-3                    |    |
|                                | Isomer 2              | 14.2                                | 883,588   | –                        | 1 NeuAc $\alpha$ 2-6<br>1 NeuAc $\alpha$ 2-3<br>1 NeuAc unknown |    |
|                                | Isomer 3              | 14.6                                | 670,502   | –                        | 1 NeuAc $\alpha$ 2-6<br>1 NeuAc $\alpha$ 2-3<br>1 NeuAc unknown |    |
|                                | Isomer 4              | 15.1                                | 214,102   | 397,994                  | 3 NeuAc $\alpha$ 2-6  |    |
| 4Ant2NeuAc                     | Isomer 1              | 16.9                                | 195,765   | –                        | 1 NeuAc $\alpha$ 2-6<br>1 NeuAc $\alpha$ 2-3                    |    |
|                                | Isomer 2              | 17.4                                | 66,762  | 1,004,143                | 2 NeuAc $\alpha$ 2-6  |    |
| 4Ant1NeuAc                     | Isomer 1              | 22.1                                | 43,075  | 648,373                  | 1 NeuAc $\alpha$ 2-6  |  |
|                                | Isomer 2 <sup>b</sup> | 22.7                                | –   | 76,451                   | –   |   |
| 4Ant0NeuAc <sup>a</sup>        | Isomer 1              | 32.4                                | –   | –                        | –   |   |
| 4Ant1LacNac4NeuAc              | Isomer 1              | 14.6                                | 425,915   | –                        | 1 NeuAc $\alpha$ 2-6<br>2 NeuAc $\alpha$ 2-3<br>1 NeuAc unknown |  |
| 4Ant1LacNac3NeuAc              | Isomer 1              | 16.2                                | 248,701   | 77,459                   | 1 NeuAc $\alpha$ 2-6<br>1 NeuAc $\alpha$ 2-3<br>1 NeuAc unknown |  |
| 4Ant1LacNac2NeuAc              | Isomer 1 <sup>b</sup> | 19.6                                | –   | 112,626                  | –   |   |
| 4Ant1LacNac1NeuAc              | Isomer 1 <sup>b</sup> | 24.7                                | –   | 32,151                   | –   |   |
| 4Ant1LacNac0NeuAc <sup>a</sup> | Isomer 1              | 33.9                                | –   | –                        | –   |   |

<sup>a</sup> Glycan not present in the native protein. Detected only after total sialidase digestion.

<sup>b</sup> Glycan isomer not present in the native protein. Detected only after  $\alpha$ 2-3 sialidase digestion.

$\alpha$ 2-3 and  $\alpha$ 2-6 linked sialic acids have also previously been described in cancer [13].

Total sialidase digestion of hAGP glycans was also performed in this work. The EICs of the resulting total desialylated glycans (non-fucosylated and fucosylated) are shown in Fig. 4. It is important to highlight that, while the non-fucosylated glycans exhibited only one chromatographic peak for each glycan, different isomers were observed in the case of fucosylated desialylated glycans. The fact that they were only present in fucosylated glycans may reveal the presence of different fucose types ( $\alpha$ 1-6 core Fuc,  $\alpha$ 1-3 antennary Fuc or  $\alpha$ 1-2 linked Fuc to Gal). Although only  $\alpha$ 1-3 antennary fucosylation was previously reported in hAGP and  $\alpha$ 1-6 core fucosylation was only detected in the glycan 2Ant2NeuAc1Fuc [12], these results confirm that fucose linkage assignment in all hAGP fucosylated glycans is still required in order to identify possible alterations of these isomers in pathological samples.

### 3.3. Characterization of fucose linkage isomers

We ruled out the presence of  $\alpha$ 1-4 fucoses, since only type-II structures (Gal $\beta$ 1-4GlcNAc) were previously described in hAGP [38]. To establish the optimum release of  $\alpha$ 1-3 antennary fucoses from hAGP glycans using  $\alpha$ 1-3,4 fucosidase, three different digestion methods were evaluated (see supplementary figure). In method 1, fucosidase digestion was carried out after PNGase F digestion by changing the digestion buffer, while in method 2, it was performed after PNGase F digestion and glycan purification and labelling. Both procedures were rejected, since any fucose residue was removed from the hAGP glycans, probably due to the steric impediment of the terminal antennary monosaccharides such as sialic acids and galactoses, which hindered fucosidase accessibility.  $\alpha$ 1-3,4 fucosidase digestion only worked successfully when method 3 was used (method 2 in combination with total sialidase and/or

**Table 3**

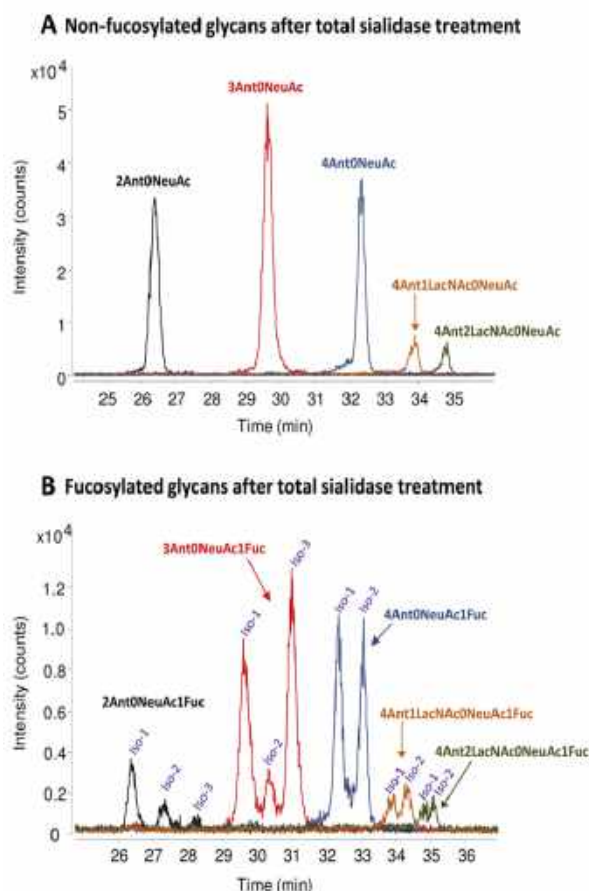
Retention time and peak areas of native fucosylated hAGP-glycans (derivatized with [<sup>13</sup>C<sub>6</sub>]AN) and fucosylated hAGP-glycans desialylated with α2-3 sialidase (derivatized with [<sup>13</sup>C<sub>6</sub>]AN) by μZIC-HILIC-TOF-MS. Sialic acid linkage assignment for each isomer.

|                             |                       | t <sub>R</sub> (min) | Glycan-[ <sup>13</sup> C <sub>6</sub> ]AN area (n = 3) | α2-3 desialylated glycan-[ <sup>13</sup> C <sub>6</sub> ]AN area (n = 3) | NeuAc linkage assignment                        |
|-----------------------------|-----------------------|----------------------|--|--|---|
| 2Ant2NeuAc1Fuc              | Isomer 1              | 13.9                 | 40,007   | –  | 1 NeuAc α2-6<br>1 NeuAc α2-3                    |
|                             | Isomer 2              | 14.2                 | 43,446   | 38,801   | 2 NeuAc α2-6                                    |
| 2Ant1NeuAc1Fuc              | Isomer 1 <sup>a</sup> | 18.6                 | –  | 31,219   | –   |
| 2Ant0NeuAc1Fuc <sup>c</sup> | Isomer 1              | 26.4                 | –  | –  | –   |
|                             | Isomer 2              | 27.2                 | –  | –  | –   |
|                             | Isomer 3              | 28.2                 | –  | –  | –   |
| 3Ant3NeuAc1Fuc              | Isomer 1              | 13.2                 | 182,067  | –  | 2 NeuAc α2-6<br>1 NeuAc α2-3                    |
|                             | Isomer 2              | 13.7                 | 2,237,472  | –  | 2 NeuAc α2-6<br>1 NeuAc α2-3                    |
| 3Ant2NeuAc1Fuc              | Isomer 1              | 15.6                 | 145,101  | –  | 1 NeuAc α2-6<br>1 NeuAc α2-3                    |
|                             | Isomer 2              | 16.2                 | 106,913  | 974,367  | 2 NeuAc α2-6                                    |
|                             | Isomer 3 <sup>b</sup> | 16.7                 | –  | 116,369  | –   |
| 3Ant1NeuAc1Fuc              | Isomer 1 <sup>b</sup> | 20.8                 | –  | 106,131  | –   |
| 3Ant0NeuAc1Fuc <sup>c</sup> | Isomer 1              | 29.6                 | –  | –  | –   |
|                             | Isomer 2              | 30.3                 | –  | –  | –   |
|                             | Isomer 3              | 30.9                 | –  | –  | –   |
| 3Ant3NeuAc2Fuc              | Isomer 1              | 14.0                 | 56,253   | –  | 2 NeuAc α2-6<br>1 NeuAc α2-3                    |
| 3Ant2NeuAc2Fuc              | Isomer 1 <sup>b</sup> | 16.6                 | –  | 30,442   | –   |
| 4Ant4NeuAc1Fuc              | Isomer 1              | 13.3                 | 140,718  | –  | 1 NeuAc α2-6<br>2 NeuAc α2-3<br>1 NeuAc unknown |
|                             | Isomer 2              | 13.6                 | 180,386  | –  | 1 NeuAc α2-6<br>2 NeuAc α2-3<br>1 NeuAc unknown |
|                             | Isomer 1              | 14.1                 | 35,632   | –  | 1 NeuAc α2-6<br>1 NeuAc α2-3<br>1 NeuAc unknown |
|                             | Isomer 2              | 14.7                 | 269,095  | –  | 1 NeuAc α2-6<br>1 NeuAc α2-3<br>1 NeuAc unknown |
| 4Ant3NeuAc1Fuc              | Isomer 3              | 15.0                 | 162,968  | –  | 1 NeuAc α2-6<br>1 NeuAc α2-3<br>1 NeuAc unknown |
|                             | Isomer 4              | 15.4                 | 51,521   | 49,623   | 3 NeuAc α2-6                                    |
|                             | Isomer 1              | 17.5                 | 47,920   | 48,603   | 2 NeuAc α2-6                                    |
| 4Ant2NeuAc1Fuc              | Isomer 2              | 18.1                 | 17,739   | 152,919  | 2 NeuAc α2-6                                    |
|                             | Isomer 3 <sup>b</sup> | 18.6                 | –  | 30,699   | –   |
|                             | Isomer 1 <sup>b</sup> | 22.1                 | –  | 23,485   | –   |
| 4Ant1NeuAc1Fuc              | Isomer 2 <sup>b</sup> | 23.0                 | –  | 161,345  | –   |
|                             | Isomer 3 <sup>b</sup> | 23.6                 | –  | 21,748   | –   |
|                             | Isomer 1              | 32.2                 | –  | –  | –   |
| 4Ant0NeuAc1Fuc <sup>c</sup> | Isomer 2              | 33.0                 | –  | –  | –   |
| 4Ant4NeuAc2Fuc              | Isomer 1              | 13.7                 | 86,151   | –  | 1 NeuAc α2-6<br>2 NeuAc α2-3<br>1 NeuAc unknown |
|                             | Isomer 1              | 15.0                 | 41,296   | –  | 1 NeuAc α2-6<br>1 NeuAc α2-3<br>1 NeuAc unknown |
| 4Ant3NeuAc2Fuc              | Isomer 2              | 15.4                 | 42,600   | 38,949   | 3 NeuAc α2-6                                    |
| 4Ant2NeuAc2Fuc              | Isomer 1 <sup>b</sup> | 18.7                 | –  | 29,613   | –   |
| 4Ant1NeuAc2Fuc              | Isomer 1 <sup>b</sup> | 23.7                 | –  | 33,914   | –   |
| 4Ant0NeuAc2Fuc <sup>c</sup> | Isomer 1              | 33.5                 | –  | –  | –   |

<sup>a</sup> Glycan not present in the native protein. Detected only after total sialidase digestion.

<sup>b</sup> Glycan isomer not present in the native protein. Detected only after α2-3 sialidase digestion.

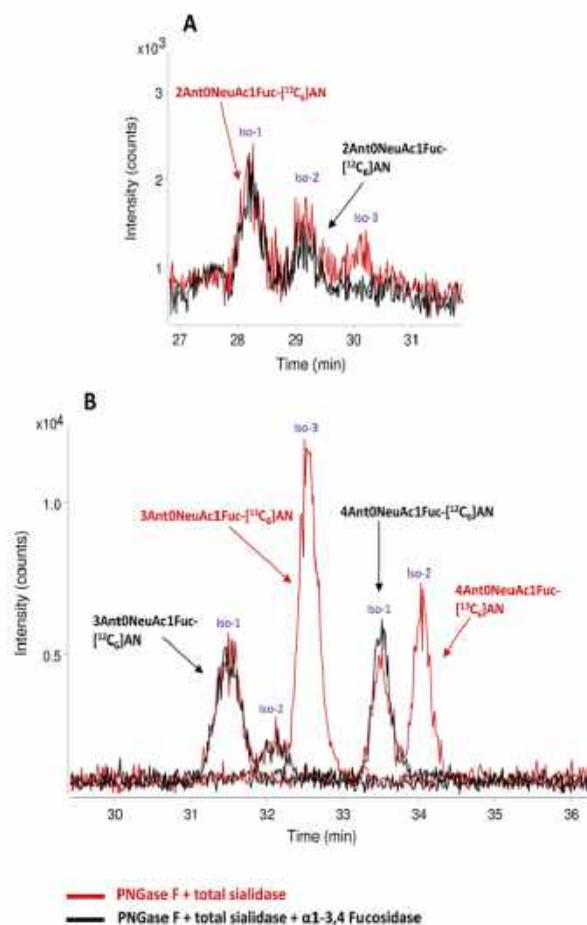




**Fig. 4.** EICs of some desialylated non-fucosylated (A) and fucosylated (B) glycans obtained from hAGP standard after  $\alpha$ 2-3,6,8 sialidase (total sialidase) digestion.

$\beta$ 1-4 galactosidase digestion). The need for sialidase or galactosidase during  $\alpha$ 1-3,4 fucosidase digestion demonstrated the failure of this enzyme to release antennary fucoses in highly decorated glycans, contrary to the manufacturer's guidelines.

Once method 3 was selected as the optimum procedure, fucose linkage-type isomer assignment was carried out using the GRIL strategy (see Fig. 1B). Three independent samples of 10  $\mu$ g of standard hAGP were digested with PNGase F and total sialidase and labelled with [ $^{13}\text{C}_6$ ]AN, while a further three samples were digested with PNGase F, total sialidase and  $\alpha$ 1-3,4 fucosidase and labelled with [ $^{12}\text{C}_6$ ]AN. A 1:1 M mixture of these samples was then prepared and analysed by  $\mu$ ZIC-HILIC-TOF-MS. Glycan fingerprints obtained for bi-, tri- and tetraantennary desialylated glycans with one fucose are illustrated in Fig. 5. The EICs, shown in red, are the hAGP glycans digested with total sialidase and labelled with [ $^{13}\text{C}_6$ ]AN, whereas the EICs shown in black are the desialylated glycans detected in hAGP, but also digested with  $\alpha$ 1-3,4 fucosidase and labelled with [ $^{12}\text{C}_6$ ]AN. With respect to the desialylated biantennary glycan with one fucose (Fig. 5A), isomer 3 was not detected after  $\alpha$ 1-3,4 fucosidase digestion, while isomer 1 and 2 were maintained at the same retention time and with similar peak areas (see also Table 4). This confirmed the notion that the first two isomers may present the  $\alpha$ 1-6 linked fucose to the core GlcNAc, as previously reported in Ref. [12], or  $\alpha$ 1-2 linked to Gal, while isomer 3 may show  $\alpha$ 1-3 antennary fucosylation. In the case of the glycan 3Ant0NeuAc1Fuc (Fig. 5B), digestion with  $\alpha$ 1-3,4 fucosidase also caused the third isomer to disappear (compare red and black EICs), which suggests  $\alpha$ 1-3 antennary fucosylation for isomer 3 and, contrary to our expectations and inconsistent with



**Fig. 5.** Equimolar mixture of aniline-labelled N-glycans released from 10  $\mu$ g of hAGP standard sample digested with PNGase F and total sialidase and labelled with [ $^{13}\text{C}_6$ ]AN (EICs in red), and 10  $\mu$ g of standard hAGP digested with PNGase F, total sialidase and  $\alpha$ 1-3,4 fucosidase and labelled with [ $^{12}\text{C}_6$ ]AN (EICs in black). EICs of the bi- (A), tri- and tetraantennary (B) desialylated fucosylated glycans. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

reported studies [12,33], an  $\alpha$ 1-6 core or  $\alpha$ 1-2 linked to Gal fucosylation for the first two isomers. Finally, the tetraantennary structures with one fucose unit (4Ant0NeuAc1Fuc, 4Ant1LacNAc0NeuAc1Fuc and 4Ant2LacNAc0NeuAc1Fuc) seemed to present one isomer with an  $\alpha$ 1-3 linked fucose and the other with an  $\alpha$ 1-6 core fucose or  $\alpha$ 1-2 fucose linked to Gal, since it was maintained after fucosidase digestion (Fig. 5B and Table 4).

As occurred in the characterization of sialic acids, fucose type assignment of the highly fucosylated glycans was more difficult to perform, since their abundance was very low in hAGP. In this case, it can be only deduced that all highly fucosylated glycan isomers show at least one  $\alpha$ 1-3 antennary fucose, with the exception of isomer 1 of the glycan 4Ant0NeuAc2Fuc (see Table 4).

The reproducibility of the EICs between runs, concerning peak areas, was already evaluated in our previous work [31]. In this work, similar RSD values were obtained ( $\sim$ 10% for major and  $\sim$ 20% for minor glycans) which demonstrated that the method was adequate for sialic linkage-type isomer assignment. However, in the case of fucose linkage-type isomer characterization, as glycan samples were more complex (containing fucosidase, BSA, etc.), the reproducibility was worse and we recommend to compare peak areas within the same run when using the GRIL method to avoid experimental variability between runs. Nevertheless, it is important to remark that the results concerning isomer characterization

**Table 4**

Retention time and peak areas of desialylated hAGP-glycans (derivatized with [ $^{13}\text{C}_6$ ]AN) and desialylated hAGP-glycans digested with  $\alpha$ 1-3,4 fucosidase (derivatized with [ $^{12}\text{C}_6$ ]AN) by  $\mu\text{ZIC-HILIC-TOF-MS}$ . Fucose linkage assignation for each fucosylated glycan isomer.

|                |          | $t_R$ (min) | Desialylated glycan-[ $^{13}\text{C}_6$ ]AN area | Desialylated + $\alpha$ 1-3,4 defucosylated glycan-[ $^{12}\text{C}_6$ ]AN area | Fucose linkage assignation                 |
|----------------|----------|-------------|--|---|--|
| 2Ant0NeuAc0Fuc | Isomer 1 | 28.2        | 392,811  | 348,791   | –  |
|                | Isomer 1 | 28.3        | 31,920   | 29,711  | 1 Fuc core                                 |
| 2Ant0NeuAc1Fuc | Isomer 2 | 29.0        | 20,008   | 17,474  | 1 Fuc core                                 |
|                | Isomer 3 | 30.2        | 12,472   | –   | 1 Fuc Antenna $\alpha$ 1-3 linked          |
| 3Ant0NeuAc0Fuc | Isomer 1 | 31.5        | 644,160  | 666,457   | –  |
|                | Isomer 1 | 31.5        | 92,394   | 95,158  | 1 Fuc core                                 |
| 3Ant0NeuAc1Fuc | Isomer 2 | 32.1        | 19,997   | 25,852  | 1 Fuc core                                 |
|                | Isomer 3 | 32.5        | 183,321  | –   | 1 Fuc Antenna $\alpha$ 1-3 linked          |
| 3Ant0NeuAc2Fuc | Isomer 1 | 32.6        | 48,452   | –   | At least 1 Fuc Antenna $\alpha$ 1-3 linked |
| 4Ant0NeuAc0Fuc | Isomer 1 | 33.5        | 298,959  | 345,043   | –  |
| 4Ant0NeuAc1Fuc | Isomer 1 | 33.5        | 64,200   | 76,963  | 1 Fuc core                                 |
|                | Isomer 2 | 34.0        | 78,593   | –   | 1 Fuc Antenna $\alpha$ 1-3 linked          |
|                | Isomer 1 | 33.5        | 24,799   | 24,713  | 2 Fuc core or $\alpha$ 1-2 linked to Gal   |
| 4Ant0NeuAc2Fuc | Isomer 2 | 34.1        | 20,435   | –   | At least 1 Fuc Antenna $\alpha$ 1-3 linked |
|                | Isomer 3 | 34.4        | 22,952   | –   | At least 1 Fuc Antenna $\alpha$ 1-3 linked |
| 4Ant0NeuAc3Fuc | Isomer 1 | 34.1        | 7247   | –   | At least 1 Fuc Antenna $\alpha$ 1-3 linked |
|                | Isomer 2 | 34.4        | 8236   | –   | At least 1 Fuc Antenna $\alpha$ 1-3 linked |

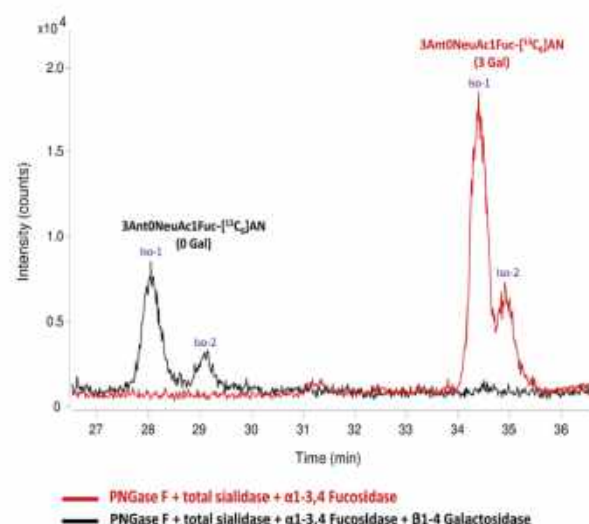
were the same between replicates.

Finally, in order to differentiate between  $\alpha$ 1-2 fucosylation linked to Gal and core fucosylation, total sialidase and  $\alpha$ 1-3,4 fucosidase digestion in combination with  $\beta$ 1-4 galactosidase was carried out. Three independent samples of 10  $\mu\text{g}$  of standard hAGP were digested with PNGase F, total sialidase and  $\alpha$ 1-3,4 fucosidase and labelled with [ $^{13}\text{C}_6$ ]AN, while a further three were digested additionally with  $\beta$ 1-4 galactosidase and labelled with [ $^{12}\text{C}_6$ ]AN. Afterwards, equimolar mixtures of these samples were prepared and analysed by  $\mu\text{ZIC-HILIC-TOF-MS}$ . Isomer profiles obtained after  $\beta$ 1-4 galactosidase digestion confirmed that none of the glycans with one fucose showed  $\alpha$ 1-2 fucoses linked to Gal, as the glycan isomers which were detected after  $\alpha$ 1-3,4 fucosidase digestion were also detected after  $\beta$ 1-4 galactosidase treatment, thereby conserving the fucose unit in their structure after removal of the galactoses (e.g. see Fig. 6 for 3Ant0NeuAc1Fuc). In the case of the highly fucosylated glycans, the absence of  $\alpha$ 1-2 fucoses could not be reliably confirmed, since the MS signals were rather low. Nonetheless, the results obtained with the glycans with one fucose revealed that the majority of hAGP glycans showed  $\alpha$ 1-3 fucosylation (Fuc Antenna), but also  $\alpha$ 1-6 fucosylation (Fuc Core). The presence of two isomers with core fucosylation in glycans 2Ant0NeuAc1Fuc and 3Ant0NeuAc1Fuc (Table 4) could not be fully explained in terms of structure. One possible explanation is the presence of  $\alpha$ 1-6 core fucoses linked to either GlcNAc of the chitobiose core, although this atypical core structure has not been described previously in hAGP. Despite the fact that core fucosylation in hAGP was reported only in the glycan 2Ant2NeuAc1Fuc [12], the presence of the other core fucosylated structures detected in this work opens up an opportunity to evaluate the fucosylation type that is truly involved in the overexpression of hAGP fucosylated glycans in pathological samples, in order to explore the potential of hAGP as a pancreatic cancer biomarker.

#### 4. Conclusions

In the first part of this paper, the conditions for separation and

detection of glycan isomers using the  $\mu\text{ZIC-HILIC-TOF-MS}$  methodology were optimized, which resulted in a significant improvement in sensitivity that could prove useful for the future analysis of biological samples or glycoproteins with low glycosylation percentages. In addition, this approach identified more glycan structures in hAGP than ever before with this methodology (2Ant1NeuAc, 4Ant1NeuAc and 4Ant2NeuAc, as well as some tetraantennary structures with one and two extra units of N-acetylglucosamine). In the second part, we established a methodology



**Fig. 6.** Equimolar mixture of aniline-labelled N-glycans released from 10  $\mu\text{g}$  of standard hAGP digested with PNGase F and total sialidase and labelled with [ $^{13}\text{C}_6$ ]-aniline, and 10  $\mu\text{g}$  of standard hAGP digested with PNGase F, total sialidase and  $\beta$ 1-4 galactosidase and labelled with [ $^{12}\text{C}_6$ ]-aniline. EIC of the triantennary desialylated fucosylated glycan (shown in red) and EIC of the triantennary desialylated fucosylated glycans also digested with  $\beta$ 1-4 galactosidase (shown in black). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

using a GRIL strategy with [ $^{12}\text{C}_6$ ]/[ $^{13}\text{C}_6$ ]-aniline and specific exoglycosidase digestions to assign sialic acid and fucose linkage-type isomers. In contrast to other reported methodologies, this GRIL strategy made it possible to reliably characterize glycan isomers by comparing peak areas and retention times, thereby avoiding the lack of reproducibility between runs and variations in the ionization of the glycans due to ion-suppression effects or tuning of the mass spectrometer. The assignment of sialic acids linked-type in hAGP glycan isomers revealed differences in the  $\alpha$ -3 and  $\alpha$ -6 sialic acid content that can be correlated with changes in the abundance of certain isomers in pathological samples. Furthermore, characterization of fucose linkage-type isomers of the most abundant hAGP glycans confirmed that, in addition to  $\alpha$ -1-3 antennary fucosylation, core fucosylation also appears to be present in several hAGP glycans. The established methodology provides a reliable way of characterizing glycan isomers in order to identify specific biomarkers in many diseases involving protein glycosylation, and in the specific case of hAGP, it provides an insight into the role of this protein in pancreatic cancer.

### Acknowledgements

This work was supported by a grant from the Spanish Ministry of Science and Innovation (CTQ2014-56777-R). Montserrat Mancera-Arteu acknowledges the University of Barcelona for an ADR fellowship.

### Appendix A. Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.aca.2016.07.043>.

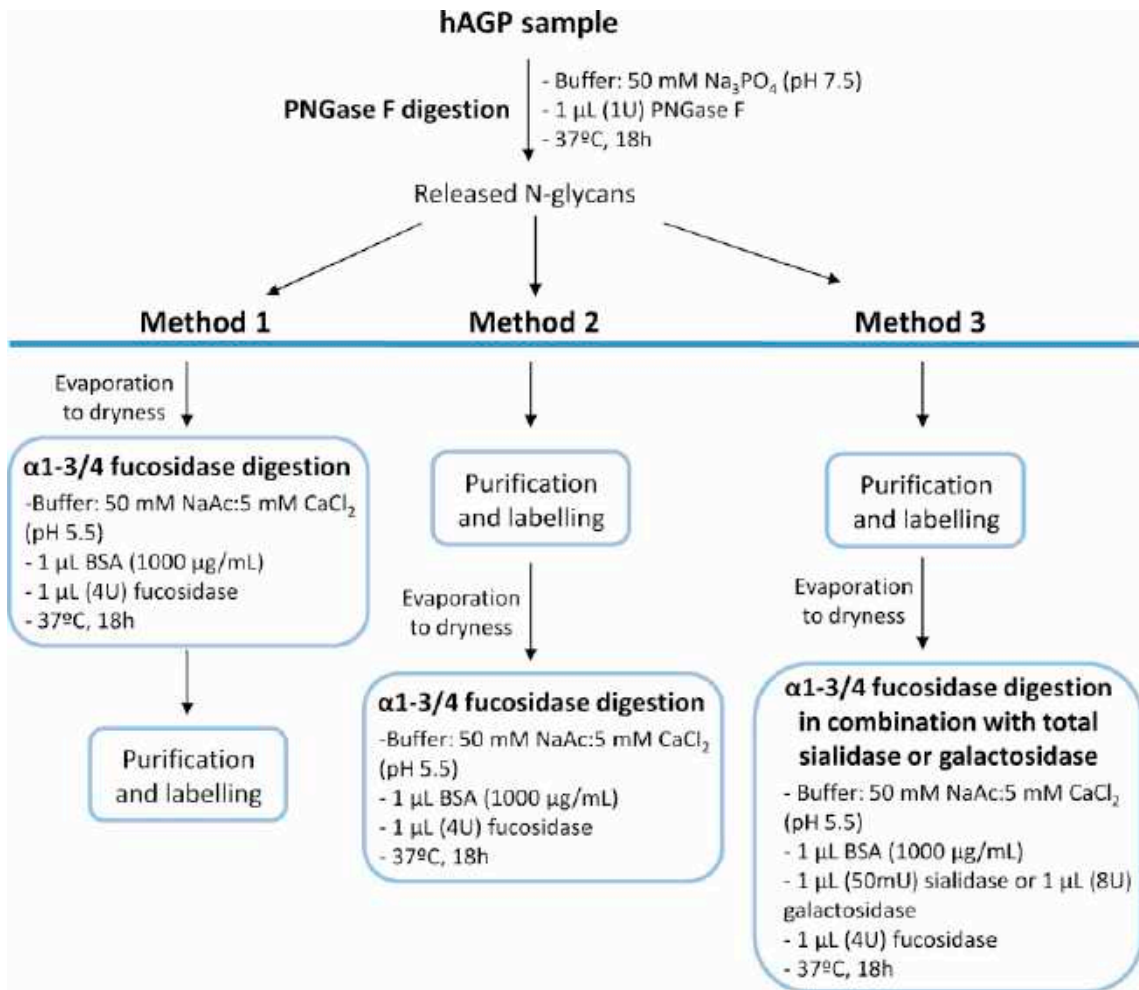
### References

- [1] R. Dwek, *Glycobiology: toward understanding the function of sugars*, Chem. Rev. 96 (1996) 683–720.
- [2] S. Gabius, H.J. Gabius, *Glycosciences. Status and Perspectives*, Chapman Hall, 1997.
- [3] W. Morelle, J.C. Michalski, Analysis of protein glycosylation by mass spectrometry, *Nat. Protoc.* 2 (2007) 1585–1602.
- [4] K. Marino, J. Bones, J.J. Kattia, P.M. Rudd, A systematic approach to protein glycosylation analysis: a path through the maze, *Nat. Chem. Biol.* 6 (2010) 713–723.
- [5] T. Cartwright, R. Schwalbe, Atypical sialylated N-glycan structures are attached to neuronal voltage-gated potassium channels, *Biosci. Rep.* 29 (2009) 301–313.
- [6] M. Nakano, R. Saldanha, A. Gobel, M. Kavallaris, N.H. Packer, Identification of glycan structure alterations on cell membrane proteins in desoxyepithelone B resistant leukemia cells, *Mol. Cell. Proteom.* 10 (2011).
- [7] C. Reis, H. Osorio, L. Silva, C. Gomes, L. David, Alterations in glycosylation as biomarkers for cancer detection, *J. Clin. Pathol.* 63 (2010) 322–329.
- [8] Y. Mechref, Y. Hu, A. Garcia, A. Hussein, Identifying cancer biomarkers by mass spectrometry-based glycomics, *Electrophoresis* 33 (2012) 1755–1767.
- [9] L.R. Ruhaak, S. Miyamoto, C.B. Lebrilla, Developments in the identification of glycan biomarkers for the detection of cancer, *Mol. Cell. Proteom.* 12 (2013) 846–855.
- [10] A. Lee, J.M. Chick, D. Kolarich, P. Haynes, G.R. Robertson, M. Tsoli, I. Janikova, S.J. Clarke, N.H. Packer, M.S. Baker, Liver membrane proteome glycosylation changes in mice bearing an extra-hepatic tumor, *Mol. Cell. Proteom.* 10 (2011).
- [11] M. Hedlund, E. Ng, A. Varki, N.M. Varki, 2-6 Linked sialic acids on N-glycans modulate carcinoma differentiation in vivo, *Cancer Res.* 68 (2008) 388–394.
- [12] A. Sarrats, R. Saldova, E. Pla, E. Fort, D.J. Harvey, W.B. Struwe, L.R. De, P.M. Rudd, R. Peracaula, Glycosylation of liver acute-phase proteins in pancreatic cancer and chronic pancreatitis, *Proteom. Clin. Appl.* 4 (2010) 432–448.
- [13] V. Padler-Karavani, Aiming at the sweet side of cancer: aberrant glycosylation as possible target for personalized-medicine, *Cancer Lett.* 352 (2014) 102–112.
- [14] S. Singh, K. Pal, J. Yadav, H. Tang, K. Partyka, D. Kletter, P. Hsueh, E. Ensink, B. KC, G. Hostetter, H.E. Xu, M. Bern, D.F. Smith, A.S. Mehta, R. Brand, K. Meicher, B.B. Haab, Upregulation of glycans containing 3' fucose in a subset of pancreatic cancers uncovered using fusion-tagged lectins, *J. Proteom. Res.* 14 (2015) 2594–2605.
- [15] M. Pabst, J. Grass, S. Toegel, E. Liebminger, R. Strasser, F. Altmann, Isomeric analysis of oligomannosidic N-glycans and their dolichol-linked precursors, *Glycobiology* 22 (2011) 389–399.
- [16] F. Toussi, J. Bones, W.S. Hancock, M. Hincapie, Differential chemical derivatization integrated with chromatographic separation for analysis of isomeric sialylated N-glycans: a nano-hydrophilic interaction liquid chromatography-MS platform, *Anal. Chem.* 85 (2013) 8421–8428.
- [17] H. Ito, K. Yamada, K. Deguchi, H. Nakagawa, S. Nishimura, Structural assignment of disialylated biantennary N-glycan isomers derivatized with 2-aminopyridine using negative-ion multistage tandem mass spectral matching, *Rapid Commun. Mass Spectrom.* 21 (2007) 212–218.
- [18] R.B. Parker, J.E. McCombs, J.J. Kohler, Sialidase specificity determined by chemoselective modification of complex sialylated glycans, *ACS Chem. Biol.* 7 (2012) 1509–1514.
- [19] C. Nwosu, H.K. Yau, S. Flecht, Assignment of core versus antenna fucosylation types in protein N-glycosylation via procainamide labeling and tandem mass spectrometry, *Anal. Chem.* 87 (2015) 5905–5913.
- [20] J.L. Abrahams, N.H. Packer, M.P. Campbell, Relative quantitation of multi-antennary N-glycan classes: combining PGC-LC-ESI-MS with exoglycosidase digestion, *Analyst* (2015) 5444–5449.
- [21] K. Deguchi, Y. Takegawa, H. Ito, N. Miura, S. Yoshioka, S. Nagai, H. Nakagawa, S.I. Nishimura, Structural assignment of isomeric 2-aminopyridine-derivatized monosialylated biantennary N-linked oligosaccharides using negative-ion multistage tandem mass spectral matching, *Rapid Commun. Mass Spectrom.* 20 (2006) 412–418.
- [22] A.V. Everest-Dass, J.L. Abrahams, D. Kolarich, N.H. Packer, M.P. Campbell, Structural feature ions for distinguishing N- and O-linked glycan isomers by LC-ESI-IT MS/MS, *J. Am. Soc. Mass Spectrom.* 24 (2013) 895–906.
- [23] D.J. Harvey, Structural determination of N-linked glycans by matrix-assisted laser desorption/ionization and electrospray ionization mass spectrometry, *Proteomics* 5 (2005) 1774–1786.
- [24] L. Mauko, N.A. Lacher, M. Pelzing, A. Nordborg, P.R. Haddad, E.F. Hilder, Comparison of ZIC-HILIC and graphitized carbon-based analytical approaches combined with exoglycosidase digestions for analysis of glycans from monoclonal antibodies, *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* 911 (2012) 93–104.
- [25] M. Pabst, J.S. Bondili, J. Stadlmann, L. Mach, F. Altmann, Mass + retention time = structure: a strategy for the analysis of N-glycans by carbon LC-ESI-MS and its application to fibrin N-glycans, *Anal. Chem.* 79 (2007) 5051–5057.
- [26] L. Royle, M.P. Campbell, C.M. Radcliffe, D.M. White, D.J. Harvey, J.L. Abrahams, Y.G. Kim, G.W. Henry, N.A. Stadlick, M.E. Weinblatt, D.M. Lee, P.M. Rudd, R.A. Dwek, HPLC-based analysis of serum N-glycans on a 96-well plate platform with dedicated database software, *Anal. Biochem.* 376 (2008) 1–12.
- [27] S. Tao, Y. Huang, B.E. Boyes, R. Orlando, Liquid chromatography-selected reaction monitoring (LC-SRM) approach for the separation and quantitation of sialylated N-Glycans linkage isomers, *Anal. Chem.* 86 (2014) 10584–10590.
- [28] J.A. Atwood, L. Cheng, G. Alvarez-Manilla, N.L. Warren, W.S. York, R. Orlando, Quantitation by isobaric labeling: applications to glycomics, *J. Proteom. Res.* 7 (2008) 367–374.
- [29] B. Xia, C.L. Feasley, G.P. Sachdev, D.F. Smith, R.D. Cummings, Glycan reductive isotope labeling (GRIL) for quantitative glycomics, *Anal. Biochem.* 387 (2010) 162–170.
- [30] G. Ridlova, J.C. Mortimer, S.L. Maslen, P. Dupree, E. Stephen, Oligosaccharide relative quantitation using isotope tagging and normal-phase liquid chromatography/mass spectrometry, *Rapid Commun. Mass Spectrom.* 22 (2008) 2723–2730.
- [31] E. Giménez, V. Sanz-Nebot, A. Rizzi, Relative quantitation of glycosylation variants by stable isotope labeling of enzymatically released N-glycans using [ $^{12}\text{C}$ ]/[ $^{13}\text{C}$ ] aniline and ZIC-HILIC-ESI-TOF-MS, *Anal. Bioanal. Chem.* 405 (2013) 7307–7319.
- [32] E. Giménez, M. Balmaña, J. Figueras, E. Fort, C. De Bolós, V. Sanz-Nebot, R. Peracaula, A. Rizzi, Quantitative analysis of N-glycans from human alpha-acid-glycoprotein using stable isotope labeling and zwitterionic hydrophilic interaction capillary liquid chromatography electrospray mass spectrometry as tool for pancreatic disease diagnosis, *Anal. Chim. Acta* 866 (2015) 59–68.
- [33] M. Balmaña, E. Giménez, A. Puerta, E. Llop, J. Figueras, E. Fort, V. Sanz-Nebot, C. de Bolós, A. Rizzi, S. Barrabés, M. de Frutos, R. Peracaula, Increased  $\alpha$ -1-3 fucosylation of  $\alpha$ -1-acid glycoprotein (AGP) in pancreatic cancer, *J. Proteom.* 132 (2016) 144–154.
- [34] F. Cecciani, V. Pocacqua, The acute phase protein alpha-1-acid glycoprotein: a model for altered glycosylation during diseases, *Curr. Protein Pept. Sci.* 8 (2007) 91–108.
- [35] K.D. Smith, J. Behan, G. Matthews Smith, A.M. Magliocco, Alpha-1-acid glycoprotein (AGP) as a potential biomarker for breast cancer, *Glycosylation* (2012) 201–222.
- [36] Typical Reaction Conditions for  $\alpha$ -1-3,4-Fucosidase (P0769), <https://www.ncbi.nlm.nih.gov/protocols/2015/07/15/typical-reaction-conditions-for-1-3-4-fucosidase-p0769>.
- [37] P.B. Kunda, F. Benavente, S. Catala-Clariana, E. Giménez, J. Barbosa, V. Sanz-Nebot, Identification of bioactive peptides in a functional yogurt by micro liquid chromatography time-of-flight mass spectrometry assisted by retention time prediction, *J. Chromatogr. A* 1229 (2012) 121–128.
- [38] T. Fournier, N. Medjoubi, D. Porquet, Alpha-1-acid glycoprotein, *Biochim. Biophys. Acta* 1482 (2000) 157–171.

## Supplementary Figures

---





**Supplementary Figure S-1.** Experimental workflow followed to evaluate the digestion of hAGP N-glycans with α1-3,4 fucosidase.





Contents lists available at ScienceDirect

Analytica Chimica Acta

journal homepage: [www.elsevier.com/locate/aca](http://www.elsevier.com/locate/aca)

## Zwitterionic-hydrophilic interaction capillary liquid chromatography coupled to tandem mass spectrometry for the characterization of human alpha-acid-glycoprotein N-glycan isomers



Montserrat Mancera-Arteu<sup>a</sup>, Estela Giménez<sup>a</sup>, José Barbosa<sup>a</sup>, Rosa Peracaula<sup>b</sup>,  
Victòria Sanz-Nebot<sup>a,\*</sup>

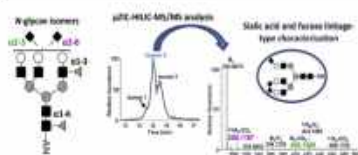
<sup>a</sup> Department of Chemical Engineering and Analytical Chemistry, University of Barcelona, Diagonal 645, E-08028 Barcelona, Spain

<sup>b</sup> Biochemistry and Molecular Biology Unit, Department of Biology, University of Girona, Campus Montilivi s/n, 17071 Girona, Spain

### HIGHLIGHTS

- ZIC-HILIC-MS/MS for a reliable characterization of complex-type N-glycan isomers.
- Determination of MS/MS ion fragments with higher diagnostic value.
- Highly sialylated tri- and tetraantennary glycans were deeply characterized.
- Analysis of serum samples demonstrated the robustness of the methodology.

### GRAPHICAL ABSTRACT



### ARTICLE INFO

#### Article history:

Received 11 April 2017

Received in revised form

27 July 2017

Accepted 28 July 2017

Available online 4 August 2017

#### Keywords:

Isomer

Glycan

AGP

Tandem mass spectrometry

ZIC-HILIC

### ABSTRACT

In this work, a  $\mu$ ZIC-HILIC-MS/MS methodology was established in negative ion mode for the characterization of glycan isomers. The possibility to separate the glycan isomers by the  $\mu$ ZIC-HILIC strategy coupled to a high resolution tandem mass spectrometry detection permitted us to obtain valuable information about each glycan structure. The most important diagnostic ion fragments previously described to characterize structural features of glycans, were evaluated in this study using hAGP as model glycoprotein. The assignation of hAGP glycan isomers performed in our previous work using the GRIL strategy in combination with exoglycosidase digestion [1] was used in this paper to confirm or discard some ion fragments reported in the literature and delve into the structural characterization of glycan isomers. Sialic acid as well as fucose linkage-type glycan isomers were assigned using this approach and daughter ions with higher diagnostic value were determined. The location of  $\alpha$ 2-3/ $\alpha$ 2-6 sialic acids on antennas and a deeper characterization of several highly sialylated tri- and tetraantennary glycans was also possible using the established MS/MS method. Moreover, relying on the characterization performed in Ref. [1], core and antenna fucosylation were differentiated in this work using specific ion fragments obtained in the tandem mass spectra. This methodology was also applied to hAGP purified from control and pathological serum samples, which corroborated its robustness and its potential for finding novel glycan-based biomarkers in patho-glycomic studies.

© 2017 Elsevier B.V. All rights reserved.

### 1. Introduction

Glycans play a fundamental role in the function and

\* Corresponding author.

E-mail address: [vsanz@ub.edu](mailto:vsanz@ub.edu) (V. Sanz-Nebot).



physicochemical properties of the proteins [2–4]. Moreover, modifications in their structure have been related to many diseases such as chronic inflammation, hereditary disorders and cancer [5–10]. Thus, the interest towards the structural analysis of these compounds has risen in recent years. Complex N-linked glycans contain a common trimannosyl-chitobiose core substituted in the two outer mannose residues with antennas, consisting mainly of N-acetylglucosamine-galactose (HexNAc-Hex) units [11,12]. N-acetylneuraminic acids (NeuAc) are commonly found attached to the terminal galactose in an  $\alpha$ 2-3 or an  $\alpha$ 2-6 linkage. Additionally, other residues such as fucose can be linked to the core or antenna HexNAc as well as to the antennary galactose. The different possible linkages of each substituent has led to the existence of isobaric isomers in glycans [13–15]. Nowadays, the in depth characterization of these isomers is of great interest to many authors as overexpression of certain glycan linkage-types were reported in cancer [6,10,16–20].

Exoglycosidase digestion has classically been used for the structural characterization of glycans by monitoring the products of digestion by gel-filtration chromatography, high-performance liquid chromatography (HPLC), matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) [21] and, more recently, liquid chromatography coupled to mass spectrometry detection (LC-MS) [22–24]. Nevertheless, commercially available enzymes can only cleave certain linkage-types and their specificity might not be absolutely guaranteed. MALDI-MS in combination with methyl esterification procedures was another technique extensively used that allows sialic acid linkage-specific derivatization to differentiate between  $\alpha$ 2-3 and  $\alpha$ 2-6 linked sialic acids. However, these methods require highly purified glycan samples, harsh conditions, long reaction times and, moreover, the complete derivatization and linkage specificity are often not achieved [25,26]. Therefore, the use of tandem mass spectrometry approaches to obtain structural information about the sequence and linkage of glycans was promoted [12,13,21,27–35]. In last years, most authors have used MS/MS in negative ion mode as it yields less ambiguous spectra and contains ion fragments that are diagnostic of specific structural features, which cannot be easily obtained from the positive ion mode [21,28–31]. Nevertheless, in many of these studies a liquid chromatography methodology able to separate the different isomers prior to the tandem mass spectrometry detection was not used [21,28,32]. Moreover, several authors did not previously characterize the linkage-types present in each glycan isomer with a complementary technique [29,30,33,34] and, therefore, the specific ion fragments to diagnose certain structural features could not be clearly determined. MS<sup>3</sup> approaches were also used to study fragmentation pathways of glycan structures. Despite the potential of these methods, creating a fingerprint catalogue is only possible for few structures and requires large amounts of purified glycan samples [13,33,35]. For all these reasons, MS<sup>n</sup> fragmentation studies are usually performed with simple structures such as mono- or bisialylated biantennary glycans.

In a previous study, human alpha-1-acid glycoprotein (hAGP) N-glycans were characterized in serum samples from healthy volunteers and from patients with pancreatic cancer (PaC) and chronic pancreatitis (ChrP) by zwitterionic-hydrophilic interaction capillary liquid chromatography coupled to mass spectrometry detection ( $\mu$ ZIC-HILIC-MS), observing an increase of several fucosylated glycans as well as changes in the proportion of certain glycan isomers in PaC samples [36]. These results agreed with those previously described in the literature for carcinomas, including changes in expression of  $\alpha$ 2-3/6 linked sialic acids and over-expression of sialyl Lewis X (SLe<sup>x</sup>) and sialyl Lewis A (SLe<sup>a</sup>) epitopes [16–20,37–39]. As main N-glycan modifications reported in cancer are associated with variations in sialylation and fucosylation,

deepening into the characterization of sialic acid and fucose linkage-type isomers is of crucial importance to find new glycan-based biomarkers. In a recent study, we developed a robust methodology to reliably identify the linkage-type of sialic acids and fucoses in hAGP N-glycan isomers using a stable isotope labelling strategy with [<sup>12</sup>C<sub>6</sub>]/[<sup>13</sup>C<sub>6</sub>] aniline in combination with exoglycosidase digestion [1]. Bi-, tri- and tetraantennary glycan isomers differing in the sialic acid linkage-type were assigned and, moreover,  $\alpha$ 1-3 antenna and  $\alpha$ 1-6 core fucosylations were both detected in hAGP glycans.

In this work, a  $\mu$ ZIC-HILIC-MS/MS method in negative ion mode has been developed using hAGP as model glycoprotein in order to determine reliable diagnostic fragments useful for sialic acid and fucose linkage-type assignments. The exact mass of the daughter ions obtained using a LTQ-Orbitrap mass spectrometer, together with the characterization of hAGP glycan isomers performed in our previous study [1] have allowed us to confirm or discard some diagnostic ion fragments proposed in the literature. Furthermore, the fragments established with a higher diagnostic character have been used to delve into the structure and linkage-type characterization of some highly sialylated tri- and tetraantennary hAGP glycans that were not completely assigned with exoglycosidase digestion. Finally, this methodology was also applied to control and pathological serum samples from patients with pancreatic diseases to corroborate its robustness.

## 2. Materials and methods

### 2.1. Chemicals

All chemicals used in the preparation of buffers and solutions were of analytical reagent grade. Acetic acid (HAc, glacial), formic acid (FA 98–100%), dimethylsulphoxide (DMSO) and acetone were supplied by Merck (Darmstadt, Germany). Sodium phosphate dodecahydrate (Na<sub>3</sub>PO<sub>4</sub>·12H<sub>2</sub>O), sodium cyanoborohydride (NaBH<sub>3</sub>CN), [<sup>12</sup>C<sub>6</sub>]-aniline ([<sup>12</sup>C<sub>6</sub>]AN), 2-mercaptoethanol ( $\beta$ -ME) and sodium dodecyl sulfate (SDS) were purchased from Sigma-Aldrich (St. Louis, MO, USA). Hydrochloric acid (HCl, 37%) was supplied by Panreac (Barcelona, Spain) and “NP-40 alternative” by Calbiochem (Darmstadt, Germany). Ammonium acetate (NH<sub>4</sub>Ac), acetonitrile (ACN) and water LC-MS quality grade from Merck and Fluka (Madrid, Spain), respectively, were used for  $\mu$ ZIC-HILIC-TOF-MS analysis. A standard sample of human  $\alpha$ 1-acid-glycoprotein (hAGP, 99%) was purchased from Sigma-Aldrich. Peptide N-glycosidase F (PNGase F) and  $\alpha$ 2-3,6,8 neuraminidase (total sialidase) were obtained from Roche Diagnostics (Basel, Switzerland) and,  $\alpha$ 2-3 neuraminidase S ( $\alpha$ 2-3 sialidase) from New England Biolabs (Ipswich, MA, USA).

### 2.2. hAGP glycan samples

Control and pathological human serum samples were provided by Hospital Universitari Dr. Josep Trueta, Girona, Spain, following the standard procedures of its Ethics Committee. Three serum samples were analyzed: a healthy volunteer serum sample as control and two pathological serum samples, one from an individual with chronic pancreatitis (ChrP) and one from a patient suffering from pancreatic cancer (PaC) with an advanced stage (stage IV). 100  $\mu$ L of each control and pathological serum samples were purified by immunoaffinity chromatography (IAC) using an anti-hAGP column and following the method described in Ref. [36]. After IAC purification, eluted hAGP was quantified at 280 nm using a Nanodrop spectrophotometer (Thermo Scientific) following the recommendations of the manufacturer for hAGP standard. Then, hAGP samples were evaporated to dryness by Speed Vac and stored

at  $-20^{\circ}\text{C}$ .

5–10  $\mu\text{g}$  of hAGP (standard or purified from serum samples) were reduced with 0.5%  $\beta$ -ME in the presence of 0.5% of SDS in 50 mM  $\text{Na}_3\text{PO}_4$  (pH 7.5) and boiled in a water bath for 30 min. When samples were at room temperature, a volume of 50 mM  $\text{Na}_3\text{PO}_4$  (pH 7.5) with 1% (v/v) of NP-40 alternative was added to achieve a final concentration of 0.1% of SDS and  $\beta$ -ME in the samples. To release the *N*-glycans, 1  $\mu\text{L}$  of PNGase F (1 U) solution was added and the mixture was carefully vortexed and incubated at  $37^{\circ}\text{C}$  for 18 h. Digestion was stopped by heating in a thermo-block at  $100^{\circ}\text{C}$  for 15 min and sample was stored at  $-20^{\circ}\text{C}$  until its use. Released *N*-glycans were purified by solid phase extraction (SPE) using Hypercarb cartridges (25 mg, 1 mL volume, Thermo Fisher Scientific). Firstly, SPE cartridges were conditioned and equilibrated with 1 mL of 60% ACN, 0.1% FA and with 2 mL of water, respectively. After dissolving the digested glycoprotein sample in  $\sim 500$   $\mu\text{L}$  of water, it was loaded and the SPE cartridge was rinsed with 1 mL of water. Retained *N*-glycans were eluted with 600  $\mu\text{L}$  of 60% ACN, 0.1% FA and the eluate was evaporated to dryness by Speed Vac. Dried *N*-glycans were stored at  $-20^{\circ}\text{C}$  until used.

Glycan labelling was carried out using a reaction mixture of 0.35 M aniline and 1 M  $\text{NaBH}_3\text{CN}$  in DMSO with 30% HAC. Dried glycans were mixed with 10  $\mu\text{L}$  of the reaction mixture and incubated at  $70^{\circ}\text{C}$  for 2 h in a water bath. Samples were cooled to room temperature and labelled glycans were precipitated with acetone as described in Ref. [40]. Finally, samples were evaporated to dryness by Speed Vac to remove the excess of acetone, and dried glycans were stored at  $-20^{\circ}\text{C}$  until their analysis. Centrifugations were performed in a Mikro 220R centrifuge (Hettich Zentrifugen, Tuttlingen, Germany).

### 2.3. Sialidase digestion of hAGP *N*-glycans

Sialic acids were released from hAGP *N*-glycans by enzymatic digestion with  $\alpha$ 2-3,6,8 sialidase (total sialidase) or  $\alpha$ 2-3 sialidase. Sialidase digestion was performed after PNGase F treatment in the same buffer solution of 50 mM  $\text{Na}_3\text{PO}_4$  (pH 7.5) by adding 1  $\mu\text{L}$  of total sialidase (50 mU) or 6.25  $\mu\text{L}$  of  $\alpha$ 2-3 sialidase (50 U). Subsequently, sample solution was incubated at  $37^{\circ}\text{C}$  for 1 h ( $\alpha$ 2-3 sialidase) or 18 h (total sialidase). Digestions were stopped by heating in a thermo-block for 10 min at  $100^{\circ}\text{C}$ , and samples were purified and labelled with aniline following the protocol described in section 2.2. Samples were evaporated to dryness by Speed Vac and stored at  $-20^{\circ}\text{C}$  until their analysis.

### 2.4. $\mu\text{ZIC-HILIC-MS/MS}$

$\mu\text{LC-MS/MS}$  experiments were performed in a 1200 series capillary liquid chromatography system coupled to a QTrap 6500 (AB Sciex) or a LTQ Orbitrap Velos (Thermo Scientific). Separation was carried out using a ZIC-HILIC column (150  $\times$  0.3 mm, 3.5  $\mu\text{m}$ , SeQuant, Umeå, Sweden). Experiments were performed at room temperature with gradient elution at a flow rate of 4  $\mu\text{L}/\text{min}$  and injecting 0.25  $\mu\text{L}$  of glycan sample labelled with aniline (glycan concentration:  $\sim 50$ – $100$  pmol/ $\mu\text{L}$ ). Eluting solvents were A: 1 mM  $\text{NH}_4\text{Ac}$  solution and B: acetonitrile. Gradient conditions were used: solvent B from 90 to 80% (within 5 min) and from 80% to 65% (within 20 min) as linear gradient, followed by cleaning and equilibration steps of B: 65  $\rightarrow$  50% (within 5 min), 50  $\rightarrow$  0% (within 5 min), 0% (over 15 min), 0  $\rightarrow$  90% (within 5 min) and 90% (over 10 min). Tuning and calibration of the mass spectrometers followed the respective manufacturer's instructions. The MS/MS experiments were performed in the negative ion mode.

In ESI-QTrap MS/MS experiments, the optimum established parameters were: curtain gas flow (CUR) 25, ion spray voltage (IS)

4500 V, source temperature  $25^{\circ}\text{C}$ , ion source gas 1 (GS1) 20 and ion source gas 2 (GS2) 0. MS spectra were collected between the mass range 600–1800  $m/z$  and MS/MS spectra between 100 and 2000  $m/z$ . An information dependent acquisition (IDA) method was used with seven different experiments: Enhanced MS (EMS), Enhanced resolution (ER) and five Enhanced product ion scans (EPI). Each EPI experiment selected automatically six precursor ions defined previously in an inclusion list with an intensity threshold of 500 and an exclusion time window of 20 s. The collision energy (CE) was set to 75% after optimization.

In ESI-LTQ-Orbitrap MS/MS experiments, the best operational conditions were: spray voltage 3 kV, capillary temperature  $275^{\circ}\text{C}$ , sheath gas flow 20, aux gas flow 10, sweep gas flow 2 and s-lens RF level 67%. MS spectra, in this case, were obtained between 100 and 2000  $m/z$ , while the MS/MS spectra were acquired between  $m/z$  150 and 2000. Fragmentations obtained using collision induced dissociation (CID) and higher-energy collision dissociation (HCD) were compared. Finally, fragmentation experiments were performed using the HCD cell with an optimized collision energy of 70–75%. Six precursor ions, defined in the method, were automatically selected with an isolation window of 1.0  $m/z$ .

Analyst 1.6.2 and Peak View 2.2 softwares (AB Sciex) were used for data acquisition and data analysis, respectively, in the ESI-QTrap instrument. Xcalibur 2.2 software (Thermo scientific) was used for ESI-LTQ-Orbitrap instrument control and data analysis. The  $\mu\text{LC}$  system was controlled by the ChemStation B.01.03 software (Agilent Technologies). Interpretation and manual annotation of ion fragment exact masses were performed using the SimGlycan 5 software (Premier Biosoft).

## 3. Results and discussion

In the last years, some authors have proposed the use of diagnostic ion fragments for the characterization of glycan isomers. However, some of these studies were performed with low mass accuracy spectrometers [29] or they did not previously separate the different isomers by liquid chromatography [21,28,32]. Moreover, some authors could not be completely sure about the diagnostic value of the selected ion fragments because they did not have enough previous information about the linkage-type composition of each glycan isomer [29,30,33,34], and most of them did not study MS/MS fragmentation of tri- and tetraantennary sialylated complex-type glycans. hAGP glycan isomers were characterized by stable isotope labelling with [ $^{12}\text{C}_6$ ]/[ $^{13}\text{C}_6$ ] aniline in combination with exoglycosidase digestion in our previous work [1], providing valuable information to confirm or discard diagnostic fragments for the identification of glycan isomers by tandem mass spectrometry. In this work, hAGP was used as model glycoprotein to establish a reliable  $\mu\text{ZIC-HILIC-MS/MS}$  methodology but also to delve into the structural characterization of its glycan isomers.

Two mass spectrometers were used in this study, i.e. a QTrap and a LTQ-Orbitrap. As can be observed in Figs. S1A and S1B for H5N4S2 glycan (isomer 2), similar ion fragments were obtained with both mass spectrometers in the analysis of hAGP glycans. Notwithstanding, the possibility to obtain the exact mass of daughter ions, the higher scan speed that permitted us to obtain more MS/MS scans for each chromatographic peak and the higher-energy collisions produced made LTQ-Orbitrap results more reliable. HCD cell was selected as more daughter ions, with high diagnostic value such as cross-ring fragment ions, were detected compared to CID fragmentation. Eight HCD energy values were evaluated ranging from 35 to 95%. Values below 55% did not produce fragmentation of the parent ion, while energy values from 85% caused the loss of A- and D-ions. By way of an example, Fig. S1 compares the tandem mass spectra of H5N4S2 glycan at 55% of

collision energy (Fig. S1C) with the mass spectra of the same glycan at the optimal value established at 75% (Fig. S1B). The ion fragments nomenclature used in this work was the proposed by Domon and Costello [41] and afterwards modified by Harvey and coworkers [31,32]. A compilation of most important characteristic ion fragments described in the literature for the characterization of complex-type *N*-glycans is summarized in Table S1 (supplementary material). Furthermore, ion fragments considered in the literature as a diagnostic of specific structural features (e.g. sialic acid or fucose linkage-type) are listed in Table S2. In both tables, we also enclosed those fragments generated after glycan desialylation. Moreover, it is important to remark that although some ion fragments were considered by several authors as specific of one structure [12,13,29,30], in these tables we included different possible fragmentations that correspond to the same ion fragment exact mass (e.g. ion fragment at  $m/z$  364.1243 could also correspond to  $B_3/Y_6$ , not only to  ${}^{3,5}A_2$  as suggested by Ref. [28]).

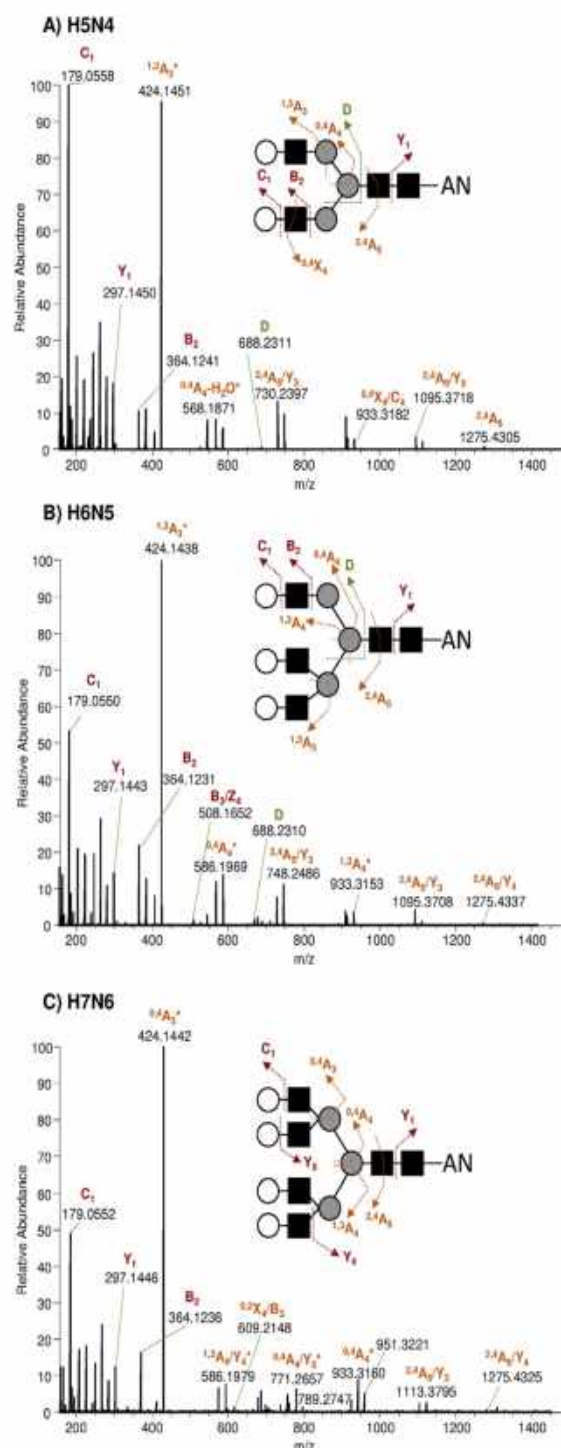
### 3.1. Ion fragments for the characterization of complex-type *N*-glycans

Prior to the characterization of the glycans in terms of sialic acid and fucose linkage-type isomers, desialylation with  $\alpha 2$ -3,6,8 sialidase (total sialidase) permitted to reveal easily information about antenna sequence and composition of complex-type glycans. In this regard, as can be observed in Fig. 1A for the desialylated biantennary glycan of hAGP (H5N4), B- and C-ions at  $m/z$  179.1 (C<sub>1</sub>) and 364.1 (B<sub>2</sub>) as well as several cross-ring fragments such as 424.1 ( ${}^{1,3}A_3$ ) were abundant and provided sequencing characterization of antennas. For higher complex structures such as tri- and tetraantennary desialylated glycans (H6N5 and H7N6, respectively), some of the specific ions listed in Table 1 revealed the existence of a branched antenna. The observed fragments that disclosed this feature were 508.2 for H6N5 glycan and 609.2, 771.3 and 789.3 for H7N6 glycan as well as the ions at  $m/z$  951.3 and 933.3 for both glycan structures (see Fig. 1B and C). Moreover, as was previously reported in many studies [12,21,29,30], D-ions provided information about the 6-antenna and were diagnostic of its composition. In this study, D and [D-H<sub>2</sub>O]<sup>+</sup> ions at  $m/z$  688.2 and 670.2 were used to identify the position of the unbranched antenna and were detected in the tandem mass spectra of H5N4 and H6N5 glycans while they were not present in H7N6 (see Table 1 and Fig. 1). The existence of these D-ions in the case of the H6N5 revealed that, in hAGP, the branched antenna is located on the 3-antenna. These results were also corroborated later for the sialylated glycans by observing D-ions at  $m/z$  979.3 and 961.3 (see Table 2). Even though it was expected to observe D-ions at  $m/z$  1053.3 or 1035.3 for the H7N6 glycan to confirm the branching also on the 6-antenna, only the absence of D-ions at  $m/z$  688.2 and 670.2 were observed in this case.

On the other hand, some authors have reported E-ions as a diagnostic of the 3-antenna composition at  $m/z$  466.1 or 831.3, in glycans with the branching located on the 3-antenna [12,21]. Nevertheless, in this study, the E-ion 466.1 was also observed in tri- and tetraantennary glycans showing its lack of specificity to unbranched 3-antenna. Furthermore, E-ions correspond to other structures, such as  ${}^{0,4}X_3/B_3$  or  ${}^{2,4}X_3/B_3$ , rather than to the  ${}^{0,4}A_3$  structure previously described [12]. For this reason, E-ions were not considered in this work as specifics of 3-antenna composition. Thus, the branched 3-antenna structure of the H6N5 glycan was concluded using only the D-ions observed.

### 3.2. Characterization of sialic acid linkage-type *N*-glycan isomers

In a previous work, a methodology to identify sialic acid linkage-

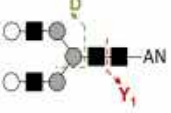
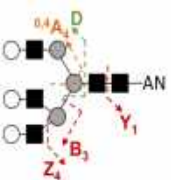
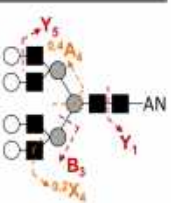
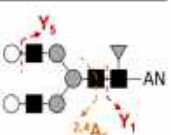
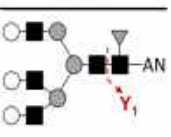
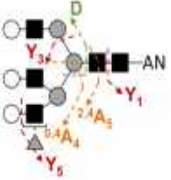


**Fig. 1.** Tandem mass spectra of the biantennary (A), triantennary (B) and tetraantennary (C) desialylated *N*-glycans obtained from hAGP standard after total sialidase digestion and labelled with [<sup>12</sup>C<sub>6</sub>]-aniline. (\*) This ion fragment could correspond to other possible structures listed in Table S1 or S2 of the supplementary material.

type glycan isomers in hAGP glycans was established combining total and  $\alpha 2$ -3 sialidase digestions with glycan reductive isotope labelling (GRIL) using [<sup>12</sup>C<sub>6</sub>]/[<sup>13</sup>C<sub>6</sub>] aniline [1]. Although some authors have pointed out problems with the specificity of enzymes to cleavage one single linkage-type, the combination of exoglycosidase digestions with the GRIL methodology enabled a reliable

**Table 1**

Characteristic ion fragments observed in the tandem mass spectra and used to characterize aniline-labelled desialylated hAGP N-glycan isomers. The nomenclature, assignment and structure of each ion fragment are also included.

| Glycan | Isomer         | Ion fragments observed ( $m/z$ , $m/z - H_2O$ )                  | Nomenclature  | Diagnostic  | Structure   |
|--------|----------------|--|---|---|---|
| H5N4   | Isomer 1       | 297.1450<br>688.2311, 670.2173                                   | $Y_1$<br>D  | Absence of core fucosylation<br>Unbranched 6-antenna  |    |
| H6N5   | Isomer 1       | 297.1443<br>508.1652<br>688.2310, 670.2164<br>951.3377, 933.3153 | $Y_1$<br>$B_3/Z_4$<br>D<br>$^{0,4}A_4^+$                              | Absence of core fucosylation<br>Branched antenna<br>Unbranched 6-antenna<br>Branched antenna                        |    |
| H7N6   | Isomer 1       | 297.1446<br>609.2148<br>789.2747, 771.2657<br>951.3221, 933.3160 | $Y_1$<br>$^{0,2}X_4/B_3$<br>$^{0,4}A_4/Y_5^+$<br>$^{0,4}A_4^+$        | Absence of core fucosylation<br>Branched antenna<br>Branched antenna<br>Branched antenna                            |    |
| H5N4F1 | Isomer 1 and 2 | 461.2251, 443.2035<br>1095.3708                                  | $Y_1$<br>$^{2,4}A_5/Y_5^+$  | Core fucosylation<br>Absence of antenna fucosylation  |   |
| H6N5F1 | Isomer 1 and 2 | 443.2040 <sup>b</sup>  | $Y_1$   | Core fucosylation   |  |
|        | Isomer 3       | 297.1441<br>570.2033<br>670.2611<br>933.3098<br>1095.3689        | $Y_1$<br>$^{0,4}A_4/Y_5^+$<br>D<br>$^{0,4}A_4^+$<br>$^{2,4}A_5/Y_3^+$ | Absence of core fucosylation<br>Fucosylated antenna<br>Unbranched 6-antenna<br>Branched antenna<br>Branched antenna |  |

<sup>a</sup> This ion fragment could correspond to other possible structures listed in Table 1 or 2 of the supplementary material.

<sup>b</sup> Detected only in half of the runs.

assignment of sialic acid linkages. Hence, hAGP glycans were used in this study to evaluate certain ion fragments for the characterization of sialic acid linkage-type isomers by tandem mass spectrometry.

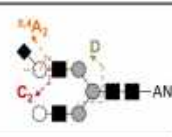
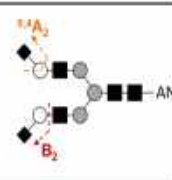
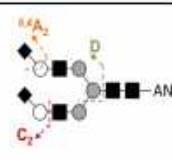
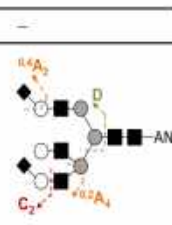
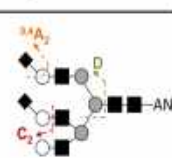
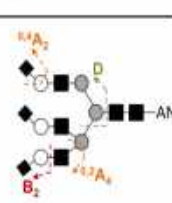
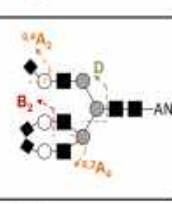
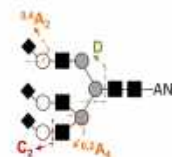
For this purpose, hAGP glycan isomers were analyzed by ZIC-HILIC-MS/MS and, moreover, some glycans were also digested with  $\alpha$ -2-3 sialidase in order to completely characterize the greatest number of isomers. The ion fragments 306.1, 408.1 and 470.1 proposed by Wheeler et al. [28] to assign sialic acid linkage-type were also detected in hAGP glycans. Nevertheless, ions 350.1 and 364.1 described afterwards for the same purpose [30], were not considered by us as several structures were found to produce the same exact mass, showing their lack of specificity (e.g.  $^{2,4}A_2$  or  $^{1,3}A_2$  fragments that could be specific of  $\alpha$ -2-3 NeuAc show the same exact mass of 350.1087 as  $^{0,4}A_2$  that could be specific of  $\alpha$ -2-6 NeuAc, see Supplemental Table S2). Moreover, the ion fragment 364.1 was

also observed in the tandem mass spectra of the desialylated N-glycans demonstrating its low diagnostic value for  $\alpha$ -2-6 NeuAc (see Fig. 1).

Although the native H5N4S2 glycan of hAGP showed two isomers, after  $\alpha$ -2-3 sialidase digestion the first isomer disappeared and turned into the H5N4S1, increasing its native peak area, as was described in more detail in our previous work [1]. Hence, isomer 2 of H5N4S2 and H5N4S1 glycan showed all sialic acids linked  $\alpha$ -2-6. Fig. 2A and B shows the tandem mass spectra of isomers 1 and 2 of the native H5N4S2 glycan, and the H5N4S1 glycan obtained after  $\alpha$ -2-3 sialidase digestion, respectively. As can be observed, similar MS/MS spectra were obtained for isomer 2 of H5N4S2 and H5N4S1 glycans (Fig. 2A(ii) and 2B(i), respectively). In both cases, an abundant ion at  $m/z$  306.1 and the ion fragment 470.1 were observed, which have been previously reported as characteristic of  $\alpha$ -2-6 linked sialic acids [28]. In addition, the absence of the ion

**Table 2**

Diagnostic ion fragments observed in the tandem mass spectra and used to characterize aniline-labelled native and  $\alpha$ 2-3 desialylated hAGP N-glycan isomers. The nomenclature, assignment and structure of each ion fragment are also included.

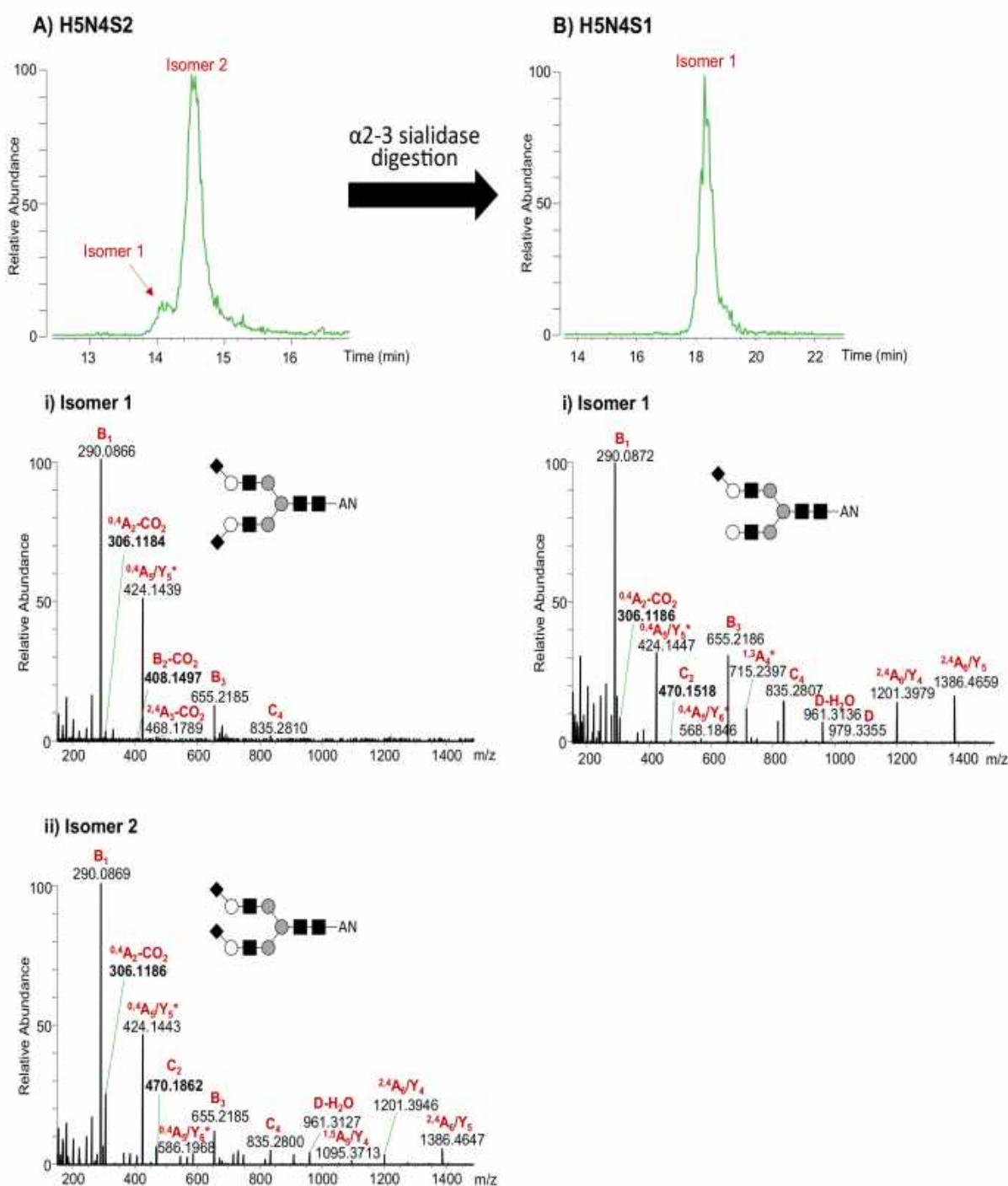
| Glycan   | Isomer                | Ion fragments observed ( $m/z$ , $m/z - H_2O$ ) | Nomenclature              | Diagnostic                      | Structure   |
|----------|-----------------------|---|---------------------------|---------------------------------|---|
| H5N4S1   | Isomer 1              | 306.1186  | $^{0,4}A_2-CO_2$          | $\alpha$ 2-6 linked NeuAc       |    |
|          |                       | 470.1518  | $C_2$                     | $\alpha$ 2-6 linked NeuAc       |   |
|          |                       | 979.3355, 961.3136                              | D                         | Sialylated unbranched 6-antenna |   |
| H5N4S2   | Isomer 1              | 306.1184  | $^{0,4}A_2-CO_2$          | $\alpha$ 2-6 linked NeuAc       |    |
|          |                       | 408.1497  | $B_2-CO_2$                | $\alpha$ 2-3 linked NeuAc       |   |
|          | Isomer 2              | 306.1186  | $^{0,4}A_2-CO_2$          | $\alpha$ 2-6 linked NeuAc       |    |
|          |                       | 470.1862  | $C_2$                     | $\alpha$ 2-6 linked NeuAc       |   |
|          |                       | 979.3333, 961.3127                              | D                         | Sialylated unbranched 6-antenna |   |
| H6N5S2   | Isomer 1 <sup>a</sup> | –   | –                         | –                               |   |
|          |                       | Isomer 2  | 306.1186                  | $^{0,4}A_2-CO_2$                |   |
|          | 470.1498              |   | $C_2$                     | $\alpha$ 2-6 linked NeuAc       |   |
|          | 775.2702              |   | $^{0,2}A_4$               | Sialylated 4-branching          |   |
|          | 979.3119, 961.3134    |   | D                         | Sialylated unbranched 6-antenna |   |
|          | Isomer 3 <sup>b</sup> | 306.1186  | $^{0,4}A_2-CO_2$          | $\alpha$ 2-6 linked NeuAc       |  |
| 470.1505 |                       | $C_2$   | $\alpha$ 2-6 linked NeuAc |                                 |   |
|          |                       | 979.3166, 961.3119                              | D                         | Sialylated unbranched 6-antenna |   |
| H6N5S3   | Isomer 1              | 306.1186  | $^{0,4}A_2-CO_2$          | $\alpha$ 2-6 linked NeuAc       |  |
|          |                       | 408.1492  | $B_2-CO_2$                | $\alpha$ 2-3 linked NeuAc       |   |
|          |                       | 775.2560  | $^{0,2}A_4$               | Sialylated 4-branching          |   |
|          |                       | 961.3204  | D                         | Sialylated unbranched 6-antenna |   |
|          | Isomer 2              | 306.1187  | $^{0,4}A_2-CO_2$          | $\alpha$ 2-6 linked NeuAc       |  |
|          |                       | 408.1546  | $B_2-CO_2$                | $\alpha$ 2-3 linked NeuAc       |   |
|          |                       | 775.2590  | $^{0,2}A_4$               | Sialylated 4-branching          |   |
|          |                       | 961.3251  | D                         | Sialylated unbranched 6-antenna |   |
|          | Isomer 3              | 306.1187  | $^{0,4}A_2-CO_2$          | $\alpha$ 2-6 linked NeuAc       |  |
|          |                       | 470.1503  | $C_2$                     | $\alpha$ 2-6 linked NeuAc       |   |
|          |                       | 775.2588  | $^{0,2}A_4$               | Sialylated 4-branching          |   |
|          |                       | 961.3140  | D                         | Sialylated unbranched 6-antenna |   |

<sup>a</sup> Glycan isomer not studied in this work.

<sup>b</sup> Glycan isomer not present in the native protein, detected only after  $\alpha$ 2-3 sialidase digestion.

fragment 408.1 in these glycans confirmed these three ion fragments (306.1, 408.1 and 470.1) as diagnostic for sialic acid linkage-type assignment. Moreover, the presence of D and D-H<sub>2</sub>O ions at  $m/z$  979.3 and 961.3 in H5N4S1 glycan revealed the localization of the

$\alpha$ 2-6 NeuAc on the 6-antenna (see Table 2 and Fig. 2B). Therefore, the first isomer of the H5N4S2 glycan, which turned into H5N4S1 glycan after  $\alpha$ 2-3 sialidase digestion, showed one  $\alpha$ 2-3 NeuAc located on the 3-antenna and one  $\alpha$ 2-6 NeuAc located on the 6-

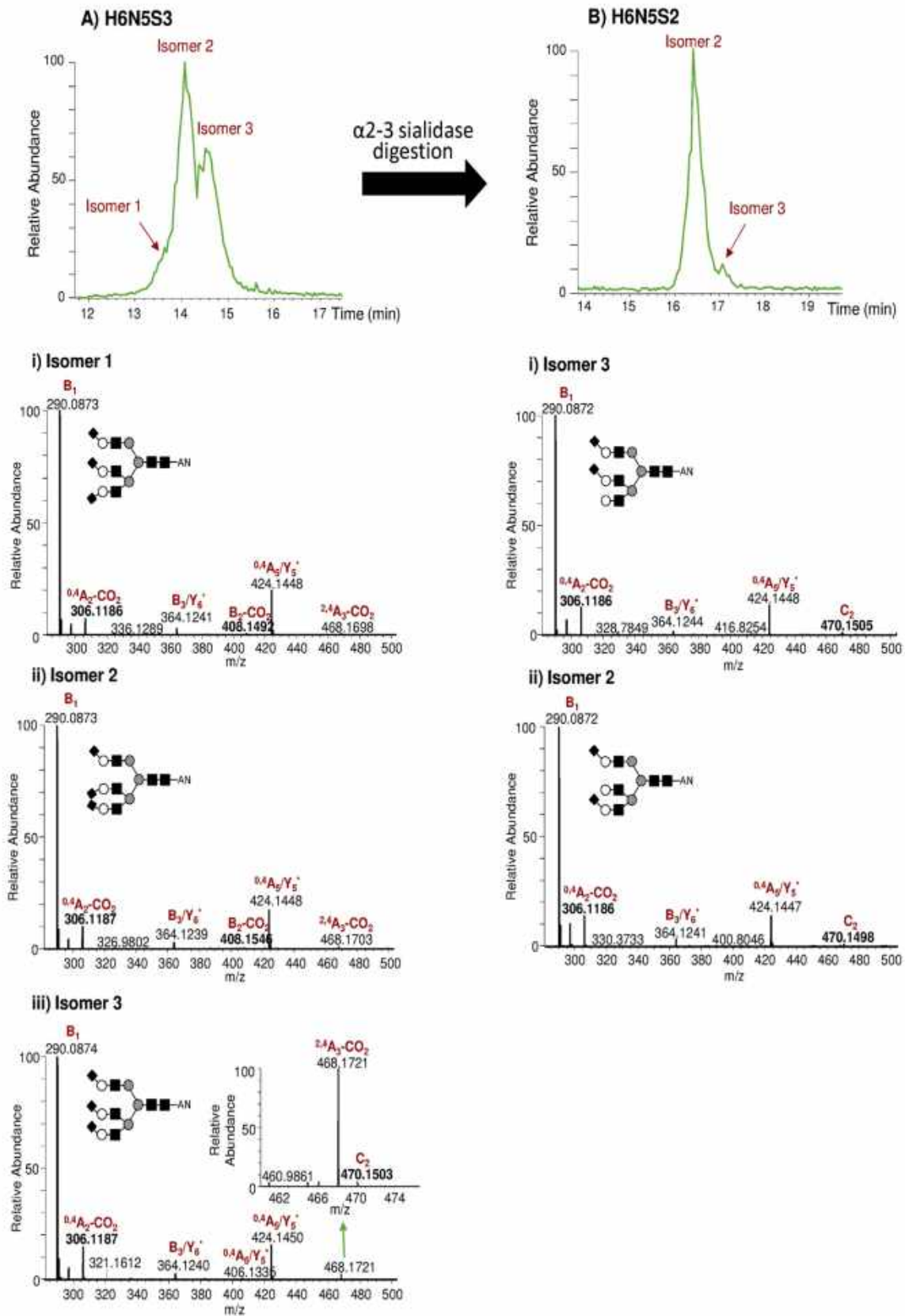


**Fig. 2.** (A) Extracted ion chromatogram (EIC) of the native H5N4S2 glycan obtained from hAGP standard and the MS/MS spectra obtained for each isomer (isomer 1 (i) and isomer 2 (ii)). (B) EIC of the H5N4S1 glycan, obtained from hAGP standard after  $\alpha$ 2-3 sialidase digestion, and its corresponding MS/MS spectrum (i). (\*) This ion fragment could correspond to other possible structures listed in Table S1 or S2 of the supplementary material.

antenna. This assignment was also corroborated by observing the ion fragments 306.1 and 408.1 in the MS/MS spectra of the native isomer 1, despite the low abundance of these ion fragments as a consequence of its coelution with the more abundant H6N5S3 glycan (see Table 2 and Fig. 2A(i)).

Regarding the native trisialylated triantennary glycan (H6N5S3), which showed three isomers, ion fragments diagnostic of  $\alpha$ 2-6 and  $\alpha$ 2-3 NeuAc were observed in isomer 1 and 2 ( $m/z$  306.1 and 408.1, see Table 2 and Fig. 3A (i) and (ii)). In contrast, isomer 3 only

presented  $\alpha$ 2-6 NeuAc fragments ( $m/z$  306.1 and 470.1, see Table 2 and Fig. 3A (iii)), showing higher relative intensity of 306.1 ion in this isomer compared to the others. These diagnostic ion fragments confirmed our previous assignment [1]. Isomers 1 and 2 presented both  $\alpha$ 2-3 and  $\alpha$ 2-6 sialic acid linkage-types and only the third isomer showed all sialic acids  $\alpha$ 2-6 linked. Additionally, some authors have pointed out that  $\alpha$ 2-3 NeuAc are more labile than  $\alpha$ 2-6 ones and they have described the relative intensity between ions 290.1 (NeuAc) and 655.2 (NeuAc-Hex-HexNAc) as a parameter to



**Fig. 3.** (A) Extracted ion chromatogram (EIC) of the native H6N5S3 glycan obtained from hAGP standard and the corresponding MS/MS spectra of each isomer (isomer 1 (i), isomer 2 (ii) and isomer 3 (iii)). (B) EIC of the H6N5S2 glycan, obtained from hAGP standard after  $\alpha 2$ -3 sialidase digestion, and the corresponding MS/MS spectra for each isomer (isomer 3 (i) and isomer 2 (ii)). (\*) This ion fragment could correspond to other possible structures listed in Table S1 or S2 of the supplementary material.

evaluate the linkage-type composition in each glycan isomer [13,29]. In our case, similar relative intensities were obtained for all isomers of H6N5S3 glycan, which demonstrated this relation is independent of the proportion of  $\alpha$ 2-3 and  $\alpha$ 2-6 linkages. On the other hand,  $[D-H_2O]^-$  ions at  $m/z$  961.3 were also observed in the three isomers (see Table 2), corroborating that H6N5S3 glycan was branched on the 3-antenna.

With respect to the native H6N5S2 glycan, two isomers were detected but, after  $\alpha$ 2-3 sialidase treatment, the first isomer disappeared, the peak area of the second isomer increased considerably and a third isomer appeared. As was justified in Ref. [1], after  $\alpha$ 2-3 sialidase digestion, isomers 1 and 2 of H6N5S3 glycan turned into isomers 3 and 2 of H6N5S2 glycan, respectively, and only the third isomer remained. Hence, isomers 1 and 2 of H6N5S3 showed two  $\alpha$ 2-6 NeuAc and one  $\alpha$ 2-3 NeuAc, while isomer 3 presented the three sialic acids  $\alpha$ 2-6 as was discussed before [1]. In this work, the tandem mass spectra of isomers 2 and 3 of H6N5S2 glycan, obtained after  $\alpha$ 2-3 sialidase digestion, were studied in order to identify the structural differences between them, which could not be determined previously with exoglycosidase digestion. Fig. 3B (i) and (ii) show the tandem mass spectra obtained for both isomers. Only diagnostic fragments of  $\alpha$ 2-6 NeuAc (306.1 and 470.1) with similar relative intensities were detected in both isomers, which corroborated the same NeuAc linkage-type composition. These results were consistent with our previous study with exoglycosidase digestion and GRIL analysis [1]. Furthermore, the MS/MS method provided additional information about the position of these  $\alpha$ 2-6 sialic acids as D-ions detected at  $m/z$  979.3 and 961.3 (see Table 2) confirmed that in both isomers one NeuAc was located on the unbranched 6-antenna, and consequently, the other one was on the branched 3-antenna. Thus, the difference among isomer 2 and 3 had to be associated with the different localization of the  $\alpha$ 2-6 NeuAc on the branched 3-antenna (2- or 4- branching). The presence of the ion at  $m/z$  775.3 ( $^{0,2}A_4$ ) in isomer 2 of the H6N5S2 glycan and its absence in isomer 3 (see Table 2) suggested that the  $\alpha$ 2-6 NeuAc present in the branched 3-antenna was located on the 4- and 2-branching respectively. Moreover, the presence of this specific ion in all isomers of H6N5S3 (see Table 2), corroborated the specific character of this ion fragment to characterize the NeuAc position on the 4-branching. Furthermore, it also confirmed that the  $\alpha$ 2-3 NeuAc in isomers 1 and 2 of this glycan was located on the 4- and 2-branching respectively (as they are related to isomers 3 and 2, respectively, of H6N5S2 glycan after  $\alpha$ 2-3 sialidase digestion, see Fig. 3). The assignment of the branching position of the  $\alpha$ 2-3 NeuAc would not have been possible without the combined strategy of specific sialidase digestion and MS/MS analysis. On the other hand, it is important to note that, while the ion 306.1 was present in all isomers containing at least one  $\alpha$ 2-6 NeuAc, and its relative abundance could be used to estimate the proportion of  $\alpha$ 2-6 NeuAc in each isomer, the ion 470.1 was only observed when all sialic acids were  $\alpha$ 2-6. This fact was concluded because the ion at  $m/z$  470.1 was not observed in isomers 1 and 2 of the native H6N5S3 but it was observed in the corresponding glycan isomers obtained after  $\alpha$ 2-3 sialidase digestion (isomers 3 and 2 of H6N5S2 glycan, respectively). Hence, ion 470.1 could be used for screening purposes to identify those glycan isomers with all sialic acids  $\alpha$ 2-6.

Characterization of highly branched sialylated glycans is still a challenge and therefore less addressed by glycomic researchers. In addition, in the particular case of hAGP, the characterization of tetraantennary glycans is even more complicated as they are less abundant and more difficult to ionize. Nevertheless, in this work, an estimation of the  $\alpha$ 2-6 NeuAc composition in H7N6S3 glycan isomers was carried out. For this purpose, only the ion fragment at  $m/z$  306.1 was used as other diagnostic ion fragments (such as 408.1 or 470.1) were not observed due to the lower abundance of this

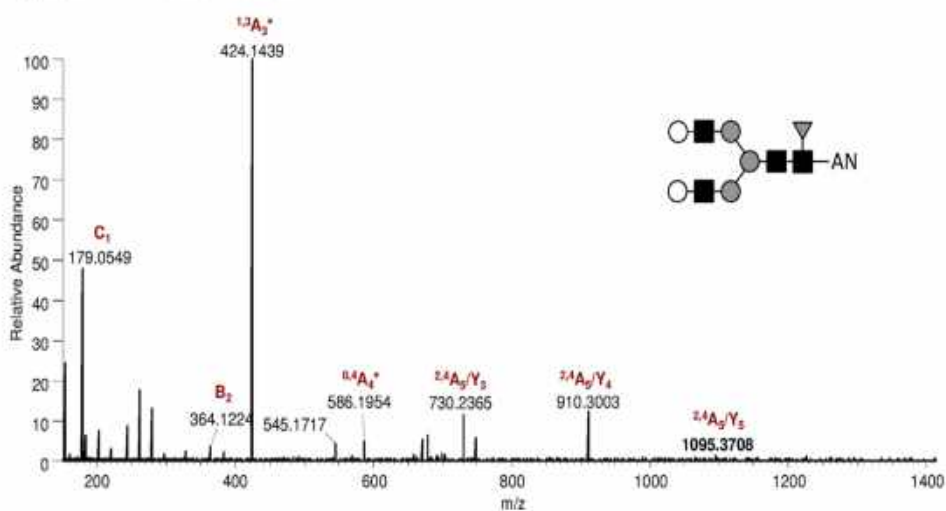
glycan. In agreement with our previous study [1], isomer 4 only presented  $\alpha$ 2-6 NeuAc with the highest intensity of 306.1 ion with respect to the other isomers (data not shown). Furthermore, the relative intensity of this diagnostic ion also suggested that the proportion of  $\alpha$ 2-6 NeuAc was higher in isomer 3 compared to isomer 2. Thus, our previous assignment [1] was complemented identifying isomer 2 with two  $\alpha$ 2-3 NeuAc and one  $\alpha$ 2-6 NeuAc, and isomer 3 with one  $\alpha$ 2-3 NeuAc and two  $\alpha$ 2-6 NeuAc. Nevertheless, contrary to the tentative characterization carried out in our previous work [1], the ion at  $m/z$  306.1 was not observed in isomer 1, which could reveal that all sialic acids present in this isomer are  $\alpha$ 2-3 linked. These results deepened the partial assignment of this glycan performed previously with exoglycosidase digestion [1], identifying tentatively the linkage-type of the three sialic acids in all isomers of H7N6S3. In addition, the characterization of all hAGP glycan isomers performed in our previous and present studies could also reveal that the proportion of  $\alpha$ 2-6 NeuAc is strongly related with the elution order of the ZIC-HILIC column, demonstrating that glycan isomers with a higher proportion of  $\alpha$ 2-6 linked NeuAc are eluted later.

### 3.3. Characterization of fucose linkage-type N-glycan isomers

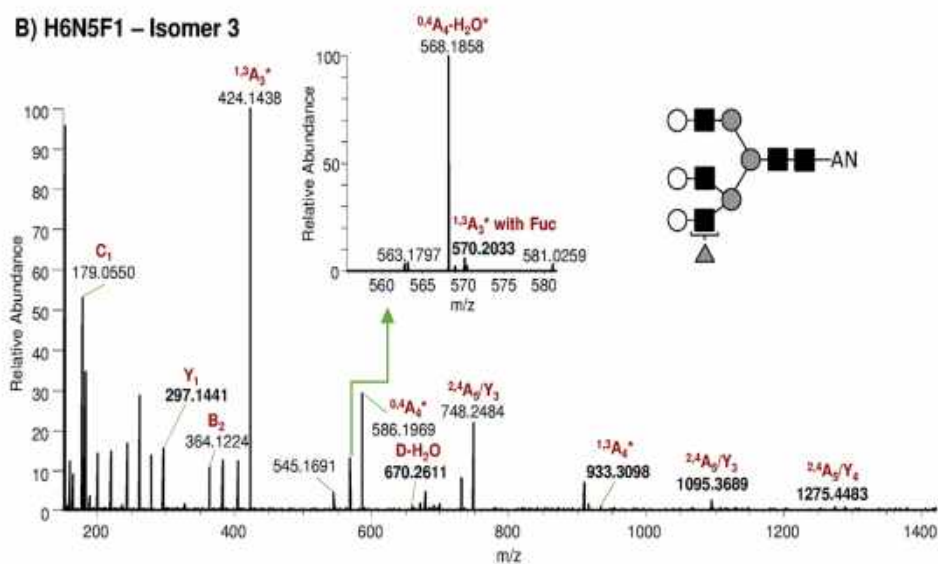
Fucose linkage-type characterization of the major glycans of hAGP was also performed in this work using several diagnostic ion fragments, some of them previously reported in the literature [12,21,29] (Table S2 of the supplementary material). In general, diagnostic fucose ion fragments are usually present at low abundance or are not directly observed in the tandem mass spectra as the fucose, mainly the antennary fucose, is preferably removed from the glycan structure during fragmentation [21]. Thus, few studies have been able to identify fucose linkage-type of complex N-glycans [12,21,42], which demonstrated finding novel specific ion fragments to differentiate between Fuc Core and Fuc Antenna is still necessary. As in the characterization of sialic acids, the assignment performed in Ref. [1] was used to verify the reliability of certain ion fragments described in the literature for fucose isomer identification. In this case, total sialidase digestion was also carried out in order to remove heterogeneity and obtain the isomers resulting only from different fucose linkages. Isomers 1 and 2 of H5N4F1 and H6N5F1 glycans, which were previously assigned as core fucosylated [1], and isomer 3 of H6N5F1 glycan, which was previously assigned as  $\alpha$ 1-3 antennary fucosylated [1], were studied in this work as they were intense enough to delve into the differentiation between Fuc Core and Fuc Antenna using tandem mass spectra ion fragments. Fig. 4 shows the tandem mass spectra of these glycans in hAGP. Fragments diagnostic of core fucosylation such as  $Y_1$  at  $m/z$  461.2 and 443.2 were found in isomers 1 and 2 of the biantennary glycan (H5N4F1) but at very low intensity and, thus, with higher error mass. Moreover, the ion at  $m/z$  443.2 was detected in isomer 1 and 2 of the triantennary glycan (H6N5F1) but only in half of the runs. These results suggested the difficulty of detecting these diagnostic ions when the complexity of the glycan increases, and thus, their absence in the MS/MS spectra could not always be indicative of lack of core fucosylation in tri- and tetraantennary glycans. However, the absence of the ion fragment  $Y_1$  at  $m/z$  297.1 in the tandem mass spectra of isomers 1 and 2 of H5N4F1 and H6N5F1 glycans (see Fig. 4A and C), which was abundant in the tandem mass spectra of the non-fucosylated glycans (Fig. 1A and B), revealed that the fucose unit could be linked to the core. Therefore, the previous assignment of these isomers as core fucosylated [1] was confirmed and the diagnostic potential of the ion fragment 297.1 to differentiate between Fuc Core and Fuc Antenna was demonstrated. Moreover, other ion fragments such as  $^{2,4}A_5/Y_5$  at  $m/z$  1095.3 suggested that the fucose unit was not located on any



## A) H5N4F1 – Isomer 1 and 2



## B) H6N5F1 – Isomer 3



## C) H6N5F1 – Isomer 1 and 2

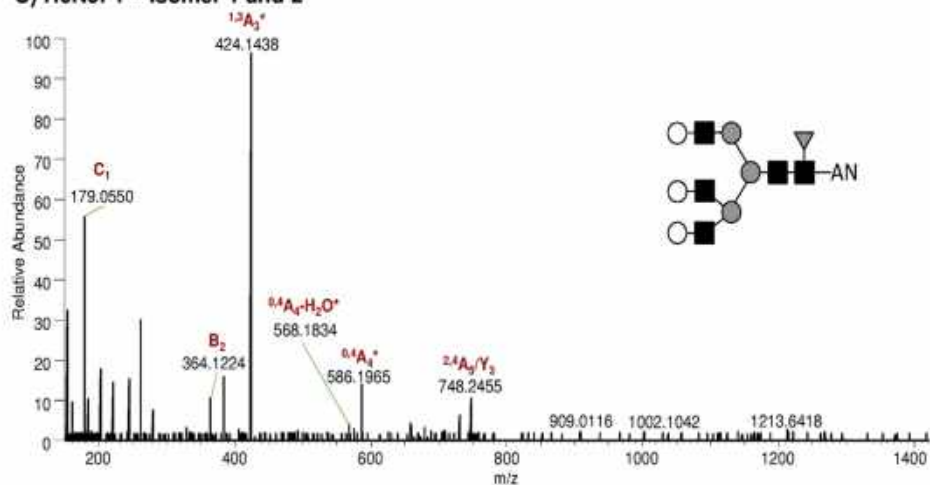
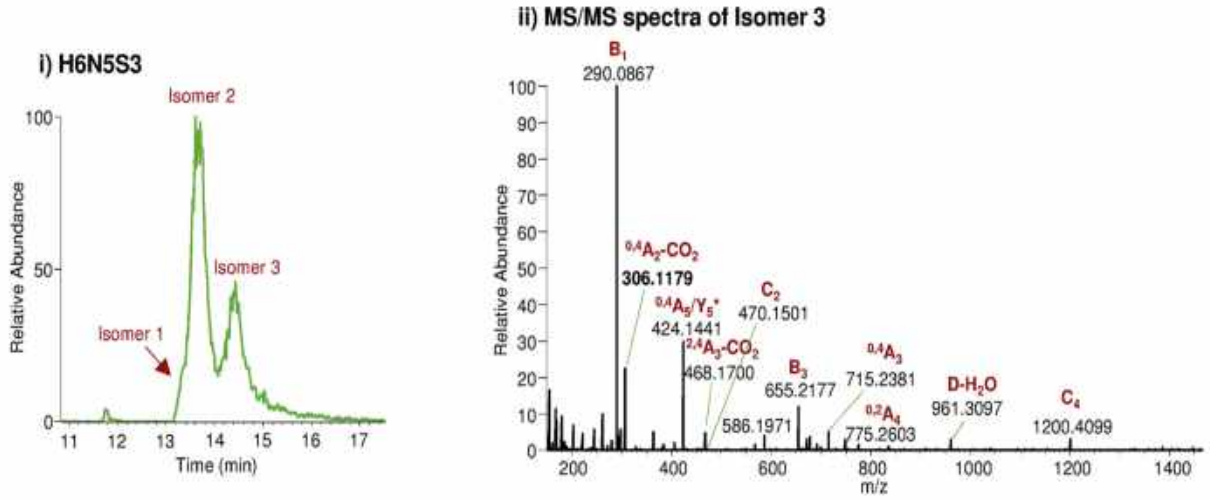
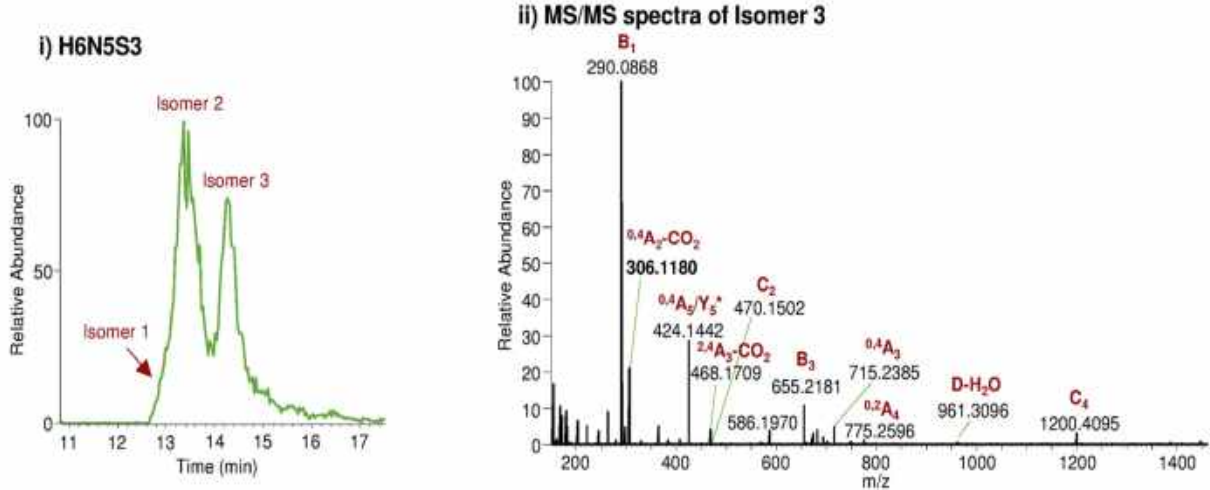


Fig. 4. Tandem mass spectra of isomers 1 and 2 of the H5N4F1 glycan (A), isomer 3 of the H6N5F1 glycan (B) and isomers 1 and 2 of the H6N5F1 glycan (C) obtained from hAGP standard after total sialidase digestion. (\*) This ion fragment could correspond to other possible structures listed in Table S1 or S2 of the supplementary material.

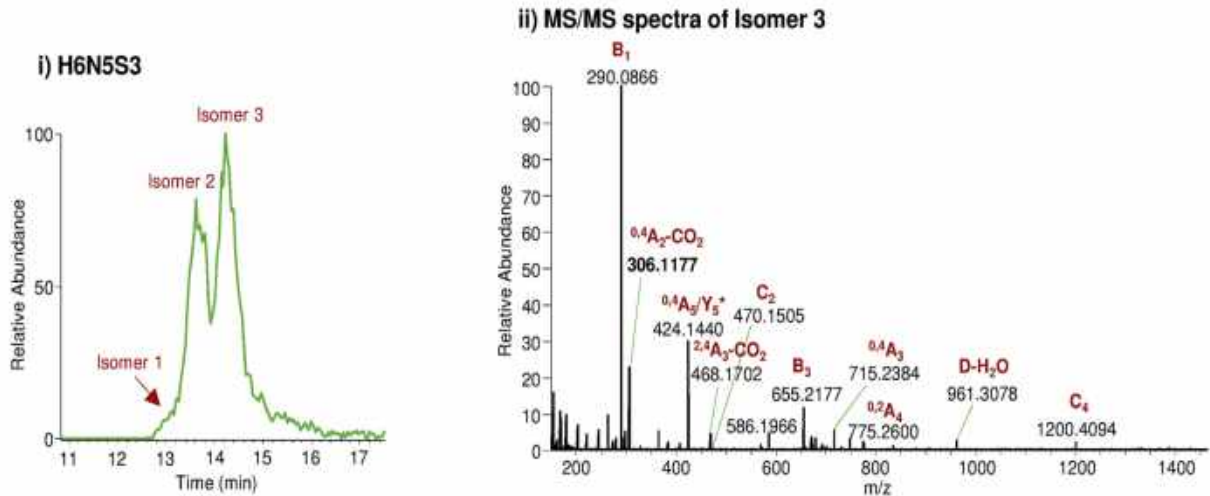
**A) Control**



**B) ChrP**



**C) PaC (stage IV)**



**Fig. 5.** EICs of the H6N5S3 glycan and the corresponding MS/MS spectra of the isomer 3 obtained from three hAGP serum samples: one control (A), one pathological sample with chronic pancreatitis (B) and one pathological sample with an advanced stage of pancreatic cancer (C). (\*) This ion fragment could correspond to other possible structures listed in Table S1 or S2 of the supplementary material.

antenna (see Table 1 and Fig. 4A). Additionally, in the tandem mass spectra of H5N4F1 and H6N5F1 (isomers 1 and 2), specific ion fragments of antennary fucosylation were not observed.

With respect to isomer 3 of H6N5F1 glycan, as can be observed in Fig. 4B, the ion fragment at  $m/z$  297.1 was observed with practically the same relative intensity than in the tandem mass spectrum of H6N5 glycan (Fig. 1B), which demonstrated, in this case, that the fucose was not linked to the core HexNAc. Furthermore, the ion 364.1 (which corresponds to the antennary Hex-HexNAc unit) presented lower relative intensity in comparison with the tandem mass spectrum of the H6N5 glycan, showing that the fucose unit could be located at one antenna. This fact was definitely confirmed with the presence of the fragment 570.2 ( $^{0,4}A_4/Y_5$ , see Fig. 4B and Table 1), despite its low intensity. Moreover, the absence of 325.1 ( $C_1$ , see Supplementary Table 2) revealed that the fucose was linked to the antennary HexNAc, as was described in our previous assignment [1], rather than to the Hex (galactose). Additionally, D-ion at  $m/z$  670.2 was observed (see Fig. 4B and Table 1) showing that the fucose could be located at 2- or 4-branching of the 3-antenna. The considerable decrease in the relative intensity of the ion fragments 933.3 ( $^{0,4}A_4$ ) and 1095.3 ( $^{2,4}A_5/Y_3$ ) with respect to the tandem mass spectrum of the H6N5 glycan corroborated that the fucose was linked to a HexNAc of the branched 3-antenna (compare Fig. 4B with Fig. 1B). The presence of these ions in the tandem mass spectra could be attributed to the preferable elimination of the fucose linked to the 3-position of the antennary HexNAc [21], as was previously mentioned.

### 3.4. Application of the MS/MS methodology to serum samples

Once the  $\mu$ ZIC-HILIC-MS/MS methodology was optimized and used to characterize sialic acid and fucose linkage-type isomers of several glycans released from hAGP standard, the same method was applied to hAGP serum samples to corroborate its robustness. Fig. 5 shows the extracted ion chromatogram (EIC) of the H6N5S3 glycan obtained in three serum samples: one control, one pathological sample with chronic pancreatitis (ChrP) and one with an advanced stage of pancreatic cancer (PaC, stage IV) and the corresponding MS/MS spectra obtained for the third isomer. As can be observed, the diagnostic ion fragments obtained in the tandem mass spectra of isomer 3 of H6N5S3 in real samples (Fig. 5 (ii)) were the same as the ones obtained in hAGP standard (see Table 2 and Fig. 3C (iii)). The same results were obtained for all studied glycans demonstrating the robustness of the established methodology. Even though in this work the analysis of these pathological samples was performed only with the aim of evaluating the MS/MS method in serum samples, the tendency to increase the relative abundance of the isomer 3 of the triantennary glycan in PaC with respect to the control and to the ChrP samples was again revealed (see Fig. 5). This fact was reported in a previous work [36] in which a larger number of PaC and ChrP samples were analyzed and could be again related with our assignment, as isomer 3 was the only one that presented all sialic acids  $\alpha$ 2-6 linked and the overexpression of this linkage-type was reported in cancer by several authors [8,20,43]. The glycans obtained from these hAGP serum samples were also treated with total sialidase in order to study the future applicability of this methodology in the characterization of fucose isomers. Similar tandem mass spectra were obtained for all fucosylated glycan isomers compared to the ones obtained previously in hAGP standard (data not shown). These results revealed again the robustness of this approach and demonstrated that the analysis of additional PaC samples is still pending to evaluate the potential of some glycan isomers as possible biomarkers of this pathology.

## 4. Conclusions

In this paper, a  $\mu$ ZIC-HILIC-MS/MS methodology was established in negative ion mode for the characterization of glycan isomers. With this strategy, the separation of the glycan isomers by  $\mu$ ZIC-HILIC combined with high resolution tandem mass spectrometry detection enabled to unequivocally assign the ion fragments to a particular glycan structure. Firstly, the most important diagnostic ion fragments, previously reported in Refs. [12,13,21,27–34] and used for the characterization of glycans, were evaluated with hAGP. Sequencing of *N*-glycan antennas was carried out using B- and C-ions and, structural features such as branching location were determined using D-ions. This method was also applied in this work for the assignment of the sialic acid linkage-type. In this regard, characterization of bi- and triantennary hAGP glycans was carried out in this study using several specific ion fragments. The exact mass of daughter ions provided by the LTQ-Orbitrap mass spectrometer and our previous results using exoglycosidase digestion in combination with the GRIL methodology [1] permitted us to confirm or discard some diagnostic ion fragments proposed in the literature. Moreover, unlike our previous work [1], the established MS/MS methodology enabled the location of  $\alpha$ 2-3 and  $\alpha$ 2-6 sialic acids on antennas and also the complete characterization of some tetraantennary glycans such as H6N7S3, which were only partially characterized with exoglycosidase digestion [1]. Furthermore, based on the characterization performed previously in Ref. [1], differentiation between Fuc core and Fuc antenna isomers was also performed in this work using specific ion fragments obtained in the tandem mass spectra.  $Y_1$  ions at  $m/z$  461.2 and 443.2 were not considered reliable diagnostic fragments of core fucosylation in the case of tri- and tetraantennary glycans as they were hardly detected in those isomers previously identified as core fucosylated [1]. On the other hand, the ion fragment 297.1, which describes the absence of core fucosylation, was established as an important diagnostic ion for fucose linkage-type assignment. Finally, robustness and future applicability of the MS/MS method was demonstrated for the study of control and pathological serum samples in order to find novel glycan isomers that could be used as biomarkers in cancer.

## Acknowledgments

This study was supported by Spanish Ministry of Economy and Competitiveness (CTQ2014-56777-R awarded to V. S-N and BIO2015-66356-R awarded to R.P.). Montserrat Mancera-Arteu acknowledges the University of Barcelona for an ADR fellowship. The authors thank the Separation Techniques Unit of the Scientific and Technological Centres of the University of Barcelona (CCiTUB) for technical assistance and advice.

## Appendix A. Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.aca.2017.07.068>.

## References

- [1] 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.
- [2] R. Dwek, Glycobiology: towards understanding the function of sugars, *Chem. Rev.* 96 (1996) 683–720.
- [3] W. Morelle, J.C. Michalski, Analysis of protein glycosylation by mass spectrometry, *Nat. Protoc.* 2 (2007) 1585–1602.
- [4] K. Martino, J. Bones, J.J. Kattila, P.M. Rudd, A systematic approach to protein glycosylation analysis: a path through the maze, *Nat. Chem. Biol.* 6 (2010)

- 713–723.
- [5] T. Cartwright, R. Schwalbe, Atypical sialylated N-glycan structures are attached to neuronal voltage-gated potassium channels, *Biosci. Rep.* 29 (2009) 301–313.
  - [6] M.J. Kallileima, D. Park, C.B. Lebrilla, Glycan and glycoproteins as specific biomarkers for cancer, *Anal. Bioanal. Chem.* 409 (2017) 395–410.
  - [7] A. Lee, J.M. Chick, D. Kolarich, P. Haynes, G.R. Robertson, M. Tsoli, L. Jankova, S.J. Clarke, N.H. Packer, M.S. Baker, Liver membrane proteome glycosylation changes in mice bearing an extra-hepatic tumor, *Mol. Cell. Proteomics* 10 (2011).
  - [8] Y. Mechref, Y. Hu, A. Garcia, A. Hussein, Identifying cancer biomarkers by mass spectrometry-based glycomics, *Electrophoresis* 33 (2012) 1755–1767.
  - [9] M. Nakano, R. Saldanha, A. Gobet, M. Kavallaris, N.H. Packer, Identification of glycan structure alterations on cell membrane proteins in desoxyepithelone B resistant leukemia cells, *Mol. Cell. Proteomics* 10 (2011).
  - [10] M.K. Sethi, W.S. Hancock, S. Fanayan, Identifying N-glycan biomarkers in colorectal cancer by mass spectrometry, *Am. Chem. Soc.* 49 (2016) 2099–2106.
  - [11] P.H. Jensen, N.G. Karlsson, D. Kolarich, N.H. Packer, Structural analysis of N- and O-glycans released from glycoproteins, *Nat. Protoc.* 7 (2012) 1299–1310.
  - [12] D.J. Harvey, L. Royle, C.M. Radcliffe, P.M. Rudd, R.A. Dwek, Structural and quantitative analysis of N-linked glycans by matrix-assisted laser desorption/ionization and negative ion nanospray mass spectrometry, *Anal. Biochem.* 376 (2008) 44–60.
  - [13] K. Deguchi, Y. Takegawa, H. Ito, N. Miura, S. Yoshioka, S. Nagai, H. Nakagawa, S.I. Nishimura, Structural assignment of isomeric 2-aminopyridine-derivatized monosialylated biantennary N-linked oligosaccharides using negative-ion multistage tandem mass spectral matching, *Rapid Commun. Mass Spectrom.* 20 (2006) 412–418.
  - [14] E. Toussi, J. BONES, W.S. Hancock, M. Hincapie, Differential chemical derivatization integrated with chromatographic separation for analysis of isomeric sialylated N-glycans: a nano-hydrophilic interaction liquid chromatography-MS platform, *Anal. Chem.* 85 (2013) 8421–8428.
  - [15] M. Pabst, J. Grass, S. Toegel, E. Liebmingler, R. Strasser, F. Altmann, Isomeric analysis of oligomannosidic N-glycans and their difucosylated precursors, *Glycobiology* 22 (2011) 389–399.
  - [16] C.A. Reis, H. Osorio, L. Silva, C. Gomes, L. David, Alterations in glycosylation as biomarkers for cancer detection, *J. Clin. Pathol.* 63 (2010) 322–329.
  - [17] M. Balmaña, E. Giménez, A. Puerta, E. Llop, J. Figueras, E. Fort, V. Sanz-Nebot, C. de Bolós, A. Rizzi, S. Barrabés, M. de Frutos, R. Peracaula, Increased  $\alpha$ 1-3 fucosylation of  $\alpha$ -1-acid glycoprotein (AGP) in pancreatic cancer, *J. Proteomics* 132 (2016) 144–154.
  - [18] S. Singh, K. Pal, J. Yadav, H. Tang, K. Partyka, D. Kletter, P. Hsueh, E. Ensink, B. KC, G. Hostetter, H.E. Xu, M. Bern, D.F. Smith, A.S. Mehta, R. Brand, K. Melcher, B.B. Haab, Upregulation of glycans containing 3' fucose in a subset of pancreatic cancers uncovered using fusion-tagged lectins, *J. Proteome Res.* 14 (2015) 2594–2605.
  - [19] L.R. Ruhaak, S. Miyamoto, C.B. Lebrilla, Developments in the identification of glycan biomarkers for the detection of cancer, *Mol. Cell. Proteomics* 12 (2013) 846–855.
  - [20] M. Hedlund, E. Ng, A. Varki, N.M. Varki,  $\alpha$ 2-6-linked sialic acids on N-glycans modulate carcinoma differentiation in vivo, *Cancer Res.* 68 (2008) 388–394.
  - [21] D.J. Harvey, Fragmentation of negative ions from carbohydrates: Part 3. Fragmentation of hybrid and complex N-linked glycans, *J. Am. Soc. Mass Spectrom.* 16 (2005) 647–659.
  - [22] J.L. Abrahams, N.H. Packer, M.P. Campbell, Relative quantitation of multi-antennary N-glycan classes: combining PGC-LC-ESI-MS with exoglycosidase digestion, *Analyst* (2015) 5444–5449.
  - [23] L. Mauko, N.A. Lacher, M. Pelzing, A. Nordborg, P.R. Haddad, E.F. Hilder, Comparison of ZIC-HILIC and graphitized carbon-based analytical approaches combined with exoglycosidase digestions for analysis of glycans from monoclonal antibodies, *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* 911 (2012) 93–104.
  - [24] S. Tao, Y. Huang, B.E. Boyes, R. Orlando, Liquid chromatography-selected reaction monitoring (LC-SRM) approach for the separation and quantitation of sialylated N-glycans linkage isomers, *Anal. Chem.* 86 (2014) 10584–10590.
  - [25] K.R. Reiding, D. Blank, D.M. Kuipper, A.M. Deelder, M. Wührer, High-throughput profiling of protein N-glycosylation by MALDI-TOF-MS employing linkage-specific sialic acid esterification, *Anal. Chem.* 86 (2014) 5784–5793.
  - [26] S.F. Wheeler, P. Domann, D.J. Harvey, Derivatization of sialic acids for stabilization in matrix-assisted laser desorption/ionization mass spectrometry and concomitant differentiation of  $\alpha$ (2-3)- and  $\alpha$ (2-6)-isomers, *Rapid Commun. Mass Spectrom.* 23 (2009) 303–312.
  - [27] D.J. Harvey, Structural determination of N-linked glycans by matrix-assisted laser desorption/ionization and electrospray ionization mass spectrometry, *Proteomics* 5 (2005) 1774–1786.
  - [28] S.F. Wheeler, D.J. Harvey, Negative ion mass spectrometry of sialylated carbohydrates: discrimination of N-acetylneuraminic acid linkages by MALDI-TOF and ESI-TOF mass spectrometry, *Anal. Chem.* 72 (2000) 5027–5039.
  - [29] A.V. Everest-Dass, J.L. Abrahams, D. Kolarich, N.H. Packer, M.P. Campbell, Structural feature ions for distinguishing N- and O-linked glycan isomers by LC-ESI-IT MS/MS, *J. Am. Soc. Mass Spectrom.* 24 (2013) 895–906.
  - [30] C. Michael, A.M. Rizzi, Tandem mass spectrometry of isomeric aniline-labeled N-glycans separated on porous graphitic carbon: revealing the attachment position of terminal sialic acids and structures of neutral glycans, *Rapid Commun. Mass Spectrom.* 29 (2015) 1268–1278.
  - [31] D.J. Harvey, Fragmentation of negative ions from carbohydrates: Part 2. Fragmentation of high-mannose N-linked glycans, *J. Am. Soc. Mass Spectrom.* 16 (2005) 631–646.
  - [32] D.J. Harvey, R.I. Martin, K.A. Jackson, C.W. Sutton, Fragmentation of N-linked glycans with a matrix-assisted laser desorption/ionization ion trap time-of-flight mass spectrometer, *Rapid Commun. Mass Spectrom.* 18 (2004) 2997–3007.
  - [33] H. Ito, K. Yamada, K. Deguchi, H. Nakagawa, S.-I. Nishimura, Structural assignment of disialylated biantennary N-glycan isomers derivatized with 2-aminopyridine using negative-ion multistage tandem mass spectral matching, *Rapid Commun. Mass Spectrom.* 21 (2007) 212–218.
  - [34] S. Zhou, X. Dong, L. Veillon, Y. Huang, Y. Mechref, LC-MS/MS analysis of permethylated N-glycans facilitating isomeric characterization, *Anal. Bioanal. Chem.* 409 (2017) 453–466.
  - [35] A.V. Everest-Dass, D. Kolarich, M.P. Campbell, N.H. Packer, Tandem mass spectra of glycan substructures enable the multistage mass spectrometric identification of determinants on oligosaccharides, *Rapid Commun. Mass Spectrom.* 27 (2013) 931–939.
  - [36] E. Giménez, M. Balmaña, J. Figueras, E. Fort, C. de Bolós, V. Sanz-Nebot, R. Peracaula, A. Rizzi, Quantitative analysis of N-glycans from human  $\alpha$ -acid-glycoprotein using stable isotope labeling and zwitterionic hydrophilic interaction capillary liquid chromatography electrospray mass spectrometry as tool for pancreatic disease diagnosis, *Anal. Chim. Acta* 866 (2015) 59–68.
  - [37] V. Padler-Karavani, Aiming at the sweet side of cancer: aberrant glycosylation as possible target for personalized-medicine, *Cancer Lett.* 352 (2014) 102–112.
  - [38] A. Sarrafs, R. Saldova, E. Pla, E. Fort, D.J. Harvey, W.B. Struwe, L.R. De, P.M. Rudd, R. Peracaula, Glycosylation of liver acute-phase proteins in pancreatic cancer and chronic pancreatitis, *Proteomics. Clin. Appl.* 4 (2010) 432–448.
  - [39] K.D. Smith, J. Behan, G. Matthews Smith, A.M. Magliocco,  $\alpha$ -1-acid glycoprotein (AGP) as a potential biomarker for breast cancer, *Glycosylation* (2012) 201–222.
  - [40] 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.
  - [41] B. Domon, C.E. Costello, A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS spectra of glycoconjugates, *Glycoconj. J.* 5 (1988) 397–409.
  - [42] C. Nwosu, H.J. Yau, S. Becht, Assignment of core versus antenna fucosylation types in protein N-glycosylation via procainamide labeling and tandem mass spectrometry, *Anal. Chem.* 87 (2015) 5905–5913.
  - [43] B.N. Vajaria, K.R. Patel, R. Begum, P.S. Patel, Sialylation: an avenue to target cancer cells, *Pathol. Oncol. Res.* 22 (2016) 443–447.



## **Supplementary Figures and Tables**

---



**Supplementary Table S-1.** Summary of the most important characteristic ion fragments obtained in negative ion mode tandem mass spectra for the characterization of sialylated and desialylated complex-type *N*-glycans. Most of these fragments have been previously described in the literature [11, 13, 21, 25-32].

| <i>N</i> -glycan ion fragments   |  | Desialylated <i>N</i> -glycan ion fragments  |  | Characteristic fragment of: |
|--|--|--|--|-----------------------------|
| Nomenclature   | <i>m/z</i> , <i>m/z</i> – H <sub>2</sub> O | Nomenclature   | <i>m/z</i> , <i>m/z</i> – H <sub>2</sub> O |                             |
|  | -  | B <sub>1</sub>   | 161.0450                                   |                             |
|  | -  | C <sub>1</sub>   | 179.0556                                   |                             |
| B <sub>1</sub>   | 290.0876                                   |  | -  |                             |
|  | -  | C <sub>2</sub>   | 382.1349                                   |                             |
| <sup>0,4</sup> A <sub>5</sub> /Y <sub>5</sub> or <sup>1,3</sup> A <sub>5</sub> /Y <sub>5</sub> <sup>a</sup> or <sup>1,5</sup> A <sub>2</sub> | 424.1455,<br>406.1349                      | <sup>0,4</sup> A <sub>4</sub> /Y <sub>5</sub> or <sup>1,3</sup> A <sub>4</sub> /Y <sub>5</sub> <sup>a</sup> or <sup>0,4</sup> A <sub>3</sub> or <sup>1,3</sup> A <sub>3</sub>              | 424.1455,<br>406.1349                      |                             |
|  |  | E  | 466.1560                                   |                             |
| <sup>2,4</sup> A <sub>3</sub> -CO <sub>2</sub>   | 468.1717                                   |  | -  |                             |
|  | -  | B <sub>3</sub> /Z <sub>4</sub>   | 508.1641                                   | Branched antenna            |
|  | -  | B <sub>3</sub>   | 526.1772                                   |                             |
|  | -  | C <sub>3</sub>   | 544.1877                                   |                             |
| <sup>0,4</sup> A <sub>5</sub> /Y <sub>6</sub> or <sup>1,3</sup> A <sub>5</sub> /Y <sub>6</sub> <sup>a</sup>                                  | 586.1983,<br>568.1877                      | <sup>0,4</sup> A <sub>4</sub> or <sup>1,3</sup> A <sub>4</sub> <sup>a</sup> or <sup>0,4</sup> A <sub>4</sub> /Y <sub>4</sub> or <sup>1,3</sup> A <sub>4</sub> /Y <sub>4</sub> <sup>a</sup> | 586.1983,<br>568.1877                      |                             |
|  | -  | <sup>0,2</sup> X <sub>4</sub> /B <sub>3</sub>  | 609.2143                                   | Branched antenna            |
| B <sub>3</sub>   | 655.2198                                   |  | -  |                             |
| <sup>0,4</sup> A <sub>4</sub> or <sup>1,3</sup> A <sub>4</sub> <sup>b</sup>  | 715.2409                                   |  | -  |                             |



|                              |                         |  |                         |                                    |
|------------------------------|-------------------------|--|-------------------------|------------------------------------|
|                              | -                       | $^{2,4}A_5/Y_3$                                      | 748.2511,<br>730.2405   |                                    |
| $^{0,2}A_4$                  | 775.2620                |  | -                       | Sialylated 4- or 6-<br>branching   |
|                              |                         | E  | 831.2882                |                                    |
| $C_4$                        | 835.2831,<br>817.2726   |  | -                       |                                    |
|                              | -                       | $^{1,5}A_3$ or $^{1,5}A_5/Y_3^b$                     | 863.3144                |                                    |
|                              |                         | $^{2,4}A_5/Y_4$                                      | 910.3039                |                                    |
| $^{1,5}A_5/Y_3$              | 951.3305,<br>933.3199   | $^{0,4}A_4$ or $^{1,3}A_4^a$ or<br>$^{0,2}X_4/C_4^b$ | 951.3305,<br>933.3199   | Branched antenna <sup>c</sup>      |
| $^{1,5}A_5/Y_4$              | 1113.3833,<br>1095.3727 | $^{2,4}A_5/Y_3$ or<br>$^{2,4}A_5/Y_5^b$              | 1113.3833,<br>1095.3727 | Branched antenna <sup>c</sup>      |
| $C_4$ or $C_6/Y_3$           | 1200.4153,<br>1182.4047 |  | -                       | Monosialylated<br>branched antenna |
| $^{2,4}A_6/Y_4$              | 1201.3993               |  | -                       |                                    |
| $^{0,4}A_5$ or $^{1,3}A_5^a$ | 1242.4259,<br>1224.4153 |  | -                       | Monosialylated<br>branched antenna |
|                              | -                       | $^{2,4}A_5/Y_4$ or $^{2,4}A_5^b$                     | 1275.4361,<br>1257.4256 |                                    |
| $^{2,4}A_6/Y_5$              | 1404.4787,<br>1386.4681 |  | -                       |                                    |

<sup>a</sup> Fragment  $^{1,3}A_5$  for sialylated *N*-glycans as well as  $^{1,3}A_4$  for desialylated ones, correspond to the same exact mass of fragment  $^{2,4}A_5$  and  $^{2,4}A_4$  respectively.

<sup>b</sup> Only possible ion fragment structure in the case of biantennary *N*-glycans.

<sup>c</sup> Characteristic of branched antenna in tri- and tetraantennary glycans.

**Supplementary Table S-2.** Summary of diagnostic ion fragments obtained in negative ion mode tandem mass spectra for the characterization of important structural features of sialylated and desialylated complex-type *N*-glycan isomers. These ion fragments have been previously described in the literature for the characterization of 6-antenna [11, 21, 25, 26], sialic acid linkage-type [28, 30] and fucose linkage-type [11, 25, 26, 32].

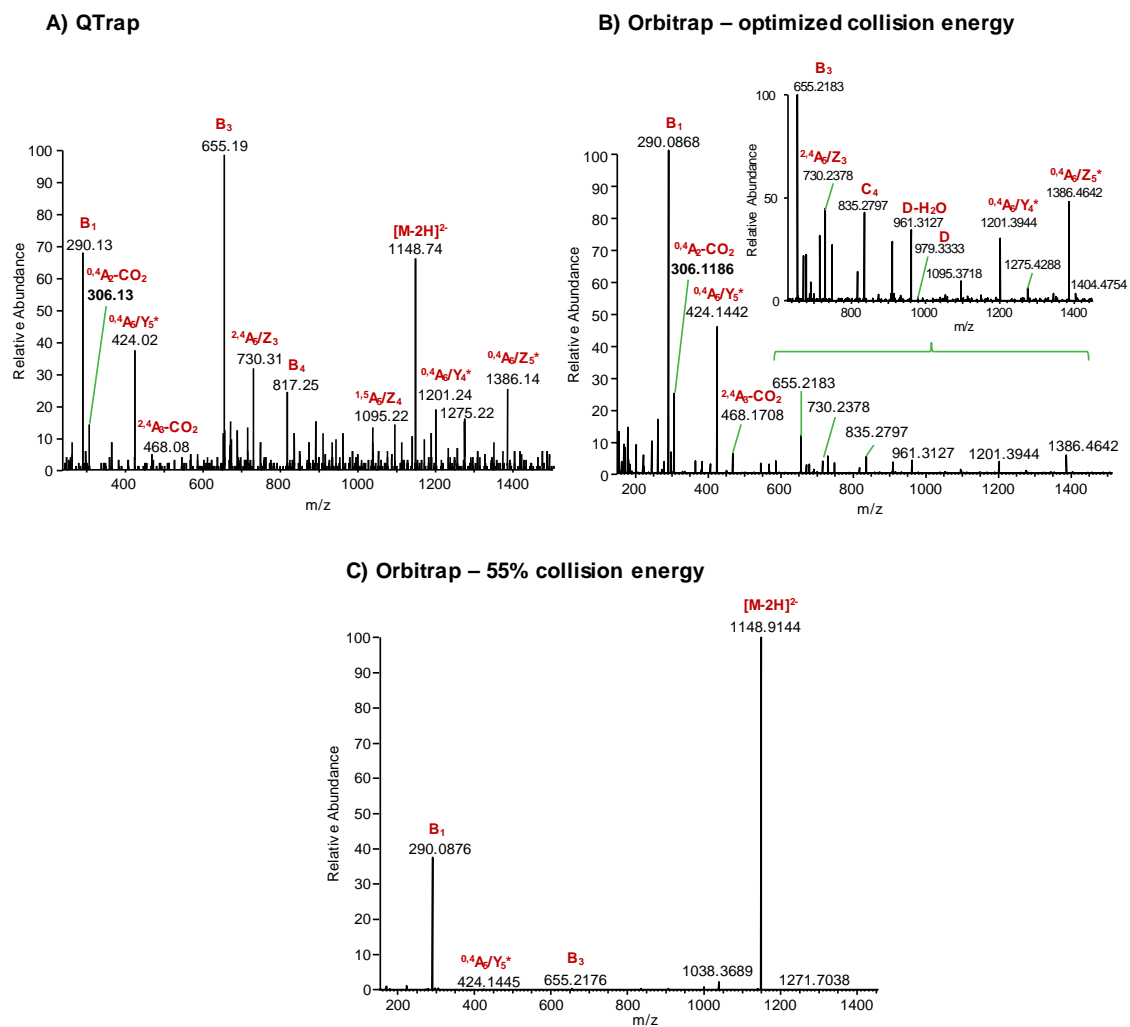
|   | N-glycan ion fragments  |                              | Desialylated N-glycan ion fragments   |                              | Diagnostic fragment of:                          |
|---|---|------------------------------|---|------------------------------|--|
|   | Nomenclature  | m/z , m/z – H <sub>2</sub> O | Nomenclature  | m/z , m/z – H <sub>2</sub> O |  |
| Ions specific to the 6-antenna              | D   | 979.3254,<br>961.3148        | D   | 688.2300,<br>670.2194        | Unbranched 6-antenna                             |
|   | D   | 1125.3833,<br>1107.3728      | D   | 834.2879,<br>816.2773        | Unbranched fucosylated 6-antenna <sup>c</sup>    |
|   | D   | 1635.5530,<br>1617.5424      | D   | 1053.3622,<br>1035.3516      | Branched 6-antenna                               |
|   | D   | 1781.6109,<br>1763.6004      | D   | 1199.4201,<br>1181.4096      | Branched fucosylated 6-antenna <sup>c</sup>      |
| Ions diagnostic of sialic acid linkage-type | <sup>0,4</sup> A <sub>2</sub> -CO <sub>2</sub>  | 306.1189                     |   | -                            | α2-6 linked NeuAc                                |
|   | <sup>0,4</sup> A <sub>2</sub> or <sup>1,3</sup> A <sub>2</sub> or <sup>2,4</sup> A <sub>2</sub>             | 350.1087                     |   | -                            | α2-3 linked NeuAc                                |
|   | B <sub>3</sub> /Y <sub>6</sub> or <sup>3,5</sup> A <sub>2</sub>   | 364.1243                     | B <sub>2</sub>  | 364.1243                     | α2-6 linked NeuAc                                |
|   | B <sub>2</sub> -CO <sub>2</sub>   | 408.1506                     |   | -                            | α2-3 linked NeuAc                                |
|   | C <sub>2</sub>  | 470.1510                     |   | -                            | α2-6 linked NeuAc                                |
| Ions diagnostic of fucose linkage-type      | Y <sub>1</sub>  | 297.1212,<br>279.1106        | Y <sub>1</sub>  | 297.1212,<br>279.1106        | Absence of core fucosylation                     |
|   |   | -                            | C <sub>1</sub>  | 325.1134                     | Fucosylation on antennary galactose <sup>c</sup> |
|   | Y <sub>1</sub>  | 461.1898,<br>443.1791        | Y <sub>1</sub>  | 461.1898,<br>443.1791        | Core fucosylation <sup>c</sup>                   |
|   | B <sub>3</sub> /Y <sub>6</sub> or <sup>3,5</sup> A <sub>2</sub>   | 528.1928,<br>510.1823        | B <sub>2</sub>  | 528.1928,<br>510.1823        | Fucosylated antenna <sup>c</sup>                 |
|   | <sup>0,4</sup> A <sub>5</sub> /Y <sub>5</sub> or <sup>1,3</sup> A <sub>5</sub> /Y <sub>5</sub> <sup>a</sup> | 570.2034                     | <sup>0,4</sup> A <sub>4</sub> /Y <sub>5</sub> or <sup>1,3</sup> A <sub>4</sub> /Y <sub>5</sub> <sup>a</sup> or <sup>0,4</sup> A <sub>3</sub> or <sup>1,3</sup> A <sub>3</sub> | 570.2034                     | Fucosylated antenna <sup>c</sup>                 |

|  |             |                       |   |                       |   |
|--|-------------|-----------------------|---|-----------------------|---|
|  | $Y_2$       | 646.2585,<br>628.2479 | $Y_2$                                   | 646.2585,<br>628.2479 | Core fucosylation <sup>c</sup>                    |
|  | $^{1,5}X_2$ | 674.2534              | $^{1,5}X_2$                             | 674.2534              | Core fucosylation <sup>c</sup>                    |
|  | $^{0,2}X_2$ | 688.2690              | $^{0,2}X_2$                             | 688.2690              | Core fucosylation <sup>c</sup>                    |
|  |             | -                     | $^{0,4}A_4/Y_3$ or<br>$^{1,3}A_4/Y_5^a$ | 789.2777,<br>771.2671 | Absence of<br>fucosylation on<br>branched antenna |
|  | $B_3$       | 801.2776              |   | -                     | Fucosylated<br>antenna <sup>c</sup>               |
|  |             | -                     | $^{0,4}A_4/Y_3$ or<br>$^{1,3}A_4/Y_5^a$ | 935.3356              | Branched<br>fucosylated antenna <sup>c</sup>      |
|  |             | -                     | $^{1,5}A_3$ or $^{1,5}A_5/Y_3^b$        | 1009.3723             | Fucosylated<br>antenna <sup>c</sup>               |

<sup>a</sup> Fragment  $^{1,3}A_5$  for sialylated *N*-glycans as well as  $^{1,3}A_4$  for desialylated ones, correspond to the same exact mass of fragment  $^{2,4}A_5$  and  $^{2,4}A_4$  respectively.

<sup>b</sup> Only possible ion fragment structure in the case of biantennary *N*-glycans.

<sup>c</sup> The exact mass of this ion fragment includes the mass of one fucose unit.



**Supplementary Figure S-1.** Tandem mass spectra of isomer 2 of H5N4S2 glycan detected in hAGP standard using (A) QTrap and (B) LTQ-Orbitrap at the optimized HCD collision energy (75%), and (C) LTQ-Orbitrap at 55% of HCD collision energy. (\*) This ion fragment could correspond to other possible structures listed in Table S1 and S2 of the supplementary material.



## **Chapter 4. Glycan-based biomarker discovery**

---



In the last years, the search of glycan structures that could be used as potential biomarkers of several pathologies has aroused great interest in biomedicine. Elevated levels of total sialic acid (TSA), different expression between  $\alpha$ 2-3/ $\alpha$ 2-6 linked sialic acids, increased branching, addition of poly N-acetyllactosamine (polylacNAc) as well as upregulation of certain fucosylated epitopes have been the most important alterations reported in glycans in inflammatory processes and cancer. Moreover, in the case of cancer, these alterations seem to be related to the tumor initiation and progress as well as to metastasis, promoting certain glycan structures the extravasation of tumor cells. Therefore, finding novel glycan-based biomarkers may have much better performance for early diagnosis, disease monitoring and prognosis.

In this chapter, the alterations occurred in hAGP glycan isomers are studied in patients with pancreatic ductal adenocarcinoma (PDAC) and chronic pancreatitis (ChrP). The carbohydrate antigen 19-9 (CA 19-9), which is currently the only biomarker recommended for PDAC, shows an inadequate sensitivity and false results, especially in patients suffering from other nonmalignant diseases such as ChrP. Certain hAGP glycan isomers are proposed in this chapter as biomarker candidates of PDAC, which allow the differentiation from ChrP. In addition, using the characterization of hAGP glycan isomers performed in the previous chapter of this thesis, the existence of a relation between specific sialic acid or fucose linkage-types and PDAC is investigated.

Similarly, changes in the expression of mouse transferrin (mTf) glycan isomers in mice with collagen-induced arthritis (CIA), an homologous disease in many aspects to rheumatoid arthritis (RA) in humans, are also studied in this chapter. Currently, there is no single test to confirm the diagnosis of RA and it is based on the symptoms and the measurement of some variables. In this chapter, certain mTf glycan isomers are proposed as biomarkers of CIA, which



could be useful, in the future, to find novel glycan biomarkers for the diagnosis of RA in humans.

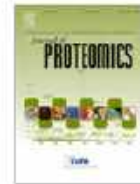
The work carried out in these studies has resulted in the publications listed below:

- **Publication 4.1.-** Multivariate data analysis for the detection of human alpha-acid glycoprotein aberrant glycosylation in pancreatic ductal adenocarcinoma. Mancera-Arteu, M., Giménez, E., Balmaña, M., Barrabés, S., Albiol-Quer, M., Fort, E., Peracaula, R., Sanz-Nebot, V. *Journal of Proteomics* (2019), 195, 76-87.
- **Publication 4.2.-** Alterations in the glycan profile of mouse transferrin: New insights in collagen-induced arthritis. Mancera-Arteu, M., Giménez, E., Sancho, J., Sanz-Nebot, V. *Journal of Proteome Research* (2020), 19, 1750-1759.



Contents lists available at ScienceDirect

Journal of Proteomics

journal homepage: [www.elsevier.com/locate/jprot](http://www.elsevier.com/locate/jprot)

## Multivariate data analysis for the detection of human alpha-acid glycoprotein aberrant glycosylation in pancreatic ductal adenocarcinoma

Montserrat Mancera-Artu<sup>a</sup>, Estela Giménez<sup>a,\*</sup>, Meritxell Balmaña<sup>b,c</sup>, Sílvia Barrabés<sup>b,d</sup>, Maite Albiol-Quer<sup>e</sup>, Esther Fort<sup>f</sup>, Rosa Peracaula<sup>b,d</sup>, Victòria Sanz-Nebot<sup>a</sup>

<sup>a</sup> Department of Chemical Engineering and Analytical Chemistry, Institute for Research on Nutrition and Food Safety (INSA UB), University of Barcelona, Barcelona, Spain

<sup>b</sup> Biochemistry and Molecular Biology Unit, Department of Biology, University of Girona, Girona, Spain

<sup>c</sup> Glycobiology in Cancer Group, I3S - Instituto de Investigação e Inovação em Saúde, Ipatimup - Institute of Molecular Pathology and Immunology, University of Porto, Porto, Portugal

<sup>d</sup> Biomedical Research Institute of Girona (IRIBGI), Salt, Spain

<sup>e</sup> Department of Surgery, Dr. Josep Trueta University Hospital, Girona, Spain

<sup>f</sup> Department of Gastroenterology, Dr. Josep Trueta University Hospital, Girona, Spain



### ARTICLE INFO

#### Keywords

Glycan isomers  
Multivariate data analysis  
Human alpha-acid glycoprotein  
Pancreatic cancer  
Sialylation  
Fucosylation

### ABSTRACT

Relative quantification of human alpha-acid glycoprotein (hAGP) glycan isomers using [<sup>13</sup>C<sub>6</sub>]/[<sup>12</sup>C<sub>6</sub>]-aniline in combination with multivariate data analysis is proposed as an efficient method for the identification of pancreatic ductal adenocarcinoma (PDAC) glycan biomarkers in serum samples. Intact and desialylated glycans from hAGP, purified from serum samples of patients with PDAC and chronic pancreatitis (ChrP), were labeled with aniline and analyzed by μZIC-HILIC-MS. Afterwards, partial least squares discriminant analysis (PLS-DA) was applied to the relative areas obtained for all glycan isomers in the different samples: pathological (ChrP or PDAC) versus healthy samples. Seven intact glycan isomers with α2-6 linked sialic acids, five of them also fucosylated, were the most meaningful to distinguish between PDAC and ChrP patients. The desialylated glycan isomers also identified by PLS-DA as potential biomarker candidates confirmed that antenna but also core fucosylation could be involved in PDAC. The analysis of intact and desialylated glycan isomers in combination with the multivariate data analysis revealed that the triantennary glycan with two fucoses of hAGP could have in the future a relevant role in the differentiation of patients with PDAC from those with ChrP.

**Significance:** Multivariate data analysis is currently being used in many omics fields for biomarker discovery. However, to date, no glycomics studies have applied chemometric tools combined with mass spectrometry in a preclinical research. In this work, this methodology has been used to identify altered glycosylation of human alpha-acid glycoprotein in pancreatic ductal adenocarcinoma (PDAC). The obtained results reveal that the triantennary glycan with two fucoses could have a great biomarker potential as it was relevant to differentiate PDAC and chronic pancreatitis (ChrP) patients.

### 1. Introduction

Pancreatic ductal adenocarcinoma (PDAC) is a highly lethal disease considered the fourth leading cause of cancer-related deaths in the world [1]. This pathology has the lowest 5-year survival rate (around 7%) due to the late-stage presentation, the lack of effective treatments and the absence of a thoroughly useful biomarker for its early detection [2–4]. The carbohydrate antigen 19-9 (CA 19-9) is currently the only biomarker recommended for PDAC by the National Comprehensive Cancer Network (NCCN) guidelines for treatment monitoring and relapse follow-up. Nonetheless, CA 19-9 presents an inadequate sensitivity, false negative

results in population with Lewis a-b- genotype and high false results in patients suffering from other nonmalignant diseases such as chronic pancreatitis (ChrP) [5,6]. Therefore, PDAC diagnosis by elevated levels of CA 19-9 should be confirmed by imaging techniques or using other carbohydrate antigens (CEA, CA 125 or CA 242).

Several studies have shown a close relationship between altered glycosylation and tumor initiation, progression and metastasis, which could be used for early diagnosis, disease monitoring and prognosis [7–12]. Most of glycan alterations have been associated with under or overexpression of glycosyltransferases that results generally in sialylation and fucosylation changes. Elevated serum total sialic acid (TSA), different expression

\* Corresponding author.

E-mail address: [estelagimenez@ub.edu](mailto:estelagimenez@ub.edu) (E. Giménez).

<https://doi.org/10.1016/j.jprot.2019.01.006>

Received 2 August 2018; Received in revised form 24 December 2018; Accepted 7 January 2019

Available online 11 January 2019

1874-3919/ © 2019 Elsevier B.V. All rights reserved.

between  $\alpha 2$ -3/ $\alpha 2$ -6 linked sialic acids, increased branching, addition of poly *N*-acetylglucosamine (polyNac) as well as upregulation of certain fucosylated epitopes have been reported in carcinomas [13–17]. Thus, early reports suggest that finding novel glycan-based biomarkers may have much better diagnostic performance [18]. Nevertheless, one of the major bottlenecks in patho-glycomics is the limited number of bioinformatic tools. Multivariate data analysis is currently being used in many *omics* fields such as metabolomics to data exploration, classification and biomarker identification [19]. Regarding glycoproteomics, there are few studies applying chemometric tools in glycan analysis [20–22] and to the best of our knowledge, none was combined with mass spectrometry data in a clinical study. We already showed the great potential of partial least squares discriminant analysis (PLS-DA) to study the glycopeptide glycoforms of human transferrin altered in congenital disorders of glycosylation (CDGs) [23]. This chemometric approach could be even more valuable when applied to the analysis of glycans as their large microheterogeneity with the presence of several isomers, requires an efficient data processing and interpretation tool of such huge and complex datasets.

Acute phase proteins (APP) can be potential markers as they show changes in both protein levels and glycan modifications in response to inflammatory processes and other diseases such as cancer [24,25]. Human alpha-acid glycoprotein (hAGP) is a positive APP which presents high *N*-linked carbohydrate content (45%, w/w) including bi-, tri- and tetraantennary sialylated and sialofucosylated glycan structures. Altered hAGP glycosylation was observed in several cancer types, including pancreatic cancer, resulting in sialyl-Lewis X epitope (NeuAc ( $\alpha 2$ -3)-Gal( $\beta 1$ -4)-(Fuc( $\alpha 1$ -3))-GlcNAc) formation in many cases, which contributes to tumor cell migration towards distant tissues and metastasis promotion [24,26,27]. In previous studies of our group, sialic acid and fucose linkage-type isomers of most hAGP glycans were characterized using exoglycosidase digestions and tandem mass spectrometry [28,29]. Moreover, hAGP glycans from patients with PDAC and ChrP were analyzed and the glycosylation patterns were compared and relatively quantified (pathological versus healthy control) using glycan reductive isotope labeling (GRIL) strategy [30,31]. In these studies, upregulation of certain sialofucosylated structures, that could be used to differentiate PDAC from ChrP, and differences in the expression of some glycan isomers were observed since initial stages of PDAC.

In the present work, intact and desialylated hAGP glycan isomers obtained from purified hAGP from serum samples of patients with PDAC and ChrP were labeled with aniline and analyzed by  $\mu$ ZIC-HILIC-MS to properly correlate the under or overexpression of certain glycan isomers with their sialic acid and fucose linkage-types. In this study, the GRIL methodology with [ $^{12}\text{C}_6$ ]/[ $^{13}\text{C}_6$ ]-aniline [32] was used to perform a reliable relative quantification isomer by isomer (area of the glycan isomer in the pathological sample versus area of the same glycan isomer in the healthy control sample). The high variability among pathological serum samples together with the large number of glycan isomers identified made necessary the use of multivariate data analysis tools to reduce the complexity of data interpretation. Therefore, PLS-DA was applied to the relative area of all hAGP glycan isomers in order to identify the most meaningful isomers in the differentiation of patients with PDAC from healthy controls (HC) and from patients with ChrP. For the first time, multivariate data analysis has been used in a glycomics study together with the relative quantification of intact and desialylated glycan isomers. This combined strategy allowed us to propose novel potential hAGP glycan isomers that could be biomarker candidates of PDAC, some of them involved in its progression which could be useful for studying the response to treatment as well as for disease monitoring.

## 2. Materials and methods

### 2.1. Chemicals

All chemicals used in the preparation of buffers and solutions were of analytical reagent grade. Acetic acid (HAc, glacial), formic acid (FA

98–100%), dimethylsulphoxide (DMSO) and acetone were supplied by Merck (Darmstadt, Germany). Sodium phosphate dodecahydrated ( $\text{Na}_3\text{PO}_4 \cdot 12\text{H}_2\text{O}$ ), sodium cyanoborohydride ( $\text{NaBH}_3\text{CN}$ ), [ $^{12}\text{C}_6$ ]-aniline, [ $^{13}\text{C}_6$ ]-aniline, 2-mercaptoethanol ( $\beta$ -ME) and sodium dodecyl sulfate (SDS) were purchased from Sigma-Aldrich (St. Louis, MO, USA). Hydrogen chloride (HCl, 37%) was supplied by Panreac (Barcelona, Spain) and “NP-40 alternative” by Calbiochem (Darmstadt, Germany). Ammonium acetate ( $\text{NH}_4\text{Ac}$ ), acetonitrile (ACN) and water LC-MS quality grade, used for  $\mu$ ZIC-HILIC-MS analysis, were obtained from Merck and Fluka (Madrid, Spain), respectively. Human alpha-1-acid-glycoprotein standard (hAGP, 99%, purified from healthy control sera) was purchased from Sigma-Aldrich. Peptide *N*-glycosidase F (PNGase F) and  $\alpha 2$ -3,6,8 neuraminidase (total sialidase) were obtained from Roche Diagnostics (Basel, Switzerland).

### 2.2. Serum samples

Control and pathological human serum samples were provided by Hospital Universitari Dr. Josep Trueta, Girona, Spain, following the standard procedures of its Ethics Committee. Serum samples comprise six healthy controls (HC), six patients with chronic pancreatitis (ChrP) and nineteen with pancreatic ductal adenocarcinoma (PDAC); six of them with resectable disease (stages IB, IIA and IIB) and thirteen at advanced stages, 7 locally advanced (stage III) and 6 with metastatic disease (stage IV). There are both females (11) and males (14) patients in the age ranged between 45 and 73 years old (Table 1). The ChrP and the stage of PDAC were diagnosed by the Digestive and Pathology Units using biopsy or image examination. On the other hand, we analyzed the patient variables age and gender of the different groups of patients (healthy control, chronic pancreatitis and pancreatic cancer stages) using the T-student and Chi square tests, respectively, and these variables did not present significant differences within patient groups. Therefore, these variables cannot discriminate among the groups of patients and are not useful to answer the clinical question.

### 2.3. Immunopurification of hAGP

Control and pathological serum samples (0.1 mL) were firstly incubated with 1% (v/v) Protease Inhibitor Cocktail for 30 min at  $-20^\circ\text{C}$  and diluted to 0.2 mL with Milli-Q water before injecting the samples in the immunaffinity chromatography (IAC) system [33]. The IAC-column was prepared as reported in [33], using a polyclonal anti-hAGP antibody bonded to an epoxy-silica chromatographic support from Waters (Protein-Pak epoxy-activated affinity products). Äkta-PPLC instrument (GE Healthcare, Waukesha, WI, USA) equipped with a UV-Vis detector was used for serum immunopurification following a two-step procedure. After the corresponding conditioning of the column using phosphate buffered saline (PBS) (10 mM sodium phosphate, 138 mM NaCl and 2.7 mM KCl, pH 6.85) over 10 min, it was cleaned by injecting three times 1.4 mL of the desorption solution (0.1 M glycine-HCl buffer, pH 2.2). Pretreated serum sample (0.2 mL) was injected and, when the absorbance signal was returned to the baseline level, the retained compounds were eluted using the desorption solution. The eluted compounds were collected and neutralized by addition of 0.1 M  $\text{Na}_2\text{HPO}_4$  (F0 fraction). Subsequently, fraction F0 was re-injected into the IAC column (as described in detail in [34]), and cleaned again by injecting 1.4 mL of desorption solution three times, i.e., at 10 min, 25 min and 40 min after injection, respectively. The corresponding eluted fractions (called F1, F2 and F3) were collected, neutralized by addition of 0.1 M  $\text{Na}_2\text{HPO}_4$ , desalted and finally concentrated by using centrifuge filter devices with 3 kDa cut off membrane (Amicon Ultra 0.5 mL 3K). The purity of these eluted fractions was evaluated by analyzing these fractions by SDS-PAGE (10% acrylamide gels) and silver staining. F3 fraction was discarded and, F1 and F2 were mixed. After IAC, the mixture of F1 and F2 fractions, containing the total purified hAGP, was quantified using a Nanodrop spectrophotometer

**Table 1**  
Clinical and pathological characteristics of the serum samples analyzed.

| Sample       | Stage <sup>a</sup> | Gender | Age | Intact glycan analysis | Desialylated glycan analysis |
|--------------|--------------------|--------|-----|------------------------|------------------------------|
| HC 44        | –                  | Male   | 59  | ✓                      |                              |
| HC 42        | –                  | Female | 63  |                        | ✓                            |
| HC 48        | –                  | Male   | 62  | ✓                      |                              |
| HC 98        | –                  | Male   | 57  |                        | ✓                            |
| HC 51        | –                  | Male   | 66  | ✓                      |                              |
| HC 200       | –                  | Female | 57  | ✓                      |                              |
| ChrP 131     | –                  | Male   | 59  | ✓                      |                              |
| ChrP 141     | –                  | Female | 62  | ✓                      |                              |
| ChrP 116     | –                  | Male   | 62  | ✓                      |                              |
| ChrP 101     | –                  | Female | 49  | ✓                      |                              |
| ChrP 118     | –                  | Male   | 65  | ✓                      |                              |
| ChrP 177     | –                  | Female | 60  | ✓                      |                              |
| PDAC IB 153  | IB (T2N0M0)        | Male   | 63  | ✓                      | ✓                            |
| PDAC IIA 164 | IIA (T3N0M0)       | Female | 62  | ✓                      | ✓                            |
| PDAC IIB 150 | IIB (T3N1M0)       | Male   | 57  | ✓                      | ✓                            |
| PDAC IIA 230 | IIA (T3N0M0)       | Male   | 61  | ✓                      | ✓                            |
| PDAC IIB 81  | IIB (T3N1M0)       | Male   | 62  | ✓                      | ✓                            |
| PDAC IIB 86  | IIB (T3N1M0)       | Female | 53  | ✓                      | ✓                            |
| PDAC III 157 | III (T4NxM0)       | Male   | 58  |                        | ✓                            |
| PDAC III 135 | III (T4NxM0)       | Female | 59  |                        | ✓                            |
| PDAC III 107 | III (T4NxM0)       | Male   | 52  |                        | ✓                            |
| PDAC III 158 | III (T4NxM0)       | Female | 69  |                        | ✓                            |
| PDAC III 47  | III (T4N1M0)       | Male   | 67  | ✓                      | ✓                            |
| PDAC III 155 | III (T4N0M0)       | Male   | 55  | ✓                      | ✓                            |
| PDAC III 103 | III (T4N0M0)       | Male   | 66  | ✓                      | ✓                            |
| PDAC IV 112  | IV (TxNxM1)        | Male   | 52  |                        | ✓                            |
| PDAC IV 109  | IV (T3N1M1)        | Female | 67  |                        | ✓                            |
| PDAC IV 30   | IV (TxNxM1)        | Female | 61  |                        | ✓                            |
| PDAC IV 90   | IV (T3N1M1)        | Female | 45  | ✓                      | ✓                            |
| PDAC IV 170  | IV (T3N1M1)        | Male   | 73  | ✓                      | ✓                            |
| PDAC IV 84   | IV (T3N1M1)        | Female | 62  | ✓                      | ✓                            |

HC, healthy control; ChrP, chronic pancreatitis; PDAC, pancreatic ductal adenocarcinoma.

<sup>a</sup> TNM system was used to determine the stage of PDAC.

(Thermo Scientific) by measuring absorbance at 278 nm and using the parameter  $E1\% = 8.93$ , following the recommendations of the manufacturer for hAGP standard. hAGP samples were evaporated to dryness by Speed Vac and stored at  $-20\text{ }^{\circ}\text{C}$  until its use.

#### 2.4. N-glycan preparation with [ $^{12}\text{C}_6$ ]/[ $^{13}\text{C}_6$ ] aniline labeling

hAGP standard as well as hAGP purified from serum samples were reduced with 0.5%  $\beta$ -ME in the presence of 0.5% of SDS in 50 mM  $\text{Na}_3\text{PO}_4$  (pH 7.5) and boiled in a water bath for 30 min. When samples were at room temperature, a volume of 50 mM  $\text{Na}_3\text{PO}_4$  (pH 7.5) with 1% (v/v) of NP-40 alternative was added to achieve a final concentration of 0.1% of SDS and  $\beta$ -ME in the samples. In order to release the N-glycans, 1  $\mu\text{L}$  of PNGase F (1 U) solution was added and the mixture was carefully vortexed and incubated at  $37\text{ }^{\circ}\text{C}$  for 18 h. Afterwards, 1  $\mu\text{L}$  more of PNGase F was added and the incubation was continued for additional 18 h, to ensure the total deglycosylation of hAGP samples. *Neuraminidase digestion*: desialylated hAGP N-glycans were obtained by using  $\alpha$ -2,3,6,8 neuraminidase (total sialidase). Total sialidase digestion was performed in conjunction with the second cycle of PNGase F digestion by adding also 1  $\mu\text{L}$  of  $\alpha$ -2,3,6,8 neuraminidase (50 mU) and incubated at  $37\text{ }^{\circ}\text{C}$  for 18 h. Digestions were stopped by adding  $\sim 3\text{ }\mu\text{L}$  of FA and samples were stored at  $-20\text{ }^{\circ}\text{C}$  until its use.

Released intact or desialylated N-glycans were purified by solid phase extraction (SPE) using Hypercarb cartridges (25 mg, 1 mL volume, Thermo Fisher Scientific). SPE cartridges were firstly conditioned and equilibrated with 1 mL of 60% ACN, 0.1% FA and with 2 mL of water, respectively. Digested sample was loaded to the SPE cartridge diluted in  $\sim 500\text{ }\mu\text{L}$  of water and then rinsed with 1 mL of water. Retained N-glycans were eluted with 600  $\mu\text{L}$  of 60% ACN, 0.1% FA and the eluate was evaporated to dryness by Speed Vac. Dried N-glycans were stored at  $-20\text{ }^{\circ}\text{C}$  until used. The labeling was carried out by adding 10  $\mu\text{L}$  of reaction mixture (0.35 M aniline and 1 M  $\text{NaCNBH}_3$  in

DMSO with 30% HAC) to the dried glycans and incubating the mixture in a water bath for 2 h at  $70\text{ }^{\circ}\text{C}$ . hAGP pathological samples were labeled with [ $^{12}\text{C}_6$ ]-aniline while hAGP standard and HC samples were labeled with [ $^{13}\text{C}_6$ ]-aniline. After incubation, samples were cooled to room temperature and labeled glycans were precipitated with acetone as described in [32]. Subsequently, equimolar mixtures of hAGP pathological sample and hAGP standard or hAGP HC pool (composed of the same amount of hAGP purified from HC44, HC48, HC51 and HC200, Table 1) were prepared. Finally, the obtained mixtures were evaporated to dryness by Speed Vac and, the dried N-glycans were stored at  $-20\text{ }^{\circ}\text{C}$  until analysis. Centrifugations were performed in a Mikro 220R centrifuge (Hettich Zentrifugen, Tuttlingen, Germany).

#### 2.5. $\mu\text{ZIC-HILIC-MS}$

1200 Series capillary liquid chromatography system coupled to a 6220 oa-TOF LC/MS mass spectrometer with an orthogonal G1385-44300 interface (Agilent Technologies) were used to perform  $\mu\text{LC-MS}$  experiments. ZIC-HILIC column packed with 3.5 mm particles,  $150 \times 0.3\text{ mm L}_T \times \text{ID}$  (SeQuant, Umeå, Sweden) was used for chromatographic separations; the stationary phase consisted of a surface with immobilized zwitterionic sulfobetaine moieties. Experiments were performed at room temperature with gradient elution at a flow rate of 4  $\mu\text{L}/\text{min}$  and injecting 0.25  $\mu\text{L}$  of each glycan sample (glycan concentration:  $\sim 100\text{--}50\text{ pmol}/\mu\text{L}$ ). Eluting solvents were A: 1 mM  $\text{NH}_4\text{Ac}$  solution and B: acetonitrile. The following gradient conditions were used: solvent B from 90% to 80% (within 5 min) and from 80% to 65% (within 20 min) as linear gradient, followed by cleaning and equilibration steps of B: 65%  $\rightarrow$  50% (within 5 min), 50%  $\rightarrow$  0% (within 5 min), 0% (over 15 min), 0%  $\rightarrow$  90% (within 5 min) and 90% (over 10 min). The mass spectrometer was equipped with a dual-nebulizer ESI source, and the orthogonal nebulizer was used for the  $\mu\text{LC-TOF-MS}$  experiments; the second nebulizer, which is generally used to introduce

the internal reference mass standard solution in conventional LC-MS experiments, was disabled to avoid any interference with the  $\mu$ LC-TOF-MS experiments [35]. Tuning and calibration of the mass spectrometer were carried out in accordance with the manufacturer's instructions. The measurement parameters were previously fine-tuned by direct infusion of maltohexaose labeled with [ $^{12}\text{C}_6$ ]-aniline to maximize the signal for the singly charged molecular ion [32]. The optimal operational conditions established in negative mode were: capillary voltage –3500 V, drying gas ( $\text{N}_2$ ) temperature 200 °C, drying gas flow rate 4 L/min, nebulizer gas ( $\text{N}_2$ ) 15 psi, fragmentor voltage 190 V, skimmer voltage 70 V and OCT 1 RF  $V_{pp}$  voltage 300 V. Data were collected in profile (continuum) at 1 spectrum/s (approx. 10,000 transients/spectrum) between  $m/z$  100 and 3200, at the highest resolution mode (4 GHz). MassHunter Workstation software (Agilent Technologies) was used for  $\mu$ LC-MS control, data acquisition and analysis.

## 2.6. Multivariate data analysis

$\mu$ LC-MS data collected for hAGP standard and hAGP purified from serum samples were processed to obtain the extracted ion chromatograms (EIC) of all glycans. Peak areas of the *N*-glycan isomers were measured from the EICs and used to calculate its relative area applying a GRIL strategy with [ $^{12}\text{C}_6$ ]/[ $^{13}\text{C}_6$ ] aniline (i.e. area of the glycan isomer in the pathological sample, labeled with [ $^{12}\text{C}_6$ ]-aniline, divided by the area of the same glycan isomer in the healthy sample, standard or control pool, labeled with [ $^{13}\text{C}_6$ ]-aniline). Relative areas of glycan isomers were used to build a matrix for multivariate data analysis (first for intact glycans and afterwards for desialylated glycans). The relative areas of each matrix were autoscaled (mean centered and scaled to unit standard deviation). Principal component analysis (PCA) was performed to explore the data for different classes and detect the presence of outliers [36]. Partial least squares discriminant analysis (PLS-DA) was applied afterwards to maximize class separation and identify which glycan isomers were the most meaningful to discriminate between classes taking into account the variable importance in the projection (VIP) scores [37,38]. Venetian blinds cross validations of the PLS-DA model was performed during calibration [39]. SOLO (Version 8.2, student edition, Eigenvector Research Inc., Wenatchee, WA, USA) was used for PCA, PLS-DA and VIP calculations. Nomenclature used for glycans correspond to their composition, in terms of number of hexoses (H), *N*-acetylglucosamines (N), fucoses (F) and sialic acids (S), followed by an index that indicates the isomer number.

## 3. Results and discussion

### 3.1. Analysis of intact glycan isomers

hAGP glycan isomers isolated from serum samples of patients with chronic pancreatitis (ChrP) and pancreatic ductal adenocarcinoma (PDAC) at different stages of the pathology were analyzed. Six ChrP samples, twelve PDAC samples (six at stages I and II, three at stage III and three at stage IV) as well as four healthy controls (HC), employed to prepare a HC pool, were used for this purpose. The reference, stage (using the 7th Edition of the TNM Classification of Malignant Tumors), gender and age of the individual of each serum sample are shown in Table 1. The characterization of hAGP glycan isomers performed in our previous works (Table S1 of the Supplementary material), using exoglycosidase digestions and tandem mass spectrometry [28,29], was used in this paper to correlate alterations observed in certain isomers with the corresponding linkage-type of the sialic acid and the fucose as well as the structure they present. The aim of this study is to find glycan isomers that could be used to discriminate patients with PDAC, since initial stages of the pathology, from those suffering from ChrP.

The isotope-coded labeling approach established previously using [ $^{12}\text{C}_6$ ]/[ $^{13}\text{C}_6$ ]-aniline [32], was used to unequivocally identify alterations in hAGP glycosylation comparing directly the glycan profile and

to reliably determine the under or overexpression of its isomers between control and patient samples. Equimolar mixtures of hAGP glycans from each pathological serum sample (labeled with [ $^{12}\text{C}_6$ ]-aniline) and hAGP glycans from HC pool (labeled with [ $^{13}\text{C}_6$ ]-aniline) were prepared and subsequently analyzed by  $\mu$ ZIC-HILIC-MS. The HC pool was composed of the same amount of hAGP purified from HC44, HC48, HC51 and HC200 (Table 1). Relative areas (pathological sample vs. HC pool) calculated for all detected glycan isomers in the different pathological samples (PDAC and ChrP patients) were difficult to be visually compared taking into account the large number of isomers identified (42 glycan isomers) and the variability obtained among serum samples. In order to reduce the complexity of data interpretation, the pathological samples were classified in PDAC or ChrP samples and, the average of the relative areas obtained from each hAGP glycan isomer in each group was calculated. Fig. 1A shows a graphic bar with these results. Standard deviations (SD) obtained in each isomer are also figured in the graphic. The overexpression of fucosylated glycan isomers, with respect to HC pool, was observed in both PDAC and ChrP patients but this upregulation was much higher in PDAC samples. In addition, generally, the glycan isomers presenting more than one fucose unit (H6N5F2S3, H7N6F2S3\_1, H7N6F2S3\_2, H7N6F2S4 and H7N6F3S4) were those more overexpressed in PDAC with respect to ChrP, and H9N8S3 glycan, which presents one extra *N*-acetylglucosamine (LacNAc) unit. On the other hand, some glycan isomers previously assigned with all or most  $\alpha$ -2-6 linked sialic acids (e.g. H6N5S3\_3, H7N6F1S3\_3 and H7N6F1S3\_4) were also more upregulated in PDAC compared to ChrP patients. By way of an example, Fig. 1B shows the extracted ion chromatograms (EICs) of H6N5S3 obtained in a PDAC sample, at initial stage of the pathology (PDAC IIA 230, see Table 1), and in a ChrP sample (ChrP I41) with respect to HC pool. The PDAC sample clearly shows different behavior between isomers: a decrease of isomers 1 and 2 previously assigned with one  $\alpha$ -3 linked sialic acid (SIA) and two  $\alpha$ -6 linked SIA, and a considerable overexpression of the third isomer (H6N5S3\_3), assigned with all  $\alpha$ -2-6 linked SIA [28,29]. Moreover, regarding H7N6F1S4 (see Fig. 1A and C), it was observed that isomers 2 and 3 were also highly upregulated in PDAC with respect to HC pool and ChrP patients. Even though isomer 3 of H7N6F1S4 (H7N6F1S4\_3) was not previously characterized [28,29] (due to its low abundance in hAGP HC and standard samples), it was assumed that this isomer presents higher proportion of  $\alpha$ -2-6 linked SIA than isomer 2, as it was more retained in the ZIC-HILIC column [28]. Therefore, certain relation seemed to exist between the overexpression of  $\alpha$ -2-6 linked SIA in hAGP glycans and PDAC, as was reported in other cancer types [8,13,15].

Although some trends could be observed, the SD values of Fig. 1A demonstrated the high variability obtained within the two patient groups (PDAC and ChrP) making necessary to compare all PDAC and ChrP samples directly, instead of using group averages. For this purpose, and due to the high amount of samples and glycan isomers to be compared, a multivariate data analysis approach was applied for data interpretation. First, PCA was used to explore the data for the unsupervised identification of trends and detection of outliers. The outlier detection plot (Q Residuals vs. Hotelling  $T^2$ ) of Fig. S1 of the Supplementary Material shows that PDAC IIB 86 was badly described by the model and thus, it was removed for the following multivariate data analysis. Subsequently, a partial least squares discriminant analysis (PLS-DA) model was built, classifying samples in PDAC and ChrP patients, to improve class separation and to identify which glycan isomers could be potentially used to differentiate these pathologies. As can be seen in the scores plot of Fig. 2A, two latent variables (LVs) allowed discrimination between the two groups (63 and 43% of the X and Y variances explained, respectively). Only PDAC IIA 164 and ChrP 131 presented an abnormal behavior, as was previously reported in [31] (named PDAC2 and ChrP1, respectively), and they were wrongly classified. The loadings plot (Fig. 2B) showed the contribution of each variable (i.e. each glycan isomer) to the different LV. As can be seen, fucosylated glycan isomers were more related to PDAC (they appeared

**Fig. 1.** A) Graphic bar of the average of the relative areas (pathological sample vs. healthy control pool) obtained for each hAGP intact glycan isomer in PDAC (red bars) and ChrP (green bars) samples with its corresponding standard deviation associated. B) Extracted ion chromatograms (EICs) obtained for H6N5S3 glycan in PDAC IIA 230 (i) and ChrP 141 (ii) samples. C) EICs of H7N6F1S4 glycan obtained in PDAC IIA 230 (i) and ChrP 141 samples. EICs of glycans in pathological samples are shown in red and EICs of glycans in HC pool are shown in black. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

in the negative side of the x-axis, in the same region as most of PDAC samples, see the scores plot in Fig. 2A). In addition, the VIP scores allowed us to quantify the influence of the different glycan isomers on the separation between PDAC and ChrP samples. In this regard, as can be seen in Fig. 2C, sixteen out of forty-two glycan isomers, most of them fucosylated, were important ( $VIP > 1$ ) to distinguish between these pathologies, especially the seven ones with VIP values higher than 1.2. Among these seven glycan isomers considered the most meaningful, five presented at least one fucose unit and all of them were characterized with most or all  $\alpha$ -2-6 linked SiA (H6N5S3\_1 and H6N5F2S3 were assigned with two  $\alpha$ -2-6 linked SiA and one  $\alpha$ -2-3 SiA while H7N6S3\_4, H6N5F1S2\_2, H7N6F1S2\_2 and H7N6F2S3\_2 were assigned with all  $\alpha$ -2-6 linked SiA, see Table S1 of the Supplementary Material) [28,29]. Moreover, these fucosylated glycan isomers with high proportion of  $\alpha$ -2-6 sialic acids (marked in red in Fig. 2B) seemed to be more associated to PDAC as they appeared in the same region as PDAC samples (compare the scores and loadings plots of Fig. 2A and B, respectively). For this reason, the formation of sialyl-Lewis X (SLe<sup>x</sup>) epitope with an  $\alpha$ -2-3 linked SiA, related to cancer metastasis [12] and previously described in hAGP [24,26,27], was not considered important for the discrimination between these two pancreatic diseases. H8N7F1S4 was also considered significant ( $VIP > 1.2$ ) but it was not previously characterized due to its low abundance [29]. Although other glycan isomers were highly overexpressed in certain PDAC samples (e.g. H6N5S3\_3 and H7N6F1S4\_3) and seemed to be very relevant in the differentiation between PDAC and ChrP patients when group averages were evaluated (see Fig. 1A), the multivariate data analysis revealed other glycan isomers that could be more reliable biomarkers, and they were masked when group averages were compared due to the high variability among samples. This effect can be observed in H6N5S3\_1, H7N6S3\_4 and H7N6F1S2\_2 glycans which show similar ratios in PDAC and ChrP sample groups (see Fig. 1A) but their biomarker potential was disclosed by the PLS-DA model. Using this approach, we concluded that fucosylated hAGP glycan isomers with most  $\alpha$ -2-6 linked SiA are essential in the differentiation between PDAC and ChrP patients and could be relevant hAGP biomarkers to distinguish between both pathologies.

### 3.2. Analysis of desialylated glycan isomers

Once it was confirmed that fucosylated glycans seemed to play an important role in the differentiation between PDAC and ChrP patients, desialylated hAGP glycan isomers obtained from PDAC, ChrP and HC serum samples were analyzed to determine which type of fucosylation was truly related to cancer. In this case, the cohort of samples was larger: six ChrP patients, seventeen PDAC patients (four at stages I and II, seven at stage III and six at stage IV) and three HC (see Table 1 for details about stage, gender and age of these patients). Most of the pathological samples analyzed in the study of intact hAGP glycans (Section 3.1) were also used to study the desialylated glycan isomers to be able to correlate the obtained results in both cases. Relative quantification of glycan isomers with [<sup>12</sup>C<sub>6</sub>]/[<sup>13</sup>C<sub>6</sub>]-aniline was also carried out but, in this case, hAGP glycans obtained from each pathological sample (labeled with [<sup>12</sup>C<sub>6</sub>]-aniline) were mixed (1:1) with hAGP glycans obtained from hAGP standard sample (labeled with [<sup>13</sup>C<sub>6</sub>]-aniline), which shows a healthy glycan profile as reported in [30]. Each HC sample was also relatively quantified with respect to hAGP standard. This GRIL strategy allowed us to reliably determine the up or downregulation of glycan isomers and thereby investigate the fucose linkage-types related to these pathologies. Serum samples were divided in HC, PDAC and ChrP. The average of the relative area obtained for each glycan isomer

in each group of samples was calculated (area of the glycan isomer in the serum sample vs. area of the same glycan isomer in hAGP standard sample). The graphic bar of Fig. 3A depicts the averages obtained for each glycan isomer with the corresponding standard deviation associated. Again, it was observed that most fucosylated glycan isomers were overexpressed in PDAC and ChrP patients with regard to HC samples. However, all glycan isomers assigned in our previous works [28,29] as antennary fucosylated (H5N4F1\_3, H6N5F1\_3, H6N5F2\_3, H7N6F1\_2, H7N6F2\_2 and H7N6F2\_3, marked with an asterisk in A) and some glycan isomers showing core fucosylation (H5N4F1\_2 and H7N6F2\_1) were higher upregulated in PDAC than in ChrP. The EICs of H6N5F1 and H7N6F2 glycans, in PDAC IB 153 and ChrP 141 samples are shown in Fig. 3B and C, respectively (EICs of glycans in pathological samples are shown in red and EICs of glycans in hAGP standard are shown in black). As can be observed, isomers assigned as antennary fucosylated were more overexpressed in PDAC with respect to ChrP (i.e. H6N5F1\_3, H7N6F2\_2 and H7N6F2\_3). In the case of H7N6F2, although isomers 2 and 3 were both only partially assigned with at least one antenna linked fucose, it could be assumed with the retention time that isomer 3 presents a higher proportion of antennary fucosylation than isomer 2 and, therefore, the overexpression of this isomer in PDAC was higher.

As happened with the intact glycans, the variability between pathological samples and the large number of glycan isomers identified made again necessary the use of multivariate data analysis tools instead of calculating group averages, which could lead to misleading conclusions. Firstly, PCA was used to explore the data and, on this occasion, outlier samples were not detected (Fig. S2 of the Supplementary Material). PLS-DA was then applied to study PDAC vs. HC samples (Fig. 4A) and ChrP vs. HC samples (Fig. 4B). As can be observed in both scores plots (Fig. 4A-i and 4B-i), pathological samples were clearly separated from HC. Two latent variables (LVs) explained 47 and 68% of the X and Y variances, respectively, in the differentiation between PDAC and HC samples (Fig. 4A) and two LVs explained 47 and 98% of the X and Y variances, respectively, in the case of ChrP and HC samples (Fig. 4B). The VIP scores showed that H5N4F1\_2, H6N5F2\_3, H7N6F1\_2, H7N6F2\_2 and H7N6F2\_3 glycans were significant in both PDAC and ChrP vs. HC differentiations ( $VIP > 1$ , see Fig. 4A-ii and 4B-ii). These results were not surprising as the overexpression of some isomers with respect to HC was observed in both PDAC and ChrP patients, although to a different extent (usually much higher in PDAC, see the graphic bar of Fig. 3A). By contrast, H5N4F1\_3, H6N5F2\_1, H6N5F2\_2 and H8N7 glycans (marked in red in Fig. 4A-ii) were more interesting as they were only meaningful in the separation between PDAC and HC samples. As the main goal of this study was to identify hAGP glycan isomers that could be used to differentiate PDAC at different stages of the pathology from ChrP, and also, to detect potential glycan biomarkers to follow-up PDAC progression, three different PLS-DA models were built. This strategy comparing ChrP samples with different stages of PDAC (stages I and II were grouped as resectable disease, stage III and stage IV constituted the other two groups as locally advanced and metastatic disease, respectively), also should improve the reliability of the PLS-DA models as similar number of samples would be compared in each group. As can be observed in Fig. 5, the first PLS-DA was applied to classify PDAC at stages I and II and ChrP samples (Fig. 5A), the second to distinguish PDAC at stage III from ChrP (Fig. 5B) and the third to differentiate PDAC patients at stage IV of the pathology from those with ChrP (Fig. 5C). In all three cases, two LVs explained between 37 and 39% and 75–85% of the X and Y variances, respectively, and the scores plots obtained showed two separated

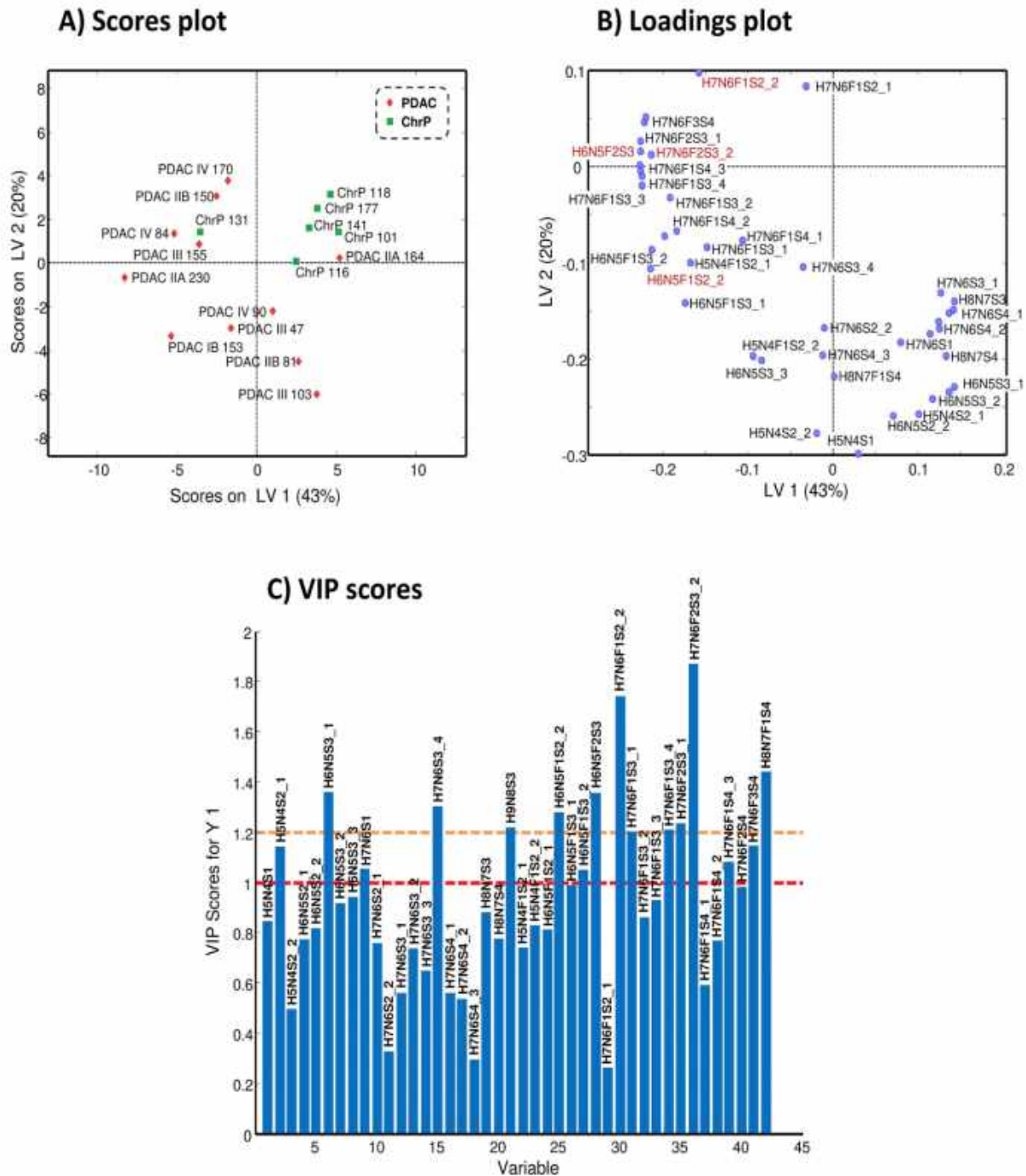
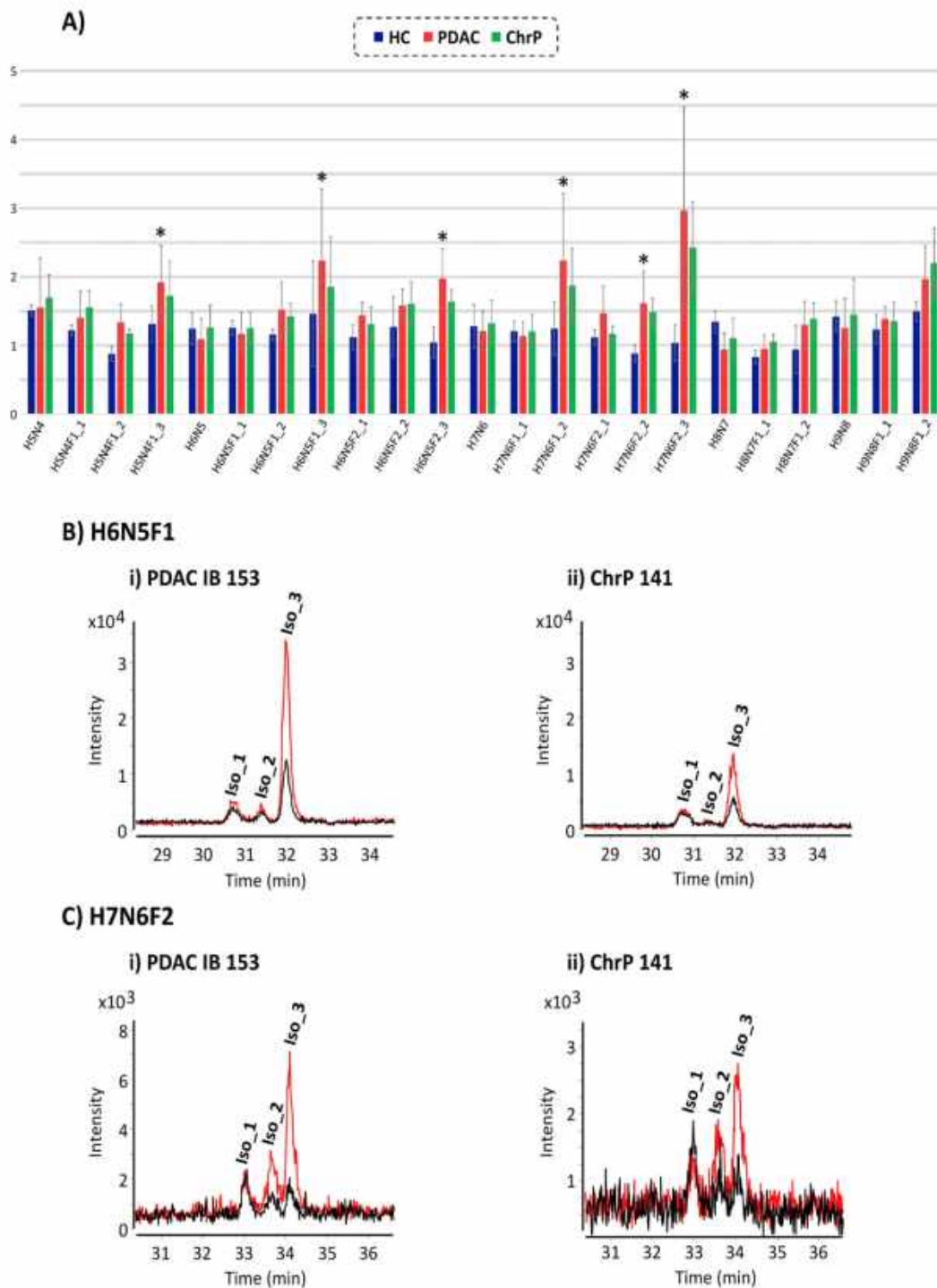


Fig. 2. A) Scores plot, B) Loadings plot and C) VIP scores of the PLS-DA model applied to the relative areas (area of the glycan isomer in the pathological samples vs. area of the same glycan isomer in the healthy control pool) obtained for each hAGP intact glycan isomer in all PDAC and ChrP serum samples analyzed.

groups of samples. Although PDAC IIA 164 and ChrP 131 acted as aberrant samples in the intact glycan analysis (see Fig. 2A), their behavior changed after glycan desialylation and they were well classified in these PLS-DA models (Fig. 5-i). The VIP scores revealed that the above mentioned glycans (H5N4F1<sub>3</sub>, H6N5F2<sub>1</sub>, H6N5F2<sub>2</sub> and H8N7) were important in the differentiation between PDAC and HC samples as well as they could be also biomarkers of certain stages of PDAC with respect to ChrP. H5N4F1<sub>3</sub> was only significant at stages I and II while H6N5F2<sub>2</sub> at stage III of PDAC (Fig. 4A-ii and 4B-ii,

respectively), and H8N7 could be mainly used to discriminate stages I and II of PDAC from ChrP (see Fig. 5A-ii). The multivariate data analysis performed separately for each group of PDAC with respect to ChrP allowed us to find which glycans could be only biomarkers of a particular stage or group of PDAC. In addition, observing the VIP scores obtained for all three PLS-DA models, it was revealed that three out of twenty-three desialylated hAGP glycan isomers were significant to distinguish PDAC at all stages of the pathology from ChrP patients (H5N4F1<sub>2</sub>, H6N5F2<sub>3</sub> and H9N8, marked in red in the VIP scores of

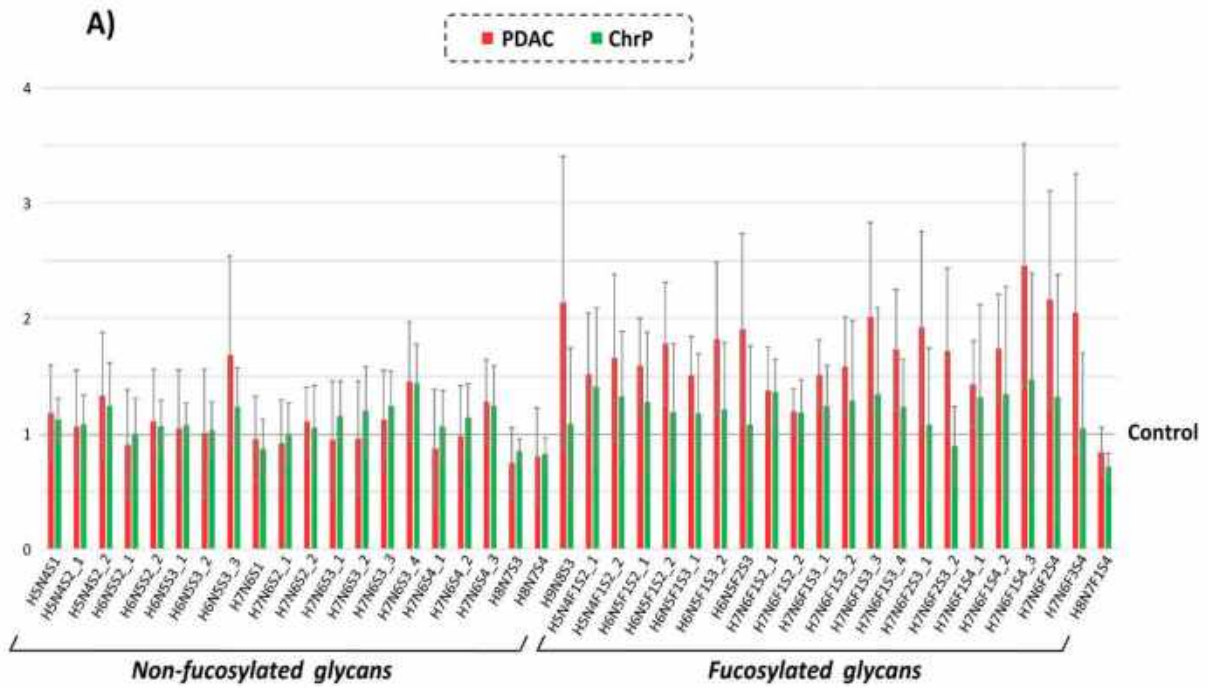


**Fig. 3.** A) Graphic bar of the average of the relative areas (pathological sample vs. hAGP standard sample) obtained for each hAGP desialylated glycan isomer in HC (blue bars), PDAC (red bars) and ChrP (green bars) samples with its corresponding standard deviation associated. B) EICs obtained for H6N5F1 glycan in PDAC IB 153 (i) and ChrP 141 (ii) samples. C) EICs of the H7N6F2 glycan obtained in PDAC IB 153 (i) and ChrP 141 samples. EICs of glycans in pathological samples are shown in red and EICs of glycans in hAGP standard are shown in black. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

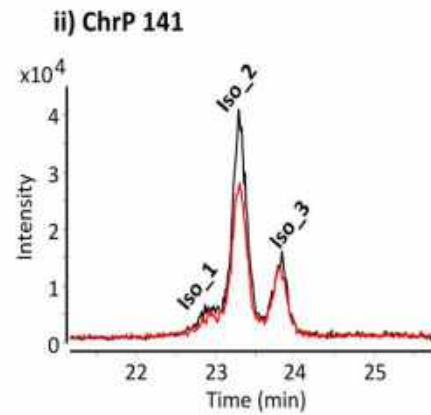
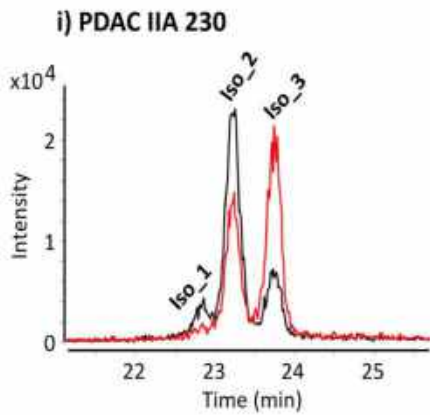
Fig. 5-ii). Although H5N4F1\_2 and H6N5F2\_3 were previously considered meaningful to separate PDAC as well as ChrP samples from HC (Fig. 4-ii), they were also crucial to differentiate PDAC and ChrP pathologies as their overexpression was higher in the case of cancer.

Therefore, H5N4F1\_2 (assigned as core fucosylated), H6N5F2\_3 (assigned with at least one antenna linked fucose) and H9N8, which presents two extra units of LacNAc, could be also potential biomarker candidates of PDAC. Nevertheless, a particular fucose linkage-type

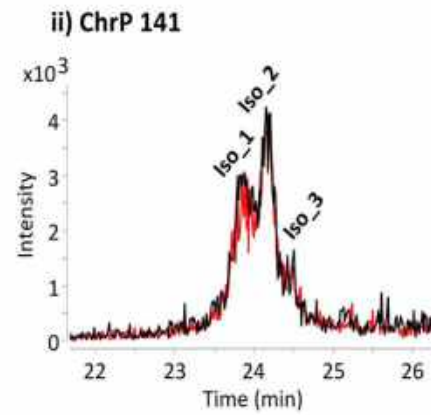
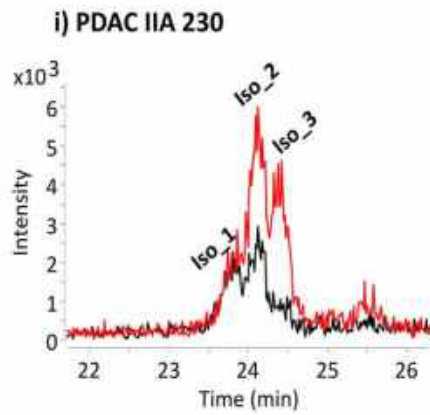




**B) H6N5S3**



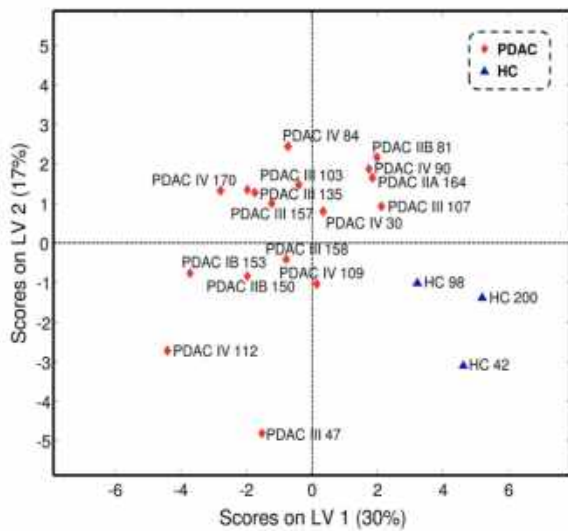
**C) H7N6F1S4**



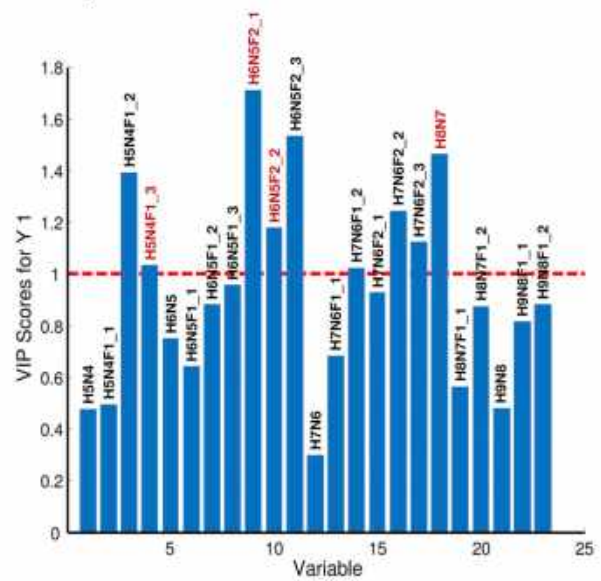
(caption on next page)

**A) PDAC vs. HC**

**i) Scores plot**

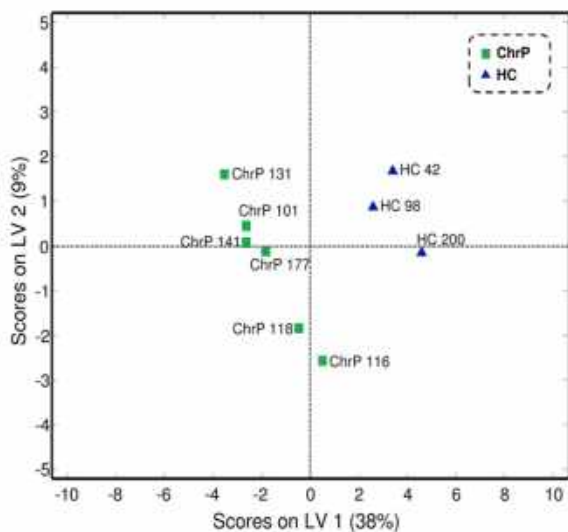


**ii) VIP scores**

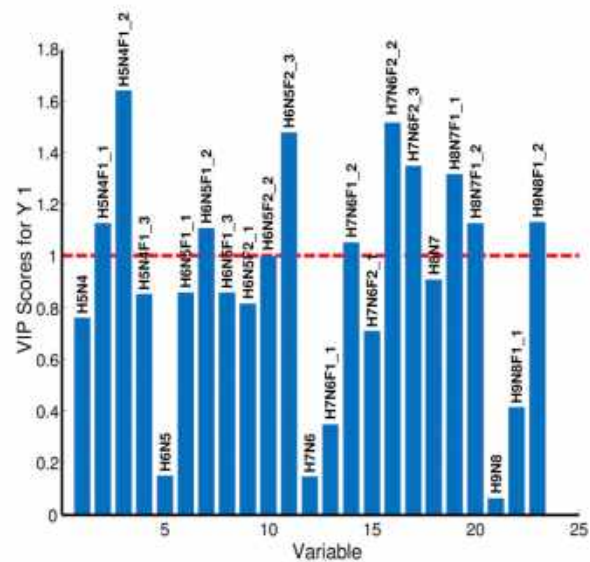


**B) ChrP vs. HC**

**i) Scores plot**



**ii) VIP scores**



**Fig. 4.** Scores plot (i) and VIP scores (ii) of the PLS-DA model applied to the relative areas (area of the glycan isomer in the pathological sample vs. area of the same glycan isomer in hAGP standard sample) obtained in A) PDAC vs. HC samples and B) ChrP vs. HC samples.

involved only in PDAC and, therefore, allowing the differentiation from ChrP patients was not identified. Taking into account the results obtained in both intact and desialylated glycan studies, it was revealed that H6N5F2 glycan (or H6N5F2S3 in the case of intact glycans) could play an important role in the search of a novel biomarker for PDAC

diagnosis as it was significant in all PLS-DA models. In order to corroborate the biomarker value of the proposed glycan a larger cohort of pathological (PDAC and ChrP) and HC samples should be studied in the future.

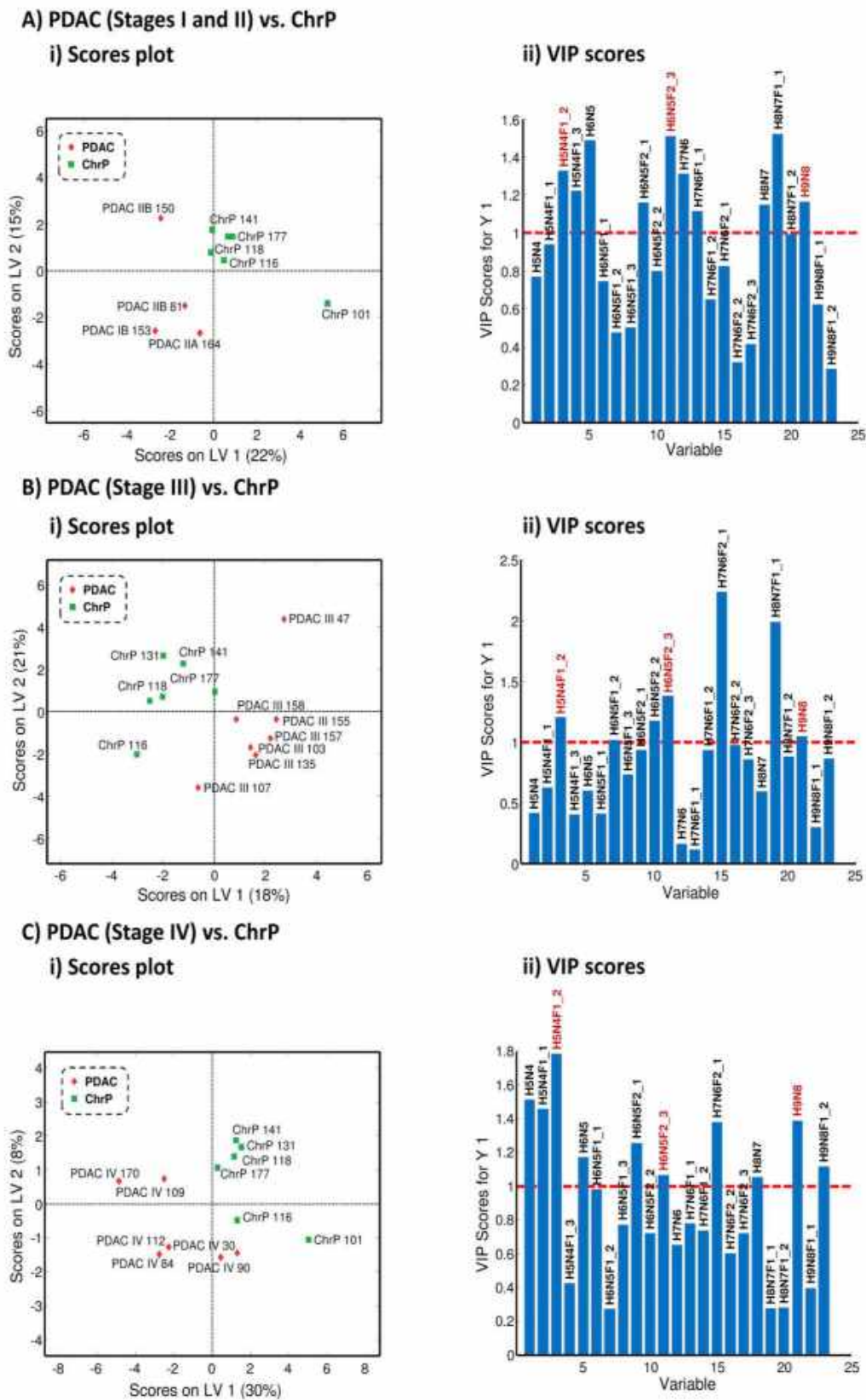


Fig. 5. Scores plot (i) and VIP scores (ii) of the PLS-DA model applied to the relative areas (area of the glycan isomer in the pathological sample vs. area of the same glycan isomer in hAGP standard sample) obtained in A) PDAC at stages I-II vs. ChrP samples, B) PDAC at stage III vs. ChrP samples and C) PDAC at stage IV vs. ChrP samples.

#### 4. Concluding remarks

The potential of multivariate data analysis was demonstrated for glycomic studies when visual data interpretation could result in misleading conclusions due to the high variability between patients and the large amount of glycan isomers to be compared. The PLS-DA model performed for intact hAGP glycans revealed that seven glycan isomers (H6N5S3\_1, H6N5F2S3, H7N6S3\_4, H6N5F1S2\_2, H7N6F1S2\_2, H7N6F2S3\_2 and H8N7F1S4) with  $\alpha$ 2-6 linked SIA, five of them also fucosylated, were the most meaningful to discriminate between PDAC and ChrP patients. On the other hand, in the analysis of desialylated hAGP glycans, the comparison of pathological (PDAC or ChrP) vs. HC samples by PLS-DA showed that H5N4F1\_3, H6N5F2\_1, H6N5F2\_2 and H8N7 glycan isomers were important as they were only significant in the differentiation between PDAC and HC. Furthermore, three out of twenty-three desialylated hAGP glycan isomers (H5N4F1\_2, H6N5F2\_3 and H9N8) were also identified as biomarker candidates because they were meaningful to distinguish all stages of PDAC from ChrP. Nonetheless, a certain fucose linkage-type only related to PDAC, which allowed the differentiation from ChrP patients, was not found. H6N5F2 glycan (or H6N5F2S3 in the case of intact glycans) could have a great biomarker potential as it was relevant in all PDAC and ChrP differentiations using PLS-DA for both intact and desialylated glycans. This study therefore could be the first step to find a hAGP glycan biomarker for PDAC and it could lead to the development of new antibody kits for the identification of these specific glycan structures in serum samples. However, this was a preliminary study and a larger cohort of pathological samples should be analyzed in the future to validate the proposed glycans as reliable biomarkers. The presented multivariate data approach provided a simple and rapid manner to identify the modifications occurred in hAGP glycosylation that could be used for PDAC diagnosis and demonstrated that this methodology could be extremely useful in patho-glycomics to find novel glycan-based biomarkers related to other diseases.

#### Acknowledgments

This study was supported by a grant from the Spanish Ministry of Economy and Competitiveness (CTQ2014-56777-R awarded to V-S-N and BIO2015-66356-R awarded to R.P.). Montserrat Mancera-Arteu acknowledges the University of Barcelona for an ADR fellowship. Meritxell Balmaña acknowledges University of Girona for a pre-doctoral fellowship.

#### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jpro.2019.01.006>.

#### References

- [1] A. Adamska, A. Domenichini, M. Falasca, Pancreatic ductal adenocarcinoma: current and evolving therapies, *Int. J. Mol. Sci.* **18** (2017), <https://doi.org/10.3390/ijms18071338>.
- [2] S. Holst, A.L. Belo, E. Giovannetti, I. Van Die, M. Wuhrer, Profiling of different pancreatic cancer cells used as models for metastatic behaviour shows large variation in their N-glycosylation, *Sci. Rep.* **7** (2017) 1–15, <https://doi.org/10.1038/s41598-017-16811-6>.
- [3] D.S. Sworeds, M.A. Firpo, C.L. Scaife, S.J. Mulvihill, Biomarkers in pancreatic adenocarcinoma: current perspectives, *Oncol. Targets. Ther.* **9** (2016) 7459–7467, <https://doi.org/10.2147/OTT.S100510>.
- [4] J.C. Chang, M. Kundranda, Novel diagnostic and predictive biomarkers in pancreatic adenocarcinoma, *Int. J. Mol. Sci.* **18** (2017) 4–6, <https://doi.org/10.3390/ijms18030467>.
- [5] E. Wu, S. Zhou, K. Bhat, Q. Ma, CA 19-9 and pancreatic cancer, *Clin. Adv. Hematol. Oncol.* **11** (2013) 53–55, <https://doi.org/10.1016/j.biotechadv.2011.08.021>.
- [6] S. Krishnan, H.J. Whitwell, J. Cuenco, A. Gentry-Maharaj, U. Menon, S.P. Pereira, M. Gaspari, J.F. Timms, Evidence of altered glycosylation of serum proteins prior to pancreatic cancer diagnosis, *Int. J. Mol. Sci.* **18** (2017) 8–10, <https://doi.org/10.3390/ijms18122670>.

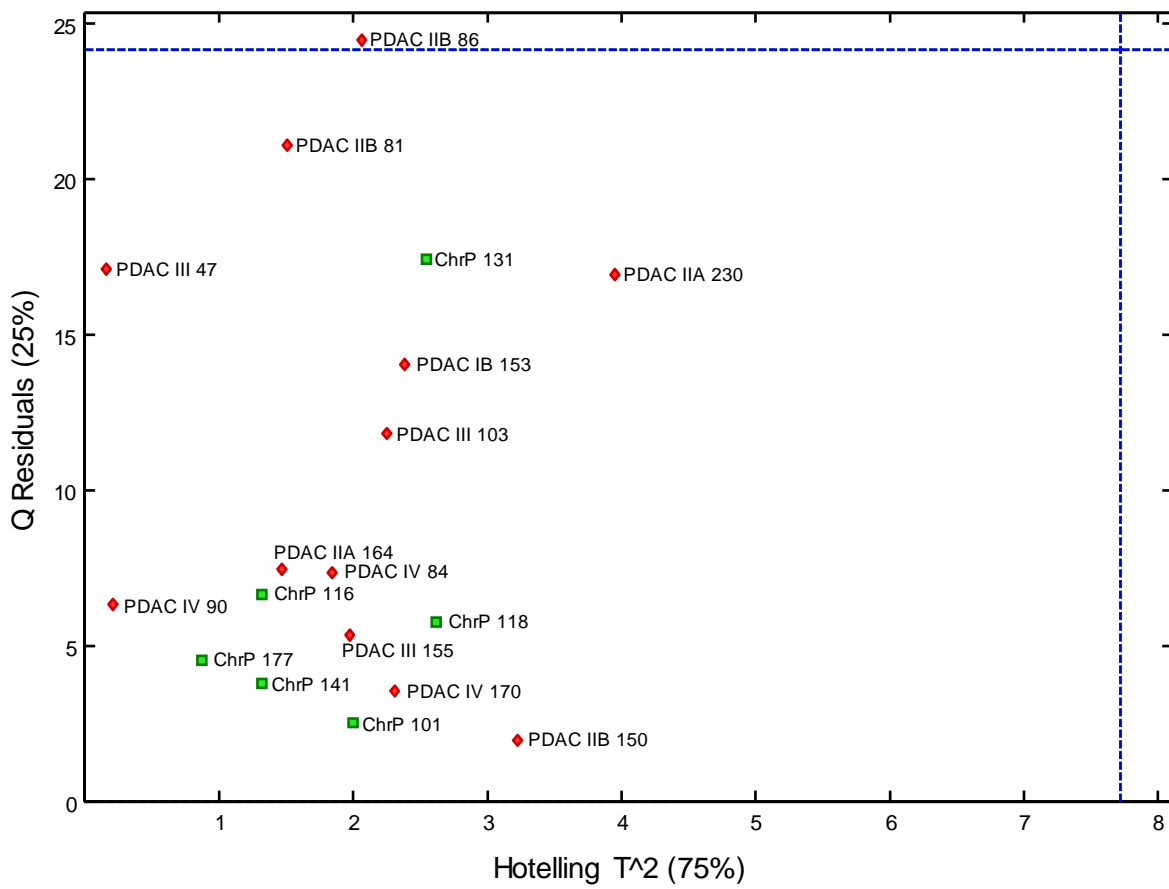
- [7] M.J. Kailemia, D. Park, C.B. Lebrilla, Glycans and glycoproteins as specific biomarkers for cancer, *Anal. Bioanal. Chem.* **409** (2017) 395–410, <https://doi.org/10.1007/s00216-016-9880-6>.
- [8] M.K. Sethi, W.S. Hancock, S. Fanayan, Identifying N-glycan biomarkers in colorectal cancer by mass spectrometry, *Am. Chem. Soc.* **49** (2016) 2099–2106, <https://doi.org/10.1021/acs.accounts.6b00193>.
- [9] N. Taniguchi, Y. Kizuka, Glycans and Cancer: Role of N-Glycans in Cancer Biomarker, Progression and Metastasis, and Therapeutics, 1st ed., Elsevier Inc, 2015, <https://doi.org/10.1016/ba.aer.2014.11.001>.
- [10] B.N. Vajaria, P.S. Patel, Glycosylation: a hallmark of cancer? *Glycoconj. J.* **34** (2017) 147–156, <https://doi.org/10.1007/s10719-016-9755-2>.
- [11] B. Adamczyk, T. Tharmalingam, P.M. Rudd, Glycans as cancer biomarkers, *Biochim. Biophys. Acta - Gen. Subj.* **1820** (2012) 1347–1353, <https://doi.org/10.1016/j.bbagen.2011.12.001>.
- [12] L. Oliveira-Ferreira, K. Legler, K. Milde-Langosch, Role of protein glycosylation in cancer metastasis, *Semin. Cancer Biol.* **44** (2017) 141–152, <https://doi.org/10.1016/j.semcancer.2017.03.002>.
- [13] B.N. Vajaria, K.R. Patel, R. Begum, P.S. Patel, Sialylation: an Avenue to Target Cancer Cells, *Pathol. Oncol. Res.* **22** (2016) 443–447, <https://doi.org/10.1007/s12253-015-0033-6>.
- [14] Z. Lin, D.M. Simone, M.A. Anderson, R.E. Brand, X. Xie, K.A. Shredlen, M.T. Ruffin, D.M. Lubman, Mass spectrometric assay for analysis of haptoglobin fucosylation in pancreatic cancer, *J. Proteome Res.* **10** (2011) 2602–2611, <https://doi.org/10.1021/ps200102h>.
- [15] Y. Mechref, Y. Hu, A. Garcia, S. Zhou, J.L. Desantos-Garcia, A. Hussein, Defining putative glycan cancer biomarkers by MS, *Bioanalysis* **4** (2012) 2457–2469, <https://doi.org/10.4155/bio.12.246>.
- [16] S. Singh, K. Pal, J. Yadav, H. Tang, K. Partyka, D. Kleiter, P. Hoeh, E. Enskik, B. Ke, G. Hostetter, E. Xu, M. Bern, D.F. Smith, A.S. Mehta, R. Brand, E. Meicher, B.B. Hanb, Upregulation of glycans containing 3' fucose in a subset of pancreatic cancers uncovered using fusion-tagged lectins, *J. Proteome Res.* **14** (2015) 2594–2605, <https://doi.org/10.1021/acs.jproteome.5b00142>.
- [17] T.F. Ørnof, E.M. Vestergaard, E. Holmes, J.S. Jakobsen, M. Mortensen, P. Johnson, P. Brøns, N. Gregersen, U.B. Jensen, L. Bolund, H. Wolf, N. Grønnet, K. Skorstengaard, Influence of Lewis  $\alpha$ -3/4-Fucosyltransferase (FUT3) Gene Mutations on Enzyme Activity, Erythrocyte Phenotyping, and Circulating Tumor Marker Sialyl-Lewis  $\alpha$  Levels, *Vol. 271* (1996), pp. 32260–32268, <https://doi.org/10.1074/jbc.271.50.32260>.
- [18] A. Drabik, A. Bodzom-Kulakowska, P. Sudler, J. Silberring, J. Kulig, M. Sierzeza, Glycosylation changes in Serum Proteins Identify patients with Pancreatic Cancer, *J. Proteome Res.* **16** (2017) 1436–1444, <https://doi.org/10.1021/acs.jproteome.6b00775>.
- [19] R. Madsen, T. Lundstedt, J. Trygg, Chemometrics in metabolomics - a review in human disease diagnosis, *Anal. Chim. Acta* **659** (2010) 23–33, <https://doi.org/10.1016/j.aca.2009.11.042>.
- [20] H. Ito Feng, S. Lim, A.K.C. Lacerda, P. Li, X. Yin, E. Simsek, S.H. Khan, S.M. Chen, S.F.Y. Li, High throughput human plasma N-glycan analysis using DNA analyzer and multivariate analysis for biomarker discovery, *Anal. Chim. Acta* **995** (2017) 106–113, <https://doi.org/10.1016/j.aca.2017.09.003>.
- [21] N. Hashii, A. Haraono, R. Kuribayashi, D. Takakura, N. Kawasaki, Characterization of N-glycan heterogeneities of erythropoietin products by liquid chromatography/mass spectrometry and multivariate analysis, *Rapid Commun. Mass Spectrom.* **28** (2014) 921–932, <https://doi.org/10.1002/rcm.6658>.
- [22] L. Zhou, J. Di Xu, S.S. Zhou, H. Shen, Q. Mao, M. Kong, Y.T. Zou, Y.Y. Xu, J. Xu, S.L. Li, Chemomics-based marker compounds mining and mimetic processing for exploring chemical mechanisms in traditional processing of herbal medicines, a continuous study on Behmarmine Radix, *J. Chromatogr. A* **1530** (2017) 232–240, <https://doi.org/10.1016/j.chroma.2017.11.036>.
- [23] A. Barros, 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, <https://doi.org/10.1016/j.talanta.2016.07.055>.
- [24] A. Sarrats, R. Saldova, E. Pla, E. Fort, D.J. Harvey, W.B. Struwe, L.R. De, P.M. Rudd, R. Peracaula, Glycosylation of liver acute-phase proteins in pancreatic cancer and chronic pancreatitis, *Proteomics Clin. Appl.* **4** (2010) 432–448, <https://doi.org/10.1002/prca.200900150>.
- [25] O. Gemik, G. Lucc, Glycosylation of serum proteins in inflammatory diseases, *Dis. Markers* **25** (2008) 267–278.
- [26] F. Cecilian, V. Pomocqua, The acute phase protein  $\alpha$ 1-acid glycoprotein: a model for altered glycosylation during diseases, *Current* (2007) 91–108.
- [27] C.L. Fernandes, R. Ligabue-Braun, H. Verli, Structural glyco-biology of human  $\alpha$ 1-acid glycoprotein and its implications for pharmacokinetics and inflammation, *Glycobiology* **25** (2015) 1125–1133, <https://doi.org/10.1093/glycob/cwv041>.
- [28] M. Mancera-Arteu, E. Giménez, J. Barbosa, R. Peracaula, V. Sanz-Nebot, Zwitterionic-hydrophilic interaction capillary liquid chromatography coupled to tandem mass spectrometry for the characterization of human  $\alpha$ 1-acid-glycoprotein N-glycan isomers, *Anal. Chim. Acta* **991** (2017) 76–88, <https://doi.org/10.1016/j.aca.2017.07.068>.
- [29] M. Mancera-Arteu, E. Giménez, J. Barbosa, V. Sanz-Nebot, Identification and characterization of isomeric N-glycans of human  $\alpha$ 1-acid glycoprotein by stable isotope labelling and ZIC-HILIC-MS in combination with exoglycosidase digestion, *Anal. Chim. Acta* **940** (2016) 92–103, <https://doi.org/10.1016/j.aca.2016.07.043>.
- [30] E. Giménez, M. Balmaña, J. Figueras, E. Fort, C. De Bolós, V. Sanz-Nebot, R. Peracaula, A. Rizzi, Quantitative analysis of N-glycans from human  $\alpha$ 1-acid-glycoprotein using stable isotope labeling and zwitterionic hydrophilic interaction

- capillary liquid chromatography electrospray mass spectrometry as tool for pancreatic disease diagnosis, *Anal. Chim. Acta* 866 (2015) 59–68, <https://doi.org/10.1016/j.aca.2015.02.008>.
- [31] M. Belmaña, E. Giménez, A. Puerta, E. Llop, J. Figueras, E. Fort, V. Sanz-Nebot, C. de Bolós, A. Rizzi, S. Barrabés, M. de Frutos, R. Peracaula, Increased  $\alpha$ 1-3 fucosylation of  $\alpha$ -1-acid glycoprotein (AGP) in pancreatic cancer, *J. Proteome* 132 (2016) 144–154, <https://doi.org/10.1016/j.jpro.2015.11.006>.
- [32] 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, <https://doi.org/10.1007/s00216-013-7178-5>.
- [33] A. Puerta, J.C. Díez-Masa, P.J. Martín-Álvarez, J.L. Martín-Ventura, C. Barbas, J. Tuñón, J. Egido, M. de Frutos, Study of the capillary electrophoresis profile of intact  $\alpha$ -1-acid glycoprotein isoforms as a biomarker of atherothrombosis, *Analyst* 136 (2011) 816–822, <https://doi.org/10.1039/C0AN00320D>.
- [34] S. Onqay, C. Neusüß, S. Vaas, J.C. Díez-Masa, M. De Frutos, Evaluation of the effect of the immunopurification-based procedures on the CZE-UV and CZE-ESI-TOF-MS determination of isoforms of intact  $\alpha$ -1-acid glycoprotein from human serum, *Electrophoresis* 31 (2010) 1796–1804, <https://doi.org/10.1002/elps.200900680>.
- [35] P.B. Karda, F. Benavente, S. Catala-Clariana, E. Giménez, J. Barbosa, V. Sanz-Nebot, Identification of bioactive peptides in a functional yogurt by micro liquid chromatography time-of-flight mass spectrometry assisted by retention time prediction, *J. Chromatogr. A* 1229 (2012) 121–128, <https://doi.org/10.1016/j.chroma.2011.12.093>.
- [36] I.T. Jolliffe, B.J. Morgis, Principal component analysis and exploratory factor analysis, *Stat. Methods Med. Res.* 1 (1992) 69–95, <https://doi.org/10.1177/096228029200100105>.
- [37] M. Barker, W. Rayens, Partial least squares for discrimination, *J. Chemom.* 17 (2003) 166–173, <https://doi.org/10.1002/cem.785>.
- [38] S. Wold, M. Sjöström, L. Eriksson, PLS-regression: a basic tool of chemometrics, *Chemom. Intell. Lab. Syst.* 58 (2001) 109–130, [https://doi.org/10.1016/S0169-7439\(01\)00155-1](https://doi.org/10.1016/S0169-7439(01)00155-1).
- [39] M. Stone, Cross-validated choice and assessment of statistical predictions, *J. R. Stat. Soc.* 36 (1974) 111–147, <https://doi.org/10.2307/2964809>.

## **Supplementary Figures and Tables**

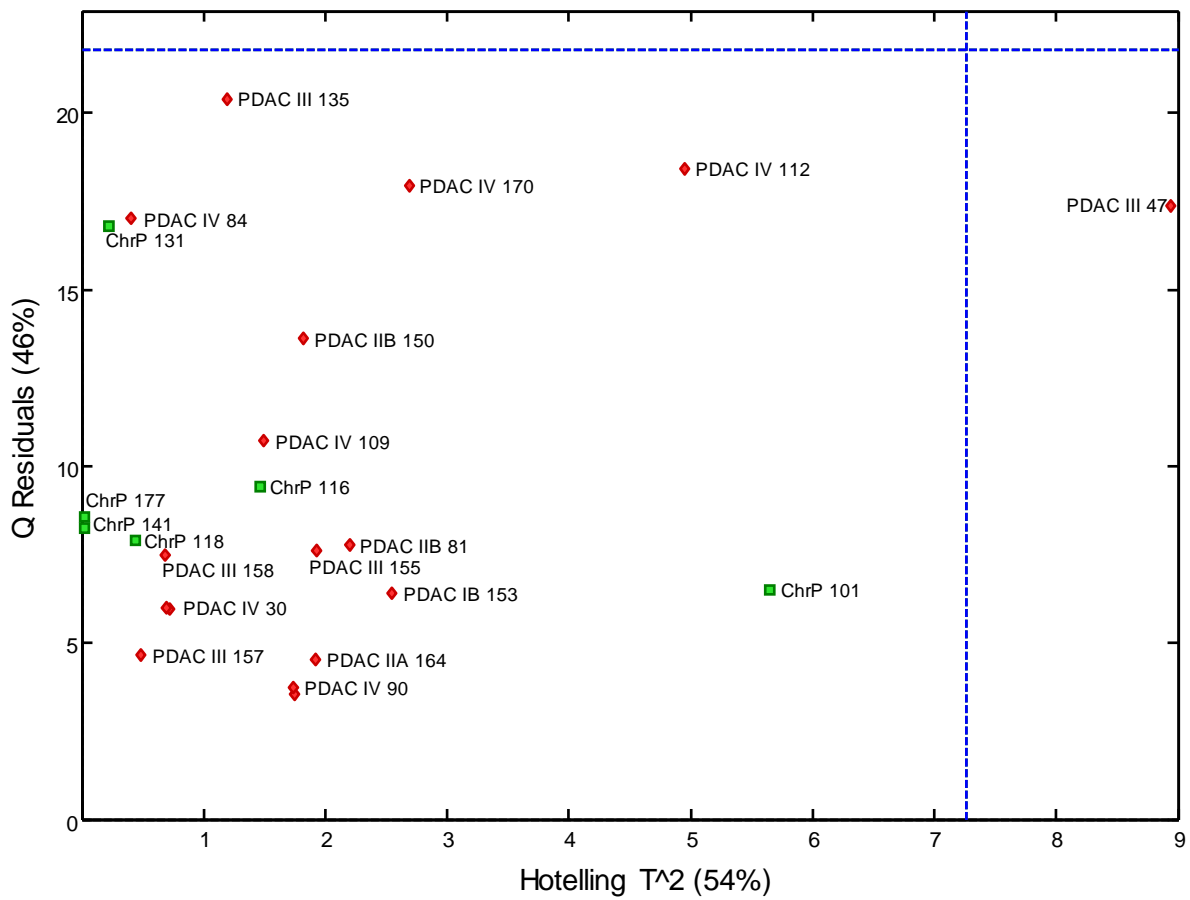
---





**Supplementary Figure S-1.** Outlier detection plot (Q Residuals vs. Hotelling T<sup>2</sup>) of the PCA model applied to the relative areas (pathological sample vs. healthy control pool) obtained for each hAGP intact glycan isomer in PDAC and ChrP samples.





**Supplementary Figure S-2.** Outlier detection plot (Q Residuals vs. Hotelling T<sup>2</sup>) of the PCA model applied to the relative areas (pathological vs. standard samples) obtained for each hAGP desialylated glycan isomer in PDAC and ChrP samples.

**Supplementary Table S-1.** Characterization of the sialic acid (SiA) and fucose (Fuc) linkage-types of most hAGP glycan isomers performed in our previous works using exoglycosidase digestions and tandem mass spectrometry [27,28].

| Glycan isomer                     | Previous characterization                                 |
|-----------------------------------|---|
| <b>Intact hAGP glycan isomers</b> |   |
| H5N4S1                            | 1 $\alpha$ 2-6 SiA  |
| H5N4S2_1                          | 1 $\alpha$ 2-3 SiA<br>1 $\alpha$ 2-6 SiA                  |
| H5N4S2_2                          | 2 $\alpha$ 2-6 SiA  |
| H6N5S2_1                          | 1 $\alpha$ 2-3 SiA<br>1 $\alpha$ 2-6 SiA                  |
| H6N5S2_2                          | 2 $\alpha$ 2-6 SiA  |
| H6N5S3_1                          | 1 $\alpha$ 2-3 SiA<br>2 $\alpha$ 2-6 SiA                  |
| H6N5S3_2                          | 1 $\alpha$ 2-3 SiA<br>2 $\alpha$ 2-6 SiA                  |
| H6N5S3_3                          | 3 $\alpha$ 2-6 SiA  |
| H7N6S1                            | 1 $\alpha$ 2-6 SiA  |
| H7N6S2_1                          | 1 $\alpha$ 2-3 SiA<br>1 $\alpha$ 2-6 SiA                  |
| H7N6S2_2                          | 2 $\alpha$ 2-6 SiA  |
| H7N6S3_1                          | 3 $\alpha$ 2-3 SiA  |
| H7N6S3_2                          | 2 $\alpha$ 2-3 SiA<br>1 $\alpha$ 2-6 SiA                  |
| H7N6S3_3                          | 1 $\alpha$ 2-3 SiA<br>2 $\alpha$ 2-6 SiA                  |
| H7N6S3_4                          | 3 $\alpha$ 2-6 SiA  |
| H7N6S4_1                          | 2 $\alpha$ 2-3 SiA<br>1 $\alpha$ 2-6 SiA<br>1 unknown SiA |
| H7N6S4_2                          | 2 $\alpha$ 2-3 SiA<br>1 $\alpha$ 2-6 SiA<br>1 unknown SiA |
| H7N6S4_3                          | 2 $\alpha$ 2-3 SiA<br>2 $\alpha$ 2-6 SiA                  |
| H5N4F1S2_1                        | 1 $\alpha$ 2-3 SiA<br>1 $\alpha$ 2-6 SiA                  |
| H5N4F1S2_2                        | 2 $\alpha$ 2-6 SiA  |
| H6N5F1S2_1                        | 1 $\alpha$ 2-3 SiA<br>1 $\alpha$ 2-6 SiA                  |
| H6N5F1S2_2                        | 2 $\alpha$ 2-6 SiA  |
| H6N5F1S3_1                        | 1 $\alpha$ 2-3 SiA<br>2 $\alpha$ 2-6 SiA                  |
| H6N5F1S3_1                        | 1 $\alpha$ 2-3 SiA<br>2 $\alpha$ 2-6 SiA                  |
| H6N5F2S3                          | 1 $\alpha$ 2-3 SiA<br>2 $\alpha$ 2-6 SiA                  |
| H7N6F1S2_1                        | 2 $\alpha$ 2-6 SiA  |
| H7N6F1S2_2                        | 2 $\alpha$ 2-6 SiA  |
| H7N6F1S3_1                        | 1 $\alpha$ 2-3 SiA<br>1 $\alpha$ 2-6 SiA<br>1 unknown SiA |
| H7N6F1S3_2                        | 1 $\alpha$ 2-3 SiA  |

|   |   |
|---|---|
|   | 1 $\alpha$ 2-6 SiA<br>1 unknown SiA                       |
| H7N6F1S3_3                              | 1 $\alpha$ 2-3 SiA<br>1 $\alpha$ 2-6 SiA<br>1 unknown SiA |
| H7N6F1S3_4                              | 3 $\alpha$ 2-6 SiA  |
| H7N6F1S4_1                              | 2 $\alpha$ 2-3 SiA<br>1 $\alpha$ 2-6 SiA<br>1 unknown SiA |
| H7N6F1S4_2                              | 2 $\alpha$ 2-3 SiA<br>1 $\alpha$ 2-6 SiA<br>1 unknown SiA |
| H7N6F2S3_1                              | 1 $\alpha$ 2-3 SiA<br>1 $\alpha$ 2-6 SiA<br>1 unknown SiA |
| H7N6F2S3_2                              | 3 $\alpha$ 2-6 SiA  |
| H7N6F2S4                                | 2 $\alpha$ 2-3 SiA<br>1 $\alpha$ 2-6 SiA<br>1 unknown SiA |
| <b>Desialylated hAGP glycan isomers</b> |   |
| H5N4F1_1                                | 1 Core Fuc  |
| H5N4F1_2                                | 1 Core Fuc  |
| H5N4F1_3                                | 1 $\alpha$ 1-3 Antenna Fuc                                |
| H6N5F1_1                                | 1 Core Fuc  |
| H6N5F1_2                                | 1 Core Fuc  |
| H6N5F1_3                                | 1 $\alpha$ 1-3 Antenna Fuc                                |
| H6N5F2_1                                | *   |
| H6N5F2_2                                | *   |
| H6N5F2_3                                | At least 1 $\alpha$ 1-3 Antenna Fuc                       |
| H7N6F1_1                                | 1 Core Fuc  |
| H7N6F1_2                                | 1 $\alpha$ 1-3 Antenna Fuc                                |
| H7N6F2_1                                | 2 Core Fuc  |
| H7N6F2_2                                | At least 1 $\alpha$ 1-3 Antenna Fuc                       |
| H7N6F2_3                                | At least 1 $\alpha$ 1-3 Antenna Fuc                       |
| H7N6F3_1                                | At least 1 $\alpha$ 1-3 Antenna Fuc                       |
| H7N6F3_1                                | At least 1 $\alpha$ 1-3 Antenna Fuc                       |

\*The characterization of these isomers was not possible.

# Alterations in the Glycan Profile of Mouse Transferrin: New Insights in Collagen-Induced Arthritis

Montserrat Mancera-Arteu, Estela Giménez, Jaime Sancho, and Victoria Sanz-Nebot\*

Cite This: *J. Proteome Res.* 2020, 19, 1750–1759

Read Online

ACCESS |

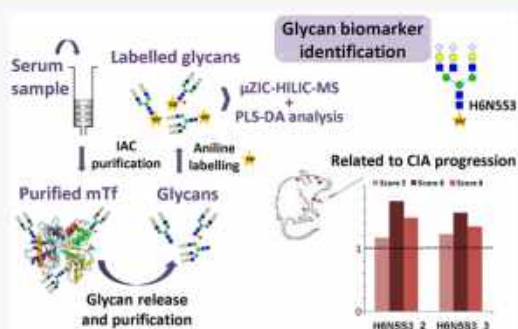
Metrics &amp; More

Article Recommendations

Supporting Information

**ABSTRACT:** Transferrin purification from mice serum samples by immunoaffinity chromatography (IAC) was optimized in order to study the possible modifications occurring in its glycans in collagen-induced arthritis (CIA) samples. SDS-PAGE and nanoLC-MS/MS were used to monitor the IAC purification performance. Afterward, a relative quantification of mouse transferrin (mTf) glycan isomers using [ $^{12}\text{C}_6$ ]/[ $^{13}\text{C}_6$ ]-aniline was used to unequivocally detect alterations in the glycan profile of CIA mice. In addition, multivariate data analysis was applied to identify the most meaningful glycan isomers for the discrimination between control and pathological samples. Partial least-squares discriminant analysis (PLS-DA) revealed that five out of fifteen mTf glycan isomers could be potential biomarkers of CIA, most of them corresponding to highly sialylated structures (H6N5S3\_2, H6N5S3\_3, and H5N4S3\_2). Moreover, some of these glycan isomers also seemed to be related with the progression of CIA, especially H6N5S2 and H6N5S3\_2, as their overexpression increased with the clinical score of the pathology. Hence, the established methodology not only provides valuable information to find glycan-based biomarkers of CIA, but also leaves the door open to evaluate, in the future, glycosylation changes of many other inflammatory diseases, in which transferrin has been described to be altered.

**KEYWORDS:** transferrin, collagen-induced arthritis, glycan isomers, immunoaffinity chromatography, multivariate data analysis, sialylation



## 1. INTRODUCTION

Changes in serum protein glycosylation are early indicators of cellular alterations in many diseases, including the inflammatory arthropathies (IA), providing a good basis for diagnosis and insights into disease progression.<sup>1,2</sup> Altered glycomic profile was reported in several proteins isolated from the sera of IA patients, including IgG, transferrin, haptoglobin, and  $\alpha$ 1-acid glycoprotein.<sup>1,3</sup> Transferrin (Tf) is a negative acute phase protein whose main role is the transport of iron through the blood plasma. Changes in sialylation, fucosylation, and glycan branching of Tf have been described in many inflammatory diseases,<sup>5</sup> such as rheumatoid arthritis (RA). RA is a chronic inflammatory autoimmune disease that affects up to 1% of population worldwide, doubling mortality rate of RA patients compared to healthy individuals. However, up to now, there is no single test to confirm the diagnosis of RA and it is based on the symptoms and the measurement of some variables (rheumatoid factor (RF), C-reactive protein (CRP), erythrocyte sedimentation rate (ESR), among others).<sup>4</sup> Therefore, with the aim of finding a reliable biomarker, several authors addressed the study of Tf glycosylation in RA analyzing the intact glycoforms by isoelectric focusing (IEF),<sup>5</sup> or more recently, by capillary electrophoresis (CE).<sup>6</sup> However, these methods provided ambiguous results: whereas an increase of

the highly sialylated glycoforms (S5 and S6) in RA patients was observed by Feelders et al.,<sup>5</sup> Gudowska et al.<sup>6</sup> reported a decrease of S3 and S5 sialoforms.

CIA is a commonly studied autoimmune mice model of arthritis that closely resembles human RA in terms of disease course, histological findings, and also in its response to commonly used antiarthritic pharmaceuticals.<sup>7,8</sup> A chronic form of CIA is induced in C57BL/6 wild-type (WT) mice by immunization with chicken type II collagen in complete Freund's adjuvant (Col-II/CFA).<sup>9</sup> Thus, it is considered an adequate model to study the efficacy of novel drugs and to evaluate glycosylation changes derived from arthritis.<sup>10</sup> Mouse transferrin (mTf) is a serum glycoprotein of about 80 kDa, which presents only one N-glycosylation site with complex type N-glycans. Due to the multiple similarities between CIA and RA previously mentioned, alterations in the glycosylation pattern of mTf in the presence of CIA could be expected. In a

Received: January 13, 2020

Published: March 12, 2020

previous work by our research group, the analysis of mouse transferrin (mTf) at the glycopeptide level by capillary liquid chromatography–mass spectrometry was carried out. That study demonstrated that mTf glycopeptides are modified in the presence of CIA, mainly observing an increase of fucosylation and glycan branching.<sup>11</sup>

However, the glycopeptide approach did not provide information about possible altered glycan isomers. Zwitterionic hydrophilic interaction capillary liquid chromatography–mass spectrometry (CapZIC-HILIC-MS) enables an excellent separation of isomeric glycans, and the use of a glycan reductive isotope labeling (GRIL) strategy with [<sup>12</sup>C<sub>6</sub>]/[<sup>13</sup>C<sub>6</sub>]-aniline allows performing a reliable quantification isomer by isomer, identifying unequivocally major and minor variations in the expression of certain glycan isomers.<sup>12,13</sup> Nevertheless, using this glycomic approach, a specific and selective purification strategy is required in order to selectively capture the target glycoprotein and avoid the contribution of glycans coming from other glycoproteins.

In this work, an immunoaffinity chromatography (IAC) column was developed and optimized to properly isolate mTf from serum samples, testing different conditions to remove the nonspecifically retained proteins. SDS-PAGE as well as nanoLC-MS/MS analysis of the eluted IAC fractions were used to monitor the performance of mTf purification. Under the optimized IAC conditions, glycan isomers of mTf purified from commercial control mouse serum were analyzed by CapZIC-HILIC-MS to establish a reference glycoprofile of this protein. Afterward, in order to detect modifications in mTf glycosylation in WT mice suffering from CIA (CIA WT samples), relative quantification of mTf glycans was carried out using the GRIL strategy with [<sup>12</sup>C<sub>6</sub>]/[<sup>13</sup>C<sub>6</sub>]-aniline (CIA or nonimmunized WT mice versus control mouse serum samples). To our knowledge, this is the first glycomic study focused on the identification of Tf aberrant glycosylation in such inflammatory diseases. Moreover, given the potential of partial least-squares discriminant analysis (PLS-DA) for the analysis of human  $\alpha$ 1-acid glycoprotein glycan isomers in pancreatic ductal adenocarcinoma (PDAC) samples,<sup>13</sup> this chemometric tool was also used in the present study to identify which glycan isomers of mTf enable the differentiation between nonimmunized and CIA WT mice. These altered glycans in CIA could be useful, in the future, to find novel glycan biomarkers for the diagnosis of RA in humans.

## 2. MATERIALS AND METHODS

### 2.1. Chemicals

All chemicals used in the preparation of buffers and solutions were of analytical reagent grade. Acetic acid (HAc, glacial), formic acid (HFor 98–100%), dimethylsulfoxide (DMSO), glycine ( $\geq 99.7\%$ ), glycerol ( $\geq 99.5\%$ ), sodium dodecyl sulfate (SDS,  $\geq 99.8\%$ ), tris(hydroxymethyl)aminomethane (Tris,  $\geq 99.9\%$ ), and acetone were supplied by Merck (Darmstadt, Germany). Sodium phosphate dodecahydrated (Na<sub>2</sub>PO<sub>4</sub>·12H<sub>2</sub>O), sodium cyanoborohydride (NaBH<sub>3</sub>CN), [<sup>12</sup>C<sub>6</sub>]-aniline, [<sup>13</sup>C<sub>6</sub>]-aniline, sodium chloride (NaCl, 99.5%), and sodium bicarbonate (NaHCO<sub>3</sub>) were purchased from Sigma-Aldrich (St. Louis, MO, USA). Hydrogen chloride (HCl, 37%) and 2-mercaptoethanol ( $\beta$ -ME) were supplied by Panreac (Barcelona, Spain) and “NP-40 alternative” by Calbiochem (Darmstadt, Germany). Sodium azide (NaN<sub>3</sub>,  $\geq 99.5\%$ ) was obtained from Fluka (Madrid, Spain). Bromophenol blue,

tetramethylethylenediamine (TEMED), acrylamide/bis solution (30%), ammonium persulfate (APS), and Bio-Safe Coomassie stain were supplied by Bio-Rad (Hercules, USA). BenchMark Protein Ladder was provided by Thermo Fisher Scientific (Waltham, USA). Ammonium acetate (NH<sub>4</sub>Ac), acetonitrile (ACN), and water LC-MS quality grade, used for CapZIC-HILIC-MS analysis, were obtained from Merck and Fluka, respectively. Mouse apo-transferrin standard (mTf,  $\geq 98\%$ , commercially purified from control sera) was purchased from Sigma-Aldrich. Peptide N-glycosidase F (PNGase F) was obtained from Roche Diagnostics (Basel, Switzerland). CNBr-activated Sepharose 4B was purchased from GE Healthcare (Waukesha, WI, USA) and goat polyclonal antibody against human Tf (immunogen affinity purified) from Abcam (Cambridge, UK).

### 2.2. Mice Serum Samples

Control mouse serum (pooled serum from normal mouse population) was purchased from Sigma-Aldrich. C57BL/6 wild-type (WT) mice were purchased from Harlan Ibérica (Barcelona, Spain). For the induction of CIA, 8–12 weeks-old male mice were immunized with Col-II/CFA 6 as previously described.<sup>9,14</sup> Four nonimmunized WT mice serum samples: WT1, WT2, WT6, WT7, all of them males of 8–10 weeks old, and five WT mice serum samples with CIA: WT1-CIA (clinical score of 5), WT3-CIA (clinical score of 7), WT4-CIA (clinical score of 6), WT6-CIA (clinical score of 8), and WT7-CIA (clinical score of 8), all of them males of 14–16 weeks old, were analyzed. The clinical score was quantified according to a graded scale of 0–3 as follows: 0 = no inflammation (normal joint); 1 = detectable local swelling and/or erythema; 2 = swelling in >1 joint and pronounced inflammation; 3 = swelling of the entire paw and/or ankylosis. Each paw was graded, and the scores were summed (with a maximum possible score 12 per mouse). All studies with live animals were approved by the IPBLN and Universidad de Cantabria Institutional Laboratory Animal Care and Use Committees.

### 2.3. Immunoaffinity Chromatography (IAC)

**2.3.1. IAC Column Preparation.** A CNBr-sepharose IAC column was prepared following our previous protocol<sup>15</sup> with some modifications proposed in the literature.<sup>16</sup> Briefly, 5-column volumes of cold activation buffer (1 mM HCl) were added to 0.15 g of cyanogen bromide-sepharose dry resin and incubated on a nutator mixer for 2 h at 4 °C. After swelling the resin (0.5 mL, approximately), it was centrifuged for 5 min and supernatant was decanted. The commercial buffer containing the anti-Tf antibody was exchanged for the coupling buffer (0.1 M NaHCO<sub>3</sub> and 0.5 M NaCl, pH 8.3) using Microcon YM-10 centrifugal filters as described in ref 17. The resultant antibody solution (650  $\mu$ L at a concentration of 1.54 mg·mL<sup>-1</sup>) was added to the resin and incubated overnight on a nutator mixer at 4 °C. Afterward, resin was washed several times with coupling buffer, centrifuged, and supernatant was removed. In order to block unreacted groups, 5-column volumes of quenching buffer (0.1 M Tris–HCl, pH 8.0) were added and incubated on a nutator mixer for 2 h at room temperature. Then, resin was extensively washed to remove uncoupled antibody repeating this cycle for three times: centrifugation and supernatant removal, addition of 10-column volumes of high pH washing buffer (0.1 M Tris–HCl and 0.5 M NaCl, pH 8.0), centrifugation and supernatant removal, and addition of 10-column volumes of low pH washing buffer (0.1 M NaAc and 0.5 M NaCl, pH 4.0). Finally, resin was

transferred to a 1.5 mL empty plastic column and stored at 4 °C filled with storage buffer (10 mM Tris-HCl and 0.01% (w/v) NaN<sub>3</sub>, pH 7.6–7.7).

**2.3.2. Albumin/IgG Depletion Kit.** ProteoExtract Albumin/IgG Removal Kit was purchased from Merck to remove serum albumin and IgG from mice serum samples. Depletion procedure was performed according to the manufacturer's instructions.<sup>18</sup> Samples were concentrated by Speed Vac until reaching a proper volume (~600–800  $\mu$ L) to be loaded later in the IAC column.

**2.3.3. mTf Purification by IAC.** The purification procedure previously described in ref 15 was modified in order to improve mTf isolation, minimizing nonspecific retention of other serum proteins. First, the IAC column was conditioned with 8-column volumes of binding buffer (10 mM Tris-HCl and 0.25 M NaCl, pH 7.6). Serum samples (50–100  $\mu$ L) were diluted 1/8 in binding buffer and passed ten times through the IAC column (albumin/IgG-depleted samples were loaded directly). After washing with 6-column volumes of washing buffer (10 mM Tris-HCl and 0.5 M NaCl, pH 7.6), retained mTf was eluted with 1-column volumes of elution buffer (100 mM glycine-HCl, pH 2.5). The eluate was collected in a tube containing 0.5 M Tris to immediately neutralize it. Finally, excess of low-molecular mass reagents was removed with Microcon YM-30 centrifugal filters as described in ref 17.

After IAC, purified mTf samples were quantified by capillary electrophoresis with ultraviolet detection (CE-UV, Agilent Technologies, Waldbronn, Germany) using a bare fused-silica capillary of 60 cm total length (LT)  $\times$  75  $\mu$ m internal diameter (I.D.)  $\times$  360  $\mu$ m outer diameter (O.D.) (Polymicro Technologies, Phoenix, AZ, USA). This protein quantification approach was used due to its simplicity, low sample consumption, quickness, and possibility of automation for consecutive quantification of many serum samples. New capillaries were activated off-line by flushing (930 mbar) sequentially for 15 min each with 1 M NaOH and water. Samples were injected for 10 s at 50 mbar and CE was only used to introduce the protein sample throughout the capillary. Experiments were performed without supplying voltage, only applying 50 mbar of pressure, and measuring absorbance at 214 nm. Calibration curve was performed with mouse apo-transferrin standard at concentrations between 50 and 1000  $\mu$ g·mL<sup>-1</sup>. Afterward, purified mTf samples were evaporated to dryness by Speed Vac and stored at -20 °C until its use.

#### 2.4. SDS-PAGE

SDS-PAGE was performed on a vertical system Mini-PROTEAN Tetra Cell with a PowerPac HC Power Supply (Bio-Rad, Hercules, USA) using in-house 10% SDS-polyacrylamide gels (30% (v/v) Acrylamide/bis solution, 0.375 M Tris-HCl pH 8.8, 10% (v/v) SDS, 10% (v/v) APS and 0.004% (v/v) TEMED). Fifteen  $\mu$ L of each IAC purified serum sample or mouse apo-transferrin standard (~10  $\mu$ g of protein) were reduced and denatured with 5  $\mu$ L of reducing-Laemmli sample buffer (0.25 M Tris-HCl pH 6.8, 4% (m/v) SDS, 20% (v/v) glycerol, 10% (v/v)  $\beta$ -ME, and 1% (v/v) bromophenol blue), incubating the mixture in a thermoshaker at 100 °C for 5 min. Then, samples were loaded into the gel to perform the protein separation. Ten  $\mu$ L of protein ladder (BenchMark Protein Ladder) were also loaded in one lane in order to assign the molecular weight to the bands. Gel electrophoresis was performed at 120 V for 2 h at room temperature using a

running buffer consisted of 25 mM Tris-base, 250 mM glycine and 0.1% SDS. After SDS-PAGE, gel was fixed in 40% (v/v) ethanol and 10% (v/v) HAc for 30 min and then rinsed in Milli-Q water (3  $\times$  5 min). Gel was incubated with Coomassie blue staining solution at room temperature for 1 h with agitation, and then rinsed in Milli-Q water until a proper degree of staining was achieved.

#### 2.5. NanoLC-MS/MS

IAC purified samples (~5  $\mu$ g of mTf) as well as mouse apo-transferrin standard (10  $\mu$ g of mTf) were reduced, alkylated and subjected to trypsin digestion in the presence of Rapigest as described in ref 15.

Prepared tryptic digests were diluted in 3% ACN, 1% HFor at a final concentration of 0.35 pmol/ $\mu$ L. One  $\mu$ L of each sample was loaded into a 300  $\mu$ m  $\times$  5 mm PepMap100, 5  $\mu$ m, 100 Å, C18  $\mu$ -precolum (Thermo Scientific) at a flow rate of 15  $\mu$ L/min using a Thermo Scientific Dionex Ultimate 3000 chromatographic system (Thermo Scientific). Peptides were separated using a C18 analytical column (Acclaim PepMap RSLC 75  $\mu$ m  $\times$  50 cm, nanoViper, C18, 2  $\mu$ m, 100Å, Thermo Scientific) with gradient elution at a flow rate of 250 nL·min<sup>-1</sup>. Eluting solvents were A: 0.1% HFor in water and B: 0.1% HFor in ACN. The following gradient conditions were used: from 3 to 35% B in 60 min, from 35 to 50% B in 5 min, and from 50% to 85% B in 2 min, followed by isocratic elution at 85% B in 5 min and stabilization to initial conditions. LC-MS coupling was performed with an Advion TriVersa NanoMate (Advion) fitted on an Orbitrap Fusion Lumos Tribrid (Thermo Scientific). The mass spectrometer operated in a data-dependent acquisition (DDA) mode. Survey MS scans were acquired in the Orbitrap with the resolution (defined at 200 *m/z*) set to 120 000. The lock mass was user-defined at 445.12 *m/z* in each Orbitrap scan. The top speed (most intense) ions per scan were fragmented by HCD and detected in the Orbitrap. The ion count target value was 400 000 and 50 000 for the survey scan and for the MS/MS scan, respectively. Target ions already selected for MS/MS were dynamically excluded for 30 s. Spray voltage in the NanoMate source was set to 1.60 kV. RF Lens were tuned to 30%. Minimal signal required to trigger MS to MS/MS switch was set to 20 000. The spectrometer was working in positive polarity mode and singly charge state precursors were rejected for fragmentation.

A database search was performed with Proteome Discoverer software v2.1.0.81 (Thermo) using Sequest HT search engine and SwissProt mouse with contaminants database. Search was run against targeted and decoy database to determine the false discovery rate (FDR). Search parameters included trypsin enzyme specificity, allowing for two missed cleavage sites, oxidation in methionine and acetylation in protein N-terminus as dynamic modifications. Peptide mass tolerance was 10 ppm and MS/MS tolerance was 0.02 Da. Peptides with a *q*-value lower than 0.1 and a FDR < 1% were considered as positive identifications with a high confidence level. The Exponentially Modified Protein Abundance Index (emPAI), based on protein coverage by the peptide matches in a database search result, was used to relatively quantify the abundance of the proteins present in the mixture.

#### 2.6. N-Glycan Preparation

mTf purified samples were reduced and digested with PNGase F to release N-glycans as explained elsewhere.<sup>19</sup> Released N-glycans were purified by solid phase extraction (SPE) using

Hypercarb cartridges (25 mg, 1 mL volume, Thermo Fisher Scientific) as described previously.<sup>13</sup> The labeling was carried out by adding 10  $\mu$ L of reaction mixture (0.35 M aniline and 1 M NaCNBH<sub>3</sub> in DMSO with 30% HAc) to the dried glycans and incubating the mixture in a thermoshaker for 2 h at 70 °C. mTf glycans obtained from WT mice serum samples (nonimmunized and CIA samples) were labeled with [<sup>12</sup>C<sub>6</sub>]-aniline, while mTf glycans obtained from commercial control mouse serum were labeled with [<sup>13</sup>C<sub>6</sub>]-aniline. After incubation, samples were cooled to room temperature and labeled glycans were precipitated with acetone as described in ref 20. Subsequently, equimolar mixtures of mTf nonimmunized or CIA samples and mTf control mouse serum samples were prepared. Finally, the obtained mixtures were evaporated to dryness by Speed Vac, and dried N-glycans were stored at -20 °C until analysis. Centrifugations were performed in a Mikro 220R centrifuge (Hettich Zentrifugen, Tuttlingen, Germany).

### 2.7. CapZIC-HILIC-MS

1200 Series capillary liquid chromatography system coupled to a 6220 oa-TOF LC/MS mass spectrometer with an orthogonal G1385-44300 interface (Agilent Technologies) were used to perform CapLC-MS experiments. ZIC-HILIC column packed with 3.5 mm particles, 150  $\times$  0.3 mm L<sub>T</sub>  $\times$  ID (SeQuant, Umeå, Sweden) was used for chromatographic separations. Experiments were performed at room temperature with gradient elution at a flow rate of 4  $\mu$ L/min and injecting 0.25  $\mu$ L of each glycan sample (glycan concentration: ~100–50 pmol/ $\mu$ L). Eluting solvents were A: 1 mM NH<sub>4</sub>Ac solution and B: acetonitrile. The following gradient conditions were used: solvent B from 90% to 80% (within 5 min) and from 80% to 65% (within 20 min) as linear gradient, followed by cleaning and equilibration steps of B: 65%  $\rightarrow$  50% (within 5 min), 50%  $\rightarrow$  0% (within 5 min), 0% (over 15 min), 0%  $\rightarrow$  90% (within 5 min) and 90% (over 10 min). Tuning and calibration of the mass spectrometer were carried out in accordance with the manufacturer's instructions. Measurement parameters and optimal operational conditions are detailed in ref 19. MassHunter Workstation software (Agilent Technologies) was used for CapLC-MS control, data acquisition and analysis.

### 2.8. Multivariate Data Analysis

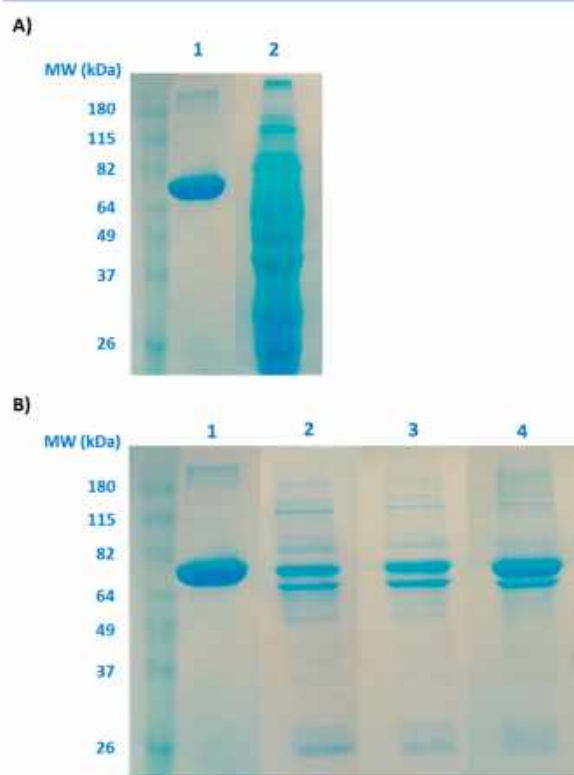
CapLC-MS data collected for mTf purified from serum samples was processed to obtain the extracted ion chromatograms (EIC) of all glycans. The EIC of each glycan was obtained based on the *m/z* of the most abundant molecular ions observed for each glycan (usually adducts corresponding to the deprotonated glycan: [M-2H]<sup>-2</sup>, [M-3H]<sup>-3</sup>, and [M-4H]<sup>-4</sup>). Peak areas of the N-glycan isomers were measured from the EICs and used to calculate its relative area applying a GRIL strategy with [<sup>12</sup>C<sub>6</sub>]/[<sup>13</sup>C<sub>6</sub>] aniline (i.e., area of the glycan isomer in the WT mice sample, nonimmunized or CIA, labeled with [<sup>12</sup>C<sub>6</sub>]-aniline, divided by the area of the same glycan isomer in the control mouse serum sample, labeled with [<sup>13</sup>C<sub>6</sub>]-aniline). Relative areas of glycan isomers were used to build a matrix for multivariate data analysis. The relative areas of the matrix were autoscaled (mean centered and scaled to unit standard deviation). Principal component analysis (PCA) was performed to explore the data for different classes and detect the presence of outliers.<sup>21</sup> Partial least-squares discriminant analysis (PLS-DA) was applied afterward to maximize class separation and identify which glycan isomers

were the most meaningful to discriminate between classes taking into account the variable importance in the projection (VIP) scores.<sup>22,23</sup> Leave-one-out cross validation of the PLS-DA model was performed during calibration.<sup>24</sup> In the prediction step, the model was used to classify the validation sample. SOLO (Version 8.7, student edition, Eigenvector Research Inc., Wenatchee, WA, USA) was used for PCA, PLS-DA, and VIP calculations. Nomenclature used for glycans correspond to their composition, in terms of number of hexoses (H), N-acetylglucosamines (N), fucoses (F), and sialic acids (S), followed by an index that indicates the isomer number.

## 3. RESULTS AND DISCUSSION

### 3.1. Purification of mTf by Immunoaffinity Chromatography

Immunoaffinity chromatography (IAC) was used to isolate Tf from mice serum samples using a cyanogen-bromide Sepharose column with an immobilized antibody against human Tf. In order to evaluate the performance of mTf purification, SDS-PAGE as well as nanoLC-MS/MS analysis of the eluted IAC fractions were carried out. First, the IAC column and the purification procedure reported previously for the analysis of human and mouse Tf glycopeptides were used.<sup>11,15</sup> Figure 1A shows the SDS-PAGE results obtained for mouse apo-



**Figure 1.** (A) SDS-PAGE results obtained for mouse apo-transferrin standard (lane 1), and commercial control mouse serum purified by the initial IAC conditions (lane 2). (B) SDS-PAGE results obtained for mouse apo-transferrin standard (lane 1) and control mouse serum purified by the new IAC column: using larger washing and elution volumes (lane 2); adding 0.25 M NaCl to the binding buffer (lane 3) and using a depletion kit before IAC (lane 4).

transferrin standard (lane 1), and for the control mouse serum purified using our initial IAC conditions (lane 2). As can be observed, other serum proteins were nonspecifically retained by the IAC column. Table 1 shows the list of proteins

**Table 1. Proteins Identified by NanoLC-MS/MS with emPAI Values Higher than 10%, in Control Mouse Serum Sample Purified by Immunoaffinity Chromatography (IAC), Using the Initial and the Optimized IAC Conditions**

| protein name <sup>a</sup>       | accession number <sup>b</sup> | MW (kDa) | N-glycosylation sites | emPAI <sup>c</sup> (%) |
|---------------------------------|-------------------------------|----------|-----------------------|------------------------|
| <b>Initial IAC conditions</b>   |                               |          |                       |                        |
| Serum albumin                   | P07724                        | 68.6     | 0                     | 93.8                   |
| Serotransferrin                 | Q92111                        | 76.7     | 1                     | 34.3                   |
| Apolipoprotein A-1              | Q00623                        | 30.6     | 0                     | 24.1                   |
| Vitamin D-binding protein       | P21614                        | 53.6     | 1                     | 23.2                   |
| Pregnancy zone protein          | Q61838                        | 165.7    | 11                    | 21.0                   |
| Alpha-1-antitrypsin 1-5         | Q00898                        | 45.9     | 3                     | 18.0                   |
| Serine protease inhibitor A3K   | P07759                        | 46.9     | 4                     | 16.3                   |
| Alpha-1-antitrypsin 1-3         | Q00896                        | 45.8     | 3                     | 14.5                   |
| Immunoglobulin kappa constant   | P01837                        | 11.8     | 0                     | 13.7                   |
| Alpha-1-antitrypsin 1-4         | Q00897                        | 46.0     | 3                     | 12.9                   |
| Alpha-1-antitrypsin 1-2         | P22599                        | 45.9     | 3                     | 11.6                   |
| Serine protease inhibitor A3M   | Q03734                        | 47.0     | 3                     | 10.1                   |
| <b>Optimized IAC conditions</b> |                               |          |                       |                        |
| Serotransferrin                 | Q92111                        | 76.7     | 1                     | 55.2                   |
| Serum albumin                   | P07724                        | 68.6     | 0                     | 21.3                   |

<sup>a</sup>Protein Name according to UniProt or to NCBI. <sup>b</sup>UniProtKB/Swiss-Prot or NCBI Accession Number. <sup>c</sup>Exponentially Modified Protein Abundance Index (emPAI): relative quantitation of the proteins in a mixture based on protein coverage by the peptide matches in a database search result.

identified by nanoLC-MS/MS with emPAI values higher than 10%, considered the most abundant ones, in the eluted fraction, which corresponds to lane 2 of the SDS-PAGE gel in Figure 1A. Most of these proteins present N-glycosylation sites, some of them with a high percentage of glycosylation, which would contribute to the total amount of glycans of the sample analyzed.

Since the purpose of this study was focused on the analysis of glycans obtained only from mTf, a more selective purification method was necessary. First, a new IAC column was prepared modifying some steps of the previous protocol reported in ref 15, as explained in section 2.3.1. Briefly, the antibody to resin ratio as well as the incubation time were substantially increased and the antibody coupling steps were performed at 4 °C as suggested by ref 16. However, mTf purification using the new IAC column also resulted in a high nonspecific retention, similar to the one showed in Figure 1A. Therefore, the purification procedure was changed increasing the washing and elution volumes trying to favor both the cleanup and the elution of the target protein. As can be observed in Figure 1B, these modifications were enough to significantly improve mTf isolation from control mouse serum by IAC (compare lane 2 of Figures 1A and 1B). Nevertheless,

some bands, corresponding to other proteins, were still present in the SDS-PAGE gel in contrast to mouse apo-transferrin standard (lane 1). To avoid this nonspecific retention, 0.25 M of NaCl was added to the binding buffer. Lane 3 of the SDS-PAGE gel of Figure 1B shows the results obtained using these conditions. Although the number of bands decreased, an intense band at around 68 kDa still remained, probably corresponding to albumin. The use of a depletion kit prior to IAC purification was then evaluated in order to eliminate albumin. Although mTf band seemed to be more intense when using a depletion kit (lane 4 of Figure 1B), the significant increase in costs and in time required to perform the purification was not worth it. In addition, the albumin band was still present. Finally, we also tested a double immunopurification, but this resulted in a very poor mTf detection for both depleted and nondepleted control mouse serum. Given the results obtained, the IAC procedure using a binding buffer containing 0.25 M NaCl, washing with 6-column volumes of washing buffer and eluting with 1-column volume of elution buffer, was established as the optimum (lane 3 of Figure 1B). Under these conditions, only transferrin and albumin were identified by nanoLC-MS/MS with emPAI values higher than 10% (see Table 1). As albumin is not a glycosylated protein, it should not contribute to the glycomic profile of the purified mTf. To corroborate this issue, the bar graph of Supplementary Figure S1 compares the glycomic profile obtained from the purified control mouse serum sample using the initial IAC conditions and the optimized protocol with the new IAC column by CapZIC-HILIC-MS. As can be observed, the profile using the initial procedure was quite different, probably because of the contribution of glycans that came from other glycosylated proteins.

### 3.2. Analysis of mTf Glycan Isomers in CIA Samples

Once optimized mTf purification from serum samples, with the aim of detecting possible modifications on its glycosylation in mice with collagen-induced arthritis (CIA), a reference glycan profile of mTf was first established by analyzing mTf glycans isolated from the control mouse serum. Table 2 lists the glycans and their corresponding isomers detected by CapZIC-HILIC-MS in the control mouse serum sample after IAC purification.

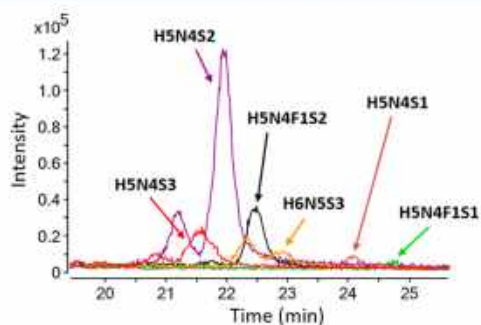
As can be observed, most glycans correspond to biantennary structures, being H5N4S2 the most abundant one. Some of them were also fucosylated, although they were less intense than the nonfucosylated counterpart. These results agreed with those reported in ref 11 for the analysis of mTf at the glycopeptide level. However, tetrantennary structures were not detected in this work, probably due to their low abundance. On the other hand, all sialic acids present in mTf glycans were N-glycolylneuramic acids (NeuGc), as reported in refs 11, 25. By way of an example, Figure 2 shows the extracted ion chromatograms (EIC) of the most abundant mTf glycans detected in the control mouse serum sample. Moreover, we compared the glycan profile of this commercially available control mouse serum with respect to a pool of nonimmunized WT mice serum samples. As can be seen in Supplementary Figure S2, both samples showed the same type of glycans and with similar abundances, which demonstrated this control mouse serum could be reliably used in this study as reference for the analysis of mTf glycans of all nonimmunized and CIA mice samples.



**Table 2.** Glycans Detected by CapZIC-HILIC-MS in the Control Mouse Serum Sample after IAC Purification

| glycan   | isomer | $t_R$<br>(min) | relative area<br>(%) <sup>a</sup> | $M_{theo}$ <sup>b</sup> | error<br>(ppm) |
|----------|--------|----------------|-----------------------------------|-------------------------|----------------|
| H5N4S1   | 1      | 24.0           | 2.41                              | 2024.7216               | 8.4            |
| H5N4F1S1 | 1      | 24.8           | 1.53                              | 2170.7796               | 9.5            |
| H5N4S2   | 1      | 21.2           | 9.99                              | 2331.8120               | 3.1            |
|          | 2      | 21.9           | 45.3                              | 2331.8120               | 2.6            |
| H5N4F1S2 | 1      | 21.8           | 1.84                              | 2477.8699               | 5.1            |
|          | 2      | 22.5           | 13.4                              | 2477.8699               | 4.3            |
| H5N4S3   | 1      | 20.8           | 1.61                              | 2638.9023               | 6.3            |
|          | 2      | 21.6           | 10.5                              | 2638.9023               | 5.2            |
| H5N4F1S3 | 1      | 21.9           | 1.21                              | 2784.9602               | 10.9           |
| H6N5S2   | 1      | 23.3           | 1.19                              | 2696.9441               | 9.8            |
| H6N5S3   | 1      | 21.7           | 0.49                              | 3004.0345               | 7.3            |
|          | 2      | 22.3           | 3.45                              | 3004.0345               | 5.2            |
|          | 3      | 22.9           | 5.90                              | 3004.0345               | 5.6            |
| H6N5F1S3 | 1      | 24.2           | 0.67                              | 3150.0924               | 9.7            |
|          | 2      | 24.6           | 0.55                              | 3150.0924               | 10.3           |

<sup>a</sup>Relative area was calculated as the peak area of each glycan divided by the sum of the peak areas of all glycans detected. <sup>b</sup>Theoretical mass of the glycan labeled with [<sup>12</sup>C<sub>6</sub>]-aniline.

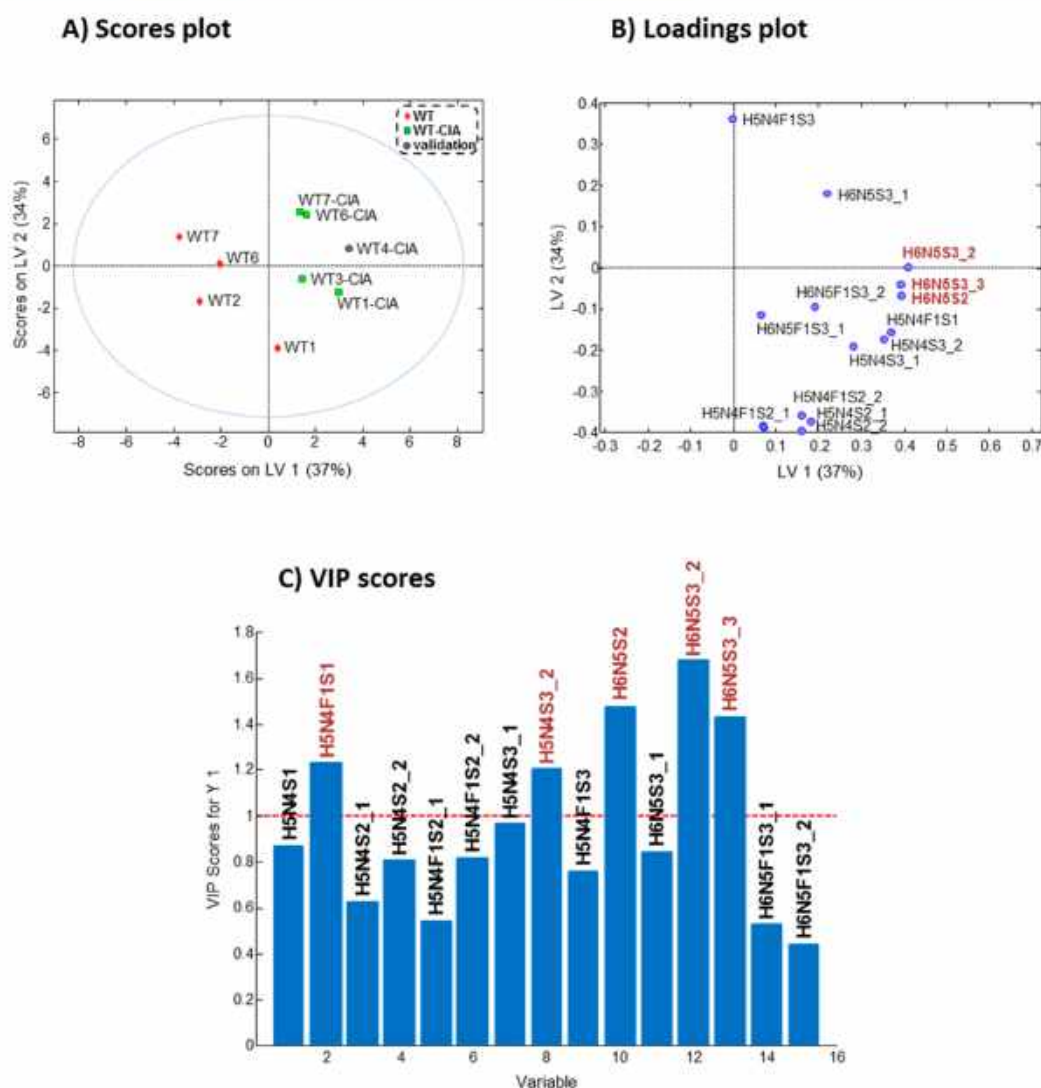
**Figure 2.** Extracted ion chromatograms (EICs) of the most abundant mTf glycan isomers detected by CapZIC-HILIC-MS in the control mouse serum after IAC purification under the optimized conditions.

To unequivocally determine the under or overexpression of mTf glycans in the pathological samples, a glycan relative isotope labeling (GRIL) approach with [<sup>12</sup>C<sub>6</sub>]/[<sup>13</sup>C<sub>6</sub>]-aniline previously established<sup>20</sup> was used. First, the reproducibility of the overall methodology (taking into account the GRIL method but also the sample treatment, including IAC purification) was evaluated. For this purpose, three independent control mouse serum samples ( $n = 3$ ) were IAC purified and labeled with [<sup>12</sup>C<sub>6</sub>]AN and another three with [<sup>13</sup>C<sub>6</sub>]AN. After derivatization, equimolar mixtures of mTf-glycan-[<sup>12</sup>C<sub>6</sub>]AN and mTf-glycan-[<sup>13</sup>C<sub>6</sub>]AN were prepared and analyzed by CapZIC-HILIC-MS by triplicate. Experimental ratios (area of the glycan labeled with [<sup>12</sup>C<sub>6</sub>]AN divided by the area of the same glycan isomer labeled with [<sup>13</sup>C<sub>6</sub>]AN) for all mTf glycans were close to 1 and relative standard deviations  $\leq 6\%$ , demonstrating the reproducibility of the established methodology. Afterward, mTf glycans isolated from each nonimmunized or CIA WT sample (labeled with [<sup>12</sup>C<sub>6</sub>]-aniline) were mixed (1:1) with mTf glycans obtained from the control mouse serum (labeled with [<sup>13</sup>C<sub>6</sub>]-aniline), and analyzed by CapZIC-HILIC-MS. Four nonimmunized and five CIA WT mice serum samples at different clinical scores were analyzed. The information about the clinical score,

gender, and age of each mouse serum sample is detailed in Materials and Methods section.

Experimental ratios (area of each glycan isomer in nonimmunized or CIA WT samples divided by the area of the same glycan isomer in the control mouse sample) for all mTf glycan isomers detected in the different samples were calculated and used for data interpretation in the multivariate data analysis approach. First, PCA was used to explore the data for the unsupervised identification of trends and detection of outliers. Four nonimmunized (WT1, WT2, WT6, and WT7) and four CIA (WT1-CIA, WT3-CIA, WT6-CIA, and WT7-CIA) WT samples were used to build the model. The scores plot for the first two principal components (a total of 77% of variance explained by the sum of PC 1 and PC 2) is shown in Supplementary Figure S3A. Despite the variability between the analyzed samples, two sample groups could be observed. The loadings plot (Supplementary Figure S3B) revealed that triantennary glycans were more related to CIA WT mice (they appeared in the right upper quadrant, the same region as most of CIA samples in the scores plot). In addition, outlier samples were not detected in the outlier detection plot (Q Residuals vs Hotelling  $T^2$ ) of Supplementary Figure S3C. However, PCA did not show the importance of each glycan isomer to differentiate between the two groups. Therefore, a partial least-squares discriminant analysis (PLS-DA) model was built in order to improve class separation and identify the glycan isomers that could be potentially used as biomarkers of CIA in mice. The same mice samples than for the PCA model were used as the calibration set classifying them in nonimmunized (WT) and CIA (WT-CIA) mice. As can be seen in the scores plot of Figure 3A, two latent variables (LVs) allowed discrimination between the two groups (37 and 34% of the X and Y variances explained, respectively). The loadings plot (Figure 3B) showed again that triantennary glycans, such as H6N5S2, H6N5S3\_2, and H6N5S3\_3, were more related to CIA. To complete this qualitative information, the VIP scores (Figure 3C) allowed us to quantify the influence of the different glycan isomers on the separation between WT and WT-CIA samples. As can be observed in Figure 3C, five out of fifteen glycan isomers were important (VIP > 1) to distinguish between nonimmunized and pathological samples. Among these glycan isomers, three of them corresponded to triantennary structures (H6N5S2, H6N5S3\_2, and H6N5S3\_3) and only one was fucosylated (H5N4F1S1). It is also worth mentioning that most of them were highly sialylated structures (H5N4S3\_2, H6N5S3\_2, and H6N5S3\_3). These results are in accordance with the previously described increase of mTf triantennary glycopeptide glycoforms in CIA samples<sup>11</sup> and also with the reported increase of highly sialylated Tf isoforms in human patients with rheumatoid arthritis (RA).<sup>6</sup> In addition, when characterizing human alpha-acid-glycoprotein (AGP) glycan isomers using exoglycosidase digestion and tandem mass spectrometry,<sup>19,20</sup> we demonstrated that isomers with higher proportion of  $\alpha$ 2-6 linked sialic acids were more retained in the ZIC-HILIC column. In this study, as most mTf glycan isomers considered biomarker candidates of CIA correspond to the most retained isomers within a glycan, it can be assumed that they present higher proportion of  $\alpha$ 2-6 linked sialic acids and hence that this linkage-type could be related to CIA diagnosis.

In order to validate the PLS-DA model, WT4-CIA sample was also analyzed and data processed following the same procedure described above. As can be observed in the scores

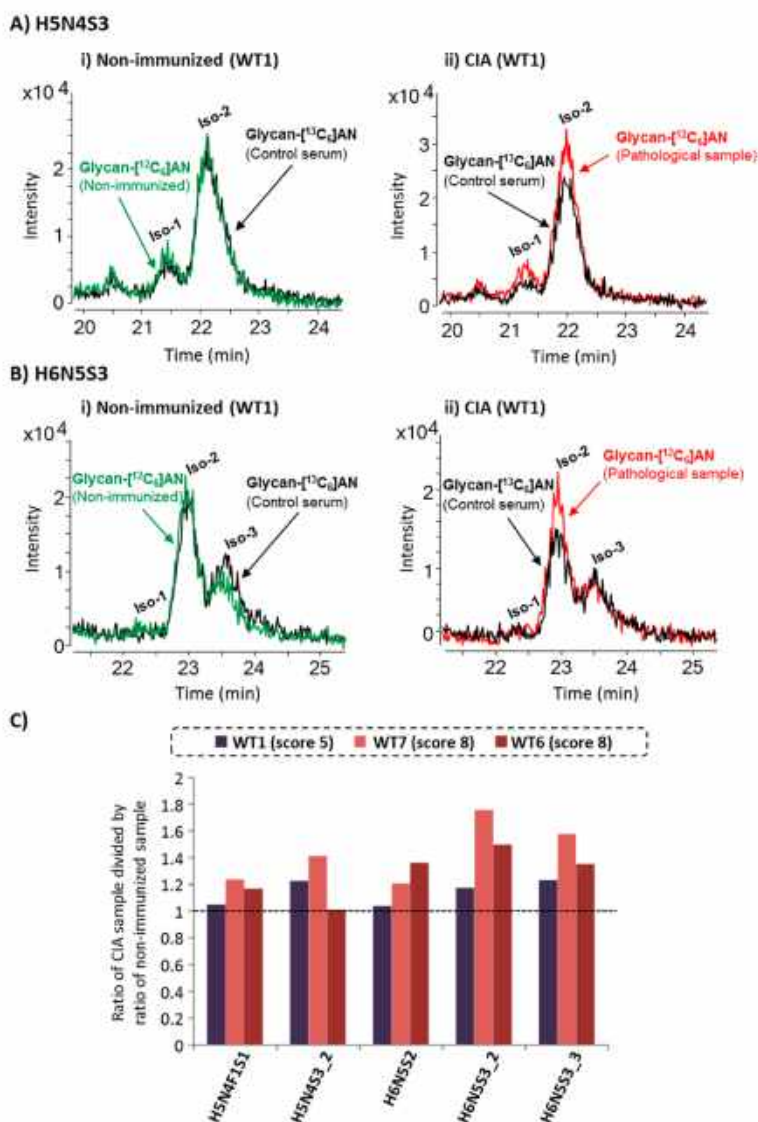


**Figure 3.** (A) Scores plot, (B) Loadings plot, and (C) VIP scores of the PLS-DA model applied to the ratios (area of the glycan isomer in the nonimmunized or CIA sample vs area of the same glycan isomer in the control mouse serum sample) obtained for each mTf glycan isomer in the nonimmunized and CIA WT mice samples analyzed. Calibration set: four nonimmunized samples (WT1, WT2, WT6, and WT7) and four CIA samples (WT1-CIA, WT3-CIA, WT6-CIA, and WT7-CIA). Validation sample: WT4-CIA.

plot of Figure 3A, class prediction was excellent, demonstrating the good performance of the model to discriminate between nonimmunized and CIA WT mice samples.

Some of the mice serum samples came from the same mouse before and after the induction of CIA (WT1, WT6, and WT7). Figure 4A and B shows, as an example, the extracted ion chromatograms (EICs) of H5N4S3 and H6N5S3 glycans, respectively, detected in WT1 mouse, before (nonimmunized) and after the induction of CIA, and compared to the control mouse serum (both glycans identified by PLS-DA as potential biomarkers of CIA). Having as a reference the peak signal of each isomer in the control mouse serum (EICs in black), both isomers of H5N4S3 as well as isomers 2 and 3 of H6N5S3 were upregulated in the CIA sample with respect to the nonimmunized one. By way of an example, the MS spectrum of isomer 2 of H6N5S3 as well as the  $[M-3H]^{3-}$  ion spectra of isomers 2 and 3 of the same glycan detected in WT1-CIA sample are shown in Supplementary Figure S4. As can be

observed, both chromatographic peaks exhibit the same isotopic envelope and identical accurate masses either in  $[^{12}C_6]AN$  and  $[^{13}C_6]AN$ , corroborating that they correspond to isobaric isomers. Moreover, in order to evaluate if the overexpression of these glycan isomers increase with the clinical score, the three pairs of serum samples, in which the same mouse was analyzed before and after the induction of CIA (WT1, WT6, and WT7), were used. In this regard, Figure 4C shows the ratio of the CIA samples (Area CIA/Area control) divided by the ratio of the nonimmunized ones (Area nonimmunized/Area control) obtained for the most meaningful mTf glycan isomers. As can be observed, all glycan isomers were more overexpressed in WT6 and WT7 (clinical score of 8) than in WT1 (clinical score of 5), with the exception of H5N4S3\_2 in WT6. These results suggested that the upregulation of these glycan isomers also seemed to be related with the CIA clinical score in mice, especially H6N5S2 and H6N5S3\_2, and therefore they could be used to study the



**Figure 4.** Extracted ion chromatograms (EICs) obtained for (A) H5N4S3 and (B) H6N5S3 glycans in nonimmunized (i) and CIA (ii) WT1 mouse samples, both with respect to control mouse serum. EICs of glycans in the nonimmunized sample are shown in green, in the CIA sample in red, and in the control mouse serum sample in black. (C) Bar graph of the ratio of the CIA samples (Area CIA/Area control) divided by the ratio of the nonimmunized ones (Area nonimmunized/Area control) obtained in WT1 (score 5), WT6 (score 8), and WT7 (score 8) mice samples.

response to novel treatments as well as to monitor this pathology.

#### 4. CONCLUDING REMARKS

Purification of Tf from mice serum samples was achieved by IAC using a CNBr-sepharose column activated with a polyclonal antihuman Tf antibody. The activation and coupling protocol as well as the IAC purification procedure were optimized in order to obtain a selective isolation of Tf from mice serum. Under the optimal conditions, SDS-PAGE and nanoLC-MS/MS results confirmed Tf was the major protein present in the eluted fraction and no other glycosylated proteins would interfere in the subsequent glycan analysis. Once established a reference glycomic profile of mTf, glycan isomers from nonimmunized and CIA serum samples were analyzed by CapZIC-HILIC-MS using the GRIL strategy with

$[^{12}\text{C}_6]/[^{13}\text{C}_6]$ -aniline. Multivariate data analysis tools showed that five out of fifteen mTf glycan isomers were important to distinguish CIA from control mice, three of them corresponding to triantennary structures (H6N5S2, H6N5S3\_2, and H6N5S3\_3), one to a highly sialylated biantennary glycan (H5N4S3\_2), and only one fucosylated (H5N4F1S1). In addition, these glycan isomers, especially H6N5S2 and H6N5S3\_2, seemed to be related with the progression of CIA, being more overexpressed in samples with higher clinical scores. Although a larger cohort of samples should be analyzed in the future to corroborate the biomarker value of these glycans, this study could be considered a starting point to find novel glycan-based biomarkers for CIA diagnosis. Moreover, the presented methodology could be also implemented to identify modifications in Tf glycosylation associated with rheumatoid arthritis or other inflammatory processes.

## ■ ASSOCIATED CONTENT

## ■ Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jproteome.0c00016>.

Figure S1: Bar graph of the glycans detected in the control mouse serum using the initial and the optimized IAC conditions; Figure S2: Bar graph of the glycans obtained from a pool of nonimmunized WT and from the control mouse serum using the optimized IAC conditions; Figure S3: PCA model applied to the ratios obtained for each mTf glycan isomer in all mice samples analyzed; Figure S4: MS spectra of isomer 2 of H6N5S3 and  $[M-3H]^{3-}$  ion spectra of isomers 2 and 3 of the same glycan in the WT1-CIA sample (PDF)

## ■ AUTHOR INFORMATION

## Corresponding Author

**Victoria Sanz-Nebot** – Department of Chemical Engineering and Analytical Chemistry, Institute for Research on Nutrition and Food Safety (INSA-UB), University of Barcelona, 08028 Barcelona, Spain; [orcid.org/0000-0002-3048-0573](https://orcid.org/0000-0002-3048-0573); Phone: (+34) 934021283; Email: [vsanz@ub.edu](mailto:vsanz@ub.edu); Fax: (+34) 934021233

## Authors

**Montserrat Mancera-Artu** – Department of Chemical Engineering and Analytical Chemistry, Institute for Research on Nutrition and Food Safety (INSA-UB), University of Barcelona, 08028 Barcelona, Spain

**Estela Giménez** – Department of Chemical Engineering and Analytical Chemistry, Institute for Research on Nutrition and Food Safety (INSA-UB), University of Barcelona, 08028 Barcelona, Spain; [orcid.org/0000-0002-0700-3701](https://orcid.org/0000-0002-0700-3701)

**Jaime Sancho** – Instituto de Parasitología y Biomedicina "López-Neyra" (IPLBN), CSIC, 18016 Granada, Spain

Complete contact information is available at:

<https://pubs.acs.org/doi/10.1021/acs.jproteome.0c00016>

## Author Contributions

Designed the study (M.M.-A., E.G., J.S., and V.S.-N.), collected mice serum samples (J.S.), carried out serum purification (M.M.-A.), performed GRIL-mass spectrometry and data analysis (M.M.-A.), wrote the manuscript (M.M.-A., E.G., and V.S.-N.). All authors reviewed the manuscript.

## Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

This work was supported by a grant from the Spanish Ministry of Economy and Competitiveness (RTI2018-097411-B-I00) and the Cathedra UB Rector Francisco Buscarons Ubeda (Forensic Chemistry and Chemical Engineering). Montserrat Mancera-Artu acknowledges the University of Barcelona for an ADR fellowship. The authors thank Dr. Marina Gay, from the Mass Spectrometry and Proteomics Unit of the Institute for Research in Biomedicine (IRB Barcelona), for her helpful assistance with data interpretation in the nanoLC-MS/MS analysis.

## ■ REFERENCES

- (1) Gornik, O.; Lauc, G. Glycosylation of serum proteins in inflammatory diseases. *Dis. Markers* **2008**, *25*, 267–278.
- (2) Reily, C.; Stewart, T. J.; Renfrow, M. B.; Novak, J. Glycosylation in health and disease. *Nat. Rev. Nephrol.* **2019**, *15*, 346–366.
- (3) Albrecht, S.; Unwin, L.; Muniyappa, M.; Rudd, P. M. Glycosylation as a marker for inflammatory arthritis. *Cancer Biomarkers* **2014**, *14*, 17–28.
- (4) Majithia, V.; Geraci, S. A. Rheumatoid arthritis: Diagnosis and management. *Am. J. Med.* **2007**, *120*, 936–939.
- (5) Feelders, R. A.; Vreugdenhi, G.; de Jong, G.; Swaak, A. J. G.; van Eijk, H. G. Transferrin microheterogeneity in rheumatoid arthritis. *Rheumatol. Int.* **1992**, *12*, 195–199.
- (6) Gudowska, M.; Gruszewska, E.; Wrona, A.; Gindzienska-Sieskiewicz, E.; Domyslawska, I.; Lipartowska-Klimuk, K.; Cylwik, B.; Sierakowski, S.; Chrostek, L. The Profile of Serum Transferrin Isoforms in Rheumatoid Arthritis. *J. Clin. Rheumatol.* **2019**, *25*, 159–162.
- (7) Brand, D. D.; Latham, K. A.; Rosloniec, E. F. Collagen-induced arthritis. *Nat. Protoc.* **2007**, *2*, 1269–1275.
- (8) Gwon, S. Y.; Rhee, K. J.; Sung, H. J. Gene and protein expression profiles in a mouse model of collagen-induced arthritis. *Int. J. Med. Sci.* **2018**, *15*, 77–85.
- (9) Inglis, J. J.; Criado, G.; Medghalchi, M.; Andrews, M.; Sandison, A.; Feldmann, M.; Williams, R. O. Collagen-induced arthritis in C57BL/6 mice is associated with a robust and sustained T-cell response to type II collagen. *Arthritis Res. Ther.* **2007**, *9*, 1–8.
- (10) Rosal-Vela, A.; García-Rodríguez, S.; Postigo, J.; Iglesias, M.; Longobardo, V.; Lario, A.; Merino, J.; Merino, R.; Zubiaur, M.; Sancho, J. Distinct serum proteome profiles associated with collagen-induced arthritis and complete Freund's adjuvant-induced inflammation in CD38<sup>-/-</sup> mice: The discriminative power of protein species or proteoforms. *Proteomics* **2015**, *15*, 3382–3393.
- (11) Rosal-Vela, A.; Barroso, A.; Giménez, E.; García-Rodríguez, S.; Longobardo, V.; Postigo, J.; Iglesias, M.; Lario, A.; Merino, J.; Merino, R.; Zubiaur, M.; Sanz-Nebot, V.; Sancho, J. Identification of multiple transferrin species in the spleen and serum from mice with collagen-induced arthritis which may reflect changes in transferrin glycosylation associated with disease activity: The role of CD38. *J. Proteomics* **2016**, *134*, 127–137.
- (12) Giménez, E.; Balmaña, M.; Figueras, J.; Fort, E.; de Bolós, C.; Sanz-Nebot, V.; Peracaula, R.; Rizzi, A. Quantitative analysis of N-glycans from human alpha-acid-glycoprotein using stable isotope labeling and zwitterionic hydrophilic interaction capillary liquid chromatography electrospray mass spectrometry as tool for pancreatic disease diagnosis. *Anal. Chim. Acta* **2015**, *866*, 59–68.
- (13) Mancera-Artu, M.; Giménez, E.; Balmaña, M.; Barrabés, S.; Albiol-Quer, M.; Fort, E.; Peracaula, R.; Sanz-Nebot, V. Multivariate data analysis for the detection of human alpha-acid glycoprotein aberrant glycosylation in pancreatic ductal adenocarcinoma. *J. Proteomics* **2019**, *195*, 76–87.
- (14) Postigo, J.; Iglesias, M.; Cerezo-Wallis, D.; Rosal-Vela, A.; García-Rodríguez, S.; Zubiaur, M.; Sancho, J.; Merino, R.; Merino, J. Mice deficient in CD38 develop an attenuated form of collagen type II-induced arthritis. *PLoS One* **2012**, *7*, 1–9.
- (15) Barroso, A.; Giménez, E.; Benavente, F.; Barbosa, J.; Sanz-Nebot, V. Analysis of human transferrin glycopeptides by capillary electrophoresis and capillary liquid chromatography-mass spectrometry. Application to diagnosis of alcohol dependence. *Anal. Chim. Acta* **2013**, *804*, 167–175.
- (16) Kavran, J. M.; Leahy, D. J. Coupling Antibody to Cyanogen Bromide-Activated Sepharose. *Methods Enzymol.* **2014**, *541*, 27–34.
- (17) Giménez, E.; Ramos-Hernán, R.; Benavente, F.; Barbosa, J.; Sanz-Nebot, V. Analysis of recombinant human erythropoietin glycopeptides by capillary electrophoresis electrospray-time of flight-mass spectrometry. *Anal. Chim. Acta* **2012**, *709*, 81–90.
- (18) *Product Information ProteoExtract—Albumin/IgG Removal Kit*; Calbiochem, 2004; pp 1–11.

(19) Mancera-Arteu, M.; Giménez, E.; Barbosa, J.; Sanz-Nebot, V. 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* **2016**, *940*, 92–103.

(20) Giménez, E.; Sanz-Nebot, V.; Rizzi, A. Relative quantitation of glycosylation variants by stable isotope labeling of enzymatically released N-glycans using  $[^{12}\text{C}]/[^{13}\text{C}]$  aniline and ZIC-HILIC-ESI-TOF-MS. *Anal. Bioanal. Chem.* **2013**, *405*, 7307–7319.

(21) Jolliffe, I. T.; Morgna, B. J. Principal Component Analysis and exploratory factor analysis. *Stat. Methods Med. Res.* **1992**, *1*, 69–95.

(22) Wold, S.; Sjöström, M.; Eriksson, L. PLS-regression: A basic tool of chemometrics. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109–130.

(23) Barker, M.; Rayens, W. Partial least squares for discrimination. *J. Chemom.* **2003**, *17*, 166–173.

(24) Cawley, G. C.; Talbot, N. L. C. Efficient leave-one-out cross-validation of kernel fisher discriminant classifiers. *Pattern Recognit.* **2003**, *36*, 2585–2592.

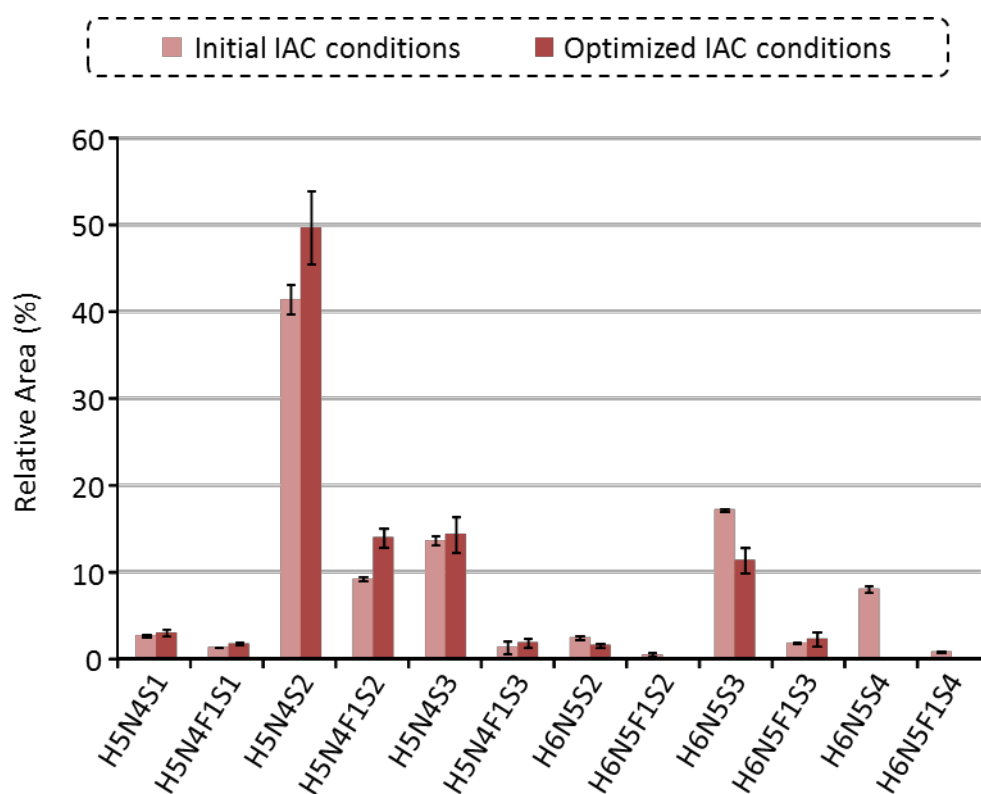
(25) Coddeville, B.; Regoeczi, E.; Strecker, G.; Plancke, Y.; Spik, G. Structural analysis of trisialylated biantennary glycans isolated from mouse serum transferrin. Characterization of the sequence Neu5Gc( $\alpha$ 2–3)Gal( $\beta$ 1–3)[Neu5Gc( $\alpha$ 2–6)]GlcNAc( $\beta$ 1–2)Man. *Biochim. Biophys. Acta, Gen. Subj.* **2000**, *1475*, 321–328.

(26) Mancera-Arteu, M.; Giménez, E.; Barbosa, J.; Peracaula, R.; Sanz-Nebot, V. Zwitterionic-hydrophilic interaction capillary liquid chromatography coupled to tandem mass spectrometry for the characterization of human alpha-acid-glycoprotein N-glycan isomers. *Anal. Chim. Acta* **2017**, *991*, 76–88.

## Supplementary Figures

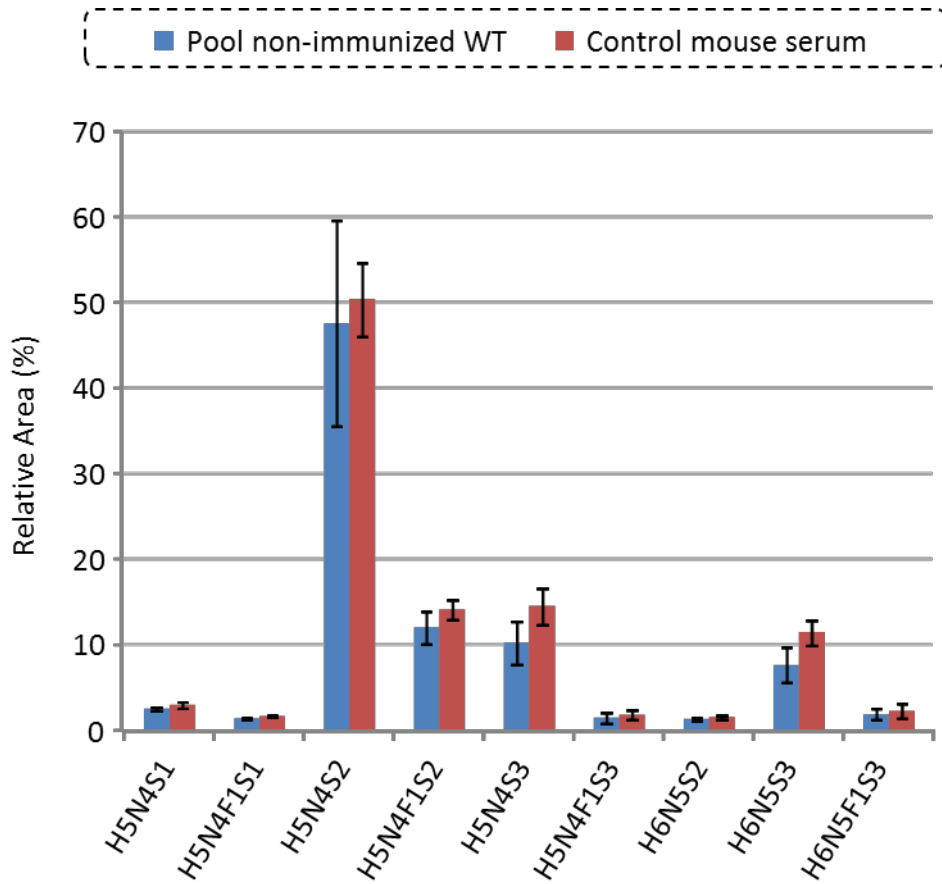
---



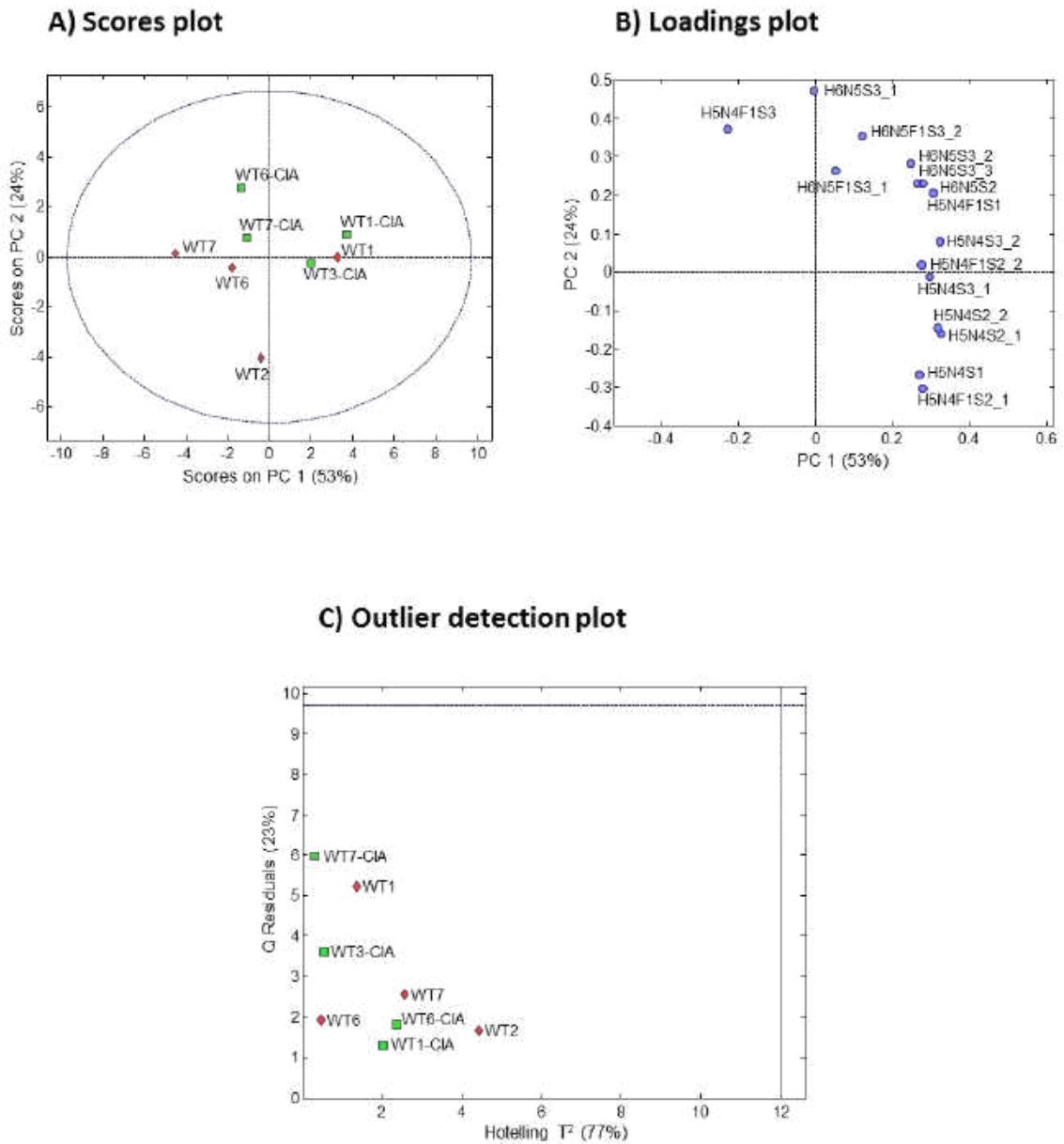


**Supplementary Figure S-1.** Bar graph of the relative area of the glycans detected by  $\mu$ ZIC-HILIC-MS in the control mouse serum using the initial and the optimized IAC conditions, with the corresponding standard deviations associated (n=3). Relative area was calculated as the peak area of each glycan divided by the sum of the peak areas of all glycans detected.





**Supplementary Figure S-2.** Bar graph of the average relative area of the glycans obtained from a pool of non-immunized WT and from the control mouse serum using the optimized IAC conditions. Relative area was calculated as the peak area of each glycan divided by the sum of the peak areas of all glycans detected.



**Supplementary Figure S-3.** A) Scores plot, B) Loadings plot and C) Outlier detection plot of the PCA model applied to the ratios (area of the glycan isomer in the non-immunized or CIA sample vs. area of the same glycan isomer in the control mouse serum sample obtained for each mTf glycan isomer in the non-immunized and CIA WT mice samples analyzed).



## Capítulo 5. Resultados y Discusión

---



## 5.1. Establecimiento de metodologías para la purificación de glicopéptidos

Como se comentó en la introducción de esta tesis doctoral, el análisis de glicopéptidos se utiliza habitualmente como estrategia *bottom-up* en estudios glicoproteómicos para caracterizar por espectrometría de masas (MS) las estructuras de los glicanos, así como los diferentes puntos de glicosilación donde se encuentran unidos a la glicoproteína de interés. Sin embargo, las dificultades para detectar estos analitos por MS, debido a su baja concentración y a la supresión iónica que experimentan en presencia de los péptidos que se encuentran en los digestos, han dado lugar al establecimiento de diversas técnicas destinadas a purificar de manera selectiva los glicopéptidos antes de ser analizados por MS [106,107,115].

En esta tesis doctoral, se han establecido dos métodos para la purificación de glicopéptidos de proteínas de interés biomédico o biofarmacéutico. Estos métodos se han aplicado al análisis de glicopéptidos por CE-MS dadas las ventajas que presenta esta técnica de separación frente a la LC-MS en el análisis de glicopéptidos [68,206]. Aunque se han analizado los digestos trípticos de diversas glicoproteínas para evaluar las metodologías de purificación establecidas, ambos métodos (descritos en los **artículos 2.1 y 2.2**) se han centrado en el enriquecimiento de las glicofomas del O-glicopéptido de la eritropoyetina humana recombinante (rhEPO) dado su interés como biofármaco y como agente dopante [79–83].

### 5.1.1. Purificación de glicopéptidos mediante precipitación con acetona

Aunque existen diversos métodos descritos en la bibliografía para la purificación de glicopéptidos empleando cromatografía de afinidad con lectinas, cromatografía de interacción hidrofílica así como extracción con partículas de dióxido de titanio (TiO<sub>2</sub>) [108,116], estas metodologías son generalmente caras, laboriosas y requieren mucho tiempo para llevarse a cabo. Con el fin de desarrollar un método rápido, simple y económico para el enriquecimiento de los glicopéptidos presentes en una muestra, Takakura *et. al.* desarrollaron un método

basado en su precipitación selectiva con acetona [118]. Este estudio se centró en la diferente solubilidad de péptidos y glicopéptidos en este disolvente para lograr su separación, pero sin investigar las diferencias de comportamiento entre N- y O-glicopéptidos.

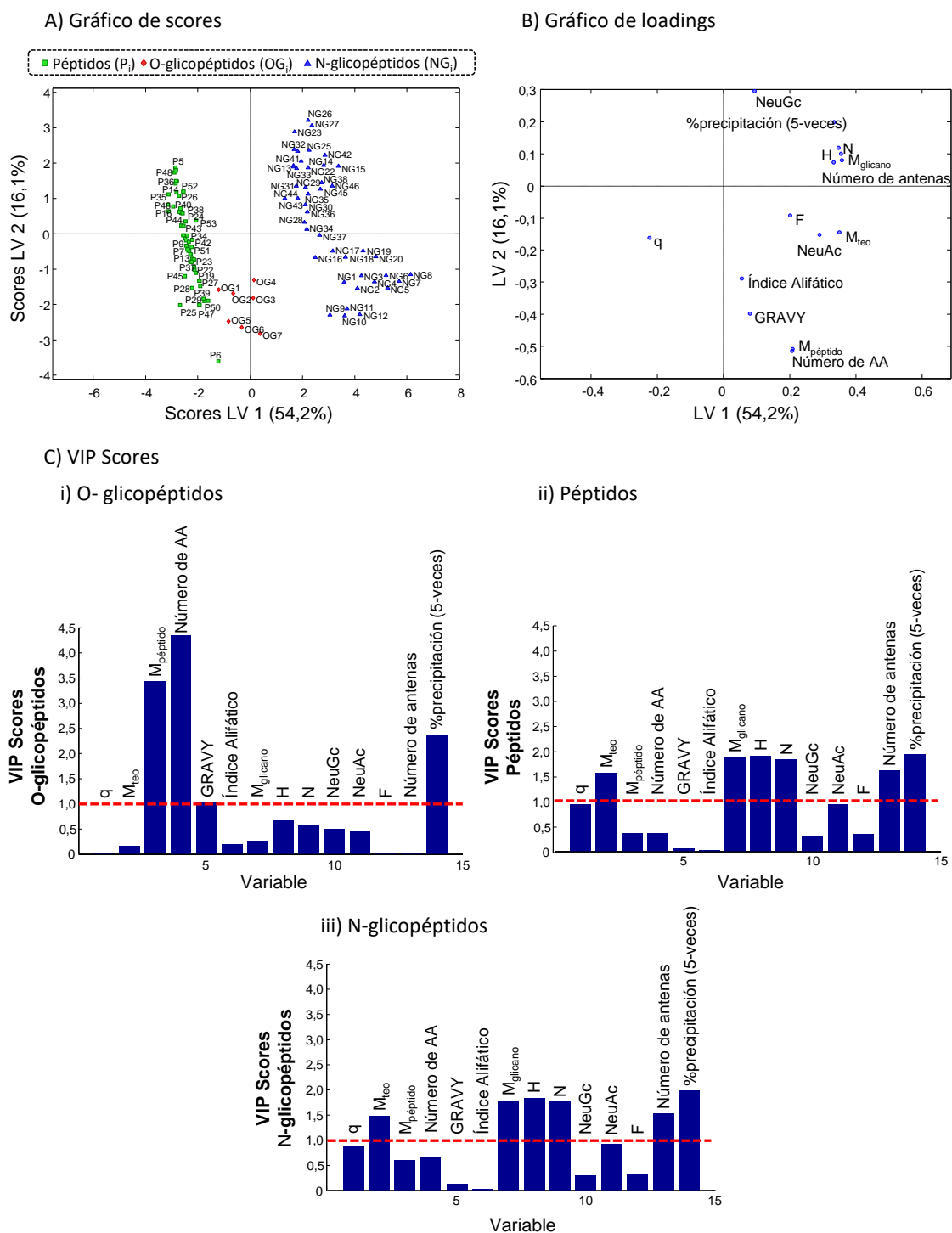
En esta tesis, se ha estudiado en profundidad la precipitación con acetona de los N- y O-glicopéptidos en función de su naturaleza, obtenidos de digestos enzimáticos de diversas glicoproteínas de interés biofarmacéutico o biomédico: rhEPO, apolipoproteína C-III (APO-C3), alfa-1-glicoproteína humana y bovina (hAGP y bAGP) y ovalbúmina (OVA) (**artículo 2.1**). Asimismo, se ha estudiado la influencia en la precipitación de diversos parámetros físico-químicos y propiedades relacionadas con la composición de los carbohidratos y la estructura del glicopéptido, mediante métodos de análisis de datos multivariantes. Finalmente, dado el interés de ciertas glicofomas del O-glicopéptido de la rhEPO en el control de calidad de este biofármaco y en el control antidopaje, se han seleccionado las condiciones para su óptima precipitación.

#### **5.1.1.1. Estudio de la precipitación de péptidos y glicopéptidos**

En primer lugar se investigaron las condiciones propuestas por Takakura *et.al.* [118] para la precipitación de los péptidos, N- y O-glicopéptidos presentes en los digestos enzimáticos de las proteínas mencionadas, utilizando un volumen de acetona 5 veces superior al de digesto. Todas las digestiones se llevaron a cabo con tripsina excepto en el caso de la OVA, cuya digestión dio lugar a un N-glicopéptido demasiado grande (secuencia peptídica de 32 aminoácidos) y difícil de detectar por MS, por lo que se utilizó quimotripsina para la digestión. Los sobrenadantes y precipitados fueron analizados por CE-MS mediante un método desarrollado previamente por nuestro grupo de investigación para el análisis de glicopéptidos [85,86]. Se detectaron un total de 53 péptidos, 46 N-glicopéptidos y 7 O-glicopéptidos con diferentes estructuras y composición de glicanos. Con estas condiciones de precipitación

(exceso de 5 veces de acetona) los N-glicopéptidos se detectaron principalmente en el precipitado, mientras que los péptidos permanecieron en el sobrenadante y los O-glicopéptidos precipitaron parcialmente. Con el fin de conocer la influencia de diversos parámetros físico-químicos y propiedades de los péptidos y glicopéptidos en su solubilidad en acetona, se emplearon herramientas quimiométricas como el análisis por componentes principales (PCA) y el análisis discriminante por mínimos cuadrados parciales (PLS-DA). Se tuvieron en cuenta variables asociadas: 1) a toda la molécula (masa teórica,  $M_{teo}$  y carga,  $q$ ), 2) a la cadena peptídica (masa del péptido,  $M_{péptido}$ ; número de aminoácidos, número de AA; promedio general de hidropatía, GRAVY y índice alifático) y 3) a los glicanos (masa del glicano ( $M_{glicano}$ ), número y tipo de monosacáridos (H, N, NeuAc, NeuGc) y número de antenas. La composición de cada uno de los péptidos y glicopéptidos identificados, los valores de cada una de las variables mencionadas anteriormente, así como el porcentaje de precipitación obtenido en cada caso a partir del análisis por CE-MS, se detallan en el material suplementario del **artículo 2.1** de esta tesis doctoral. En primer lugar, para explorar los datos y estudiar la existencia de tendencias o grupos así como detectar la presencia de *outliers*, se aplicó PCA al conjunto de parámetros físico-químicos y propiedades de cada péptido/glicopéptido así como a los porcentajes de precipitación obtenidos. Tal y como se explica en detalle en el **artículo 2.1**, se observó la presencia de tres clases, correspondientes a péptidos ( $P_i$ ), N-glicopéptidos ( $NG_i$ ) y O-glicopéptidos ( $OG_i$ ), mientras que los glicopéptidos NG39 y NG40 de la OVA fueron considerados *outliers*. A continuación se utilizó PLS-DA con el objetivo de mejorar la separación entre clases y facilitar la identificación de las variables más importantes en la precipitación de péptidos y glicopéptidos. La **Figura 5.1** muestra los gráficos de *scores* (A), *loadings* (B) y *VIP scores* (C) obtenidos en dicho modelo.





**Figura 5.1.-** (A) Gráfico de scores y (B) gráfico de loadings del modelo de PLS-DA aplicado al porcentaje de precipitación, añadiendo un exceso de cinco veces de acetona, y a los diferentes parámetros físico-químicos y propiedades de péptidos/glicopéptidos obtenidos de los digestos de OVA, hAGP, bAGP, APO-C3 y rhEPO (ver artículo 2.1). (C) VIP scores de las diferentes variables cuando se considera la separación de (i) O-glicopéptidos con respecto a péptidos y N-glicopéptidos, (ii) péptidos con respecto a N- y O-glicopéptidos y (iii) N-glicopéptidos con respecto a péptidos y O-glicopéptidos.

Como puede observarse en el gráfico de *scores*, dos variables latentes (LVs) fueron suficientes para separar las tres clases de compuestos con un total del 70.3% de la variancia explicada, siendo LV1 la más útil para diferenciar entre clases de compuestos. En este sentido, el gráfico de *loadings* mostró que las variables relacionadas con la estructura de los glicanos y su composición, tales como  $M_{\text{teo}}$ ,  $M_{\text{glicano}}$ , número de monosacáridos H, N y NeuAc, número de antenas y q, eran las más importantes para la separación entre péptidos, N- y O-glicopéptidos. Además, los gráficos de *VIP scores*, desvelaron que  $M_{\text{teo}}$ ,  $M_{\text{glicano}}$ , número de H y N así como el número de antenas eran las variables más relevantes para diferenciar el comportamiento tanto de péptidos como de N-glicopéptidos del resto (**Figuras 5.1C-ii y 1-iii**). Por el contrario, tal y como muestra la **Figura 5.1C-i**,  $M_{\text{péptido}}$ , número de AA y, en menor medida, GRAVY fueron las variables más significativas en la diferenciación de los O-glicopéptidos respecto al resto de compuestos. Teniendo en cuenta estos resultados y los obtenidos en el modelo de PLS-DA aplicado sólo a péptidos y O-glicopéptidos (ver **artículo 2.1**), se puede concluir que prácticamente todos los N-glicopéptidos precipitan cuantitativamente empleando un exceso de cinco veces de acetona ya que presentan un elevado porcentaje de glicosilación ( $\geq 45\%$  (m/m)) y una mayor  $M_{\text{teo}}$ , mientras que los péptidos permanecen mayoritariamente en disolución. En el caso de los O-glicopéptidos, el número de AA,  $M_{\text{péptido}}$ , así como su hidrofobicidad (GRAVY e índice alifático) y menor porcentaje de glicosilación ( $\leq 40\%$  (m/m)) respecto a los N-glicopéptidos, los hace más parecidos estructuralmente a los péptidos, lo que podría explicar su diferente comportamiento presentando una precipitación parcial con esta proporción de acetona.

#### 5.1.1.2. Estudio de la precipitación de O-glicopéptidos

Para estudiar con más detalle el comportamiento de los O-glicopéptidos, se analizaron por CE-MS tanto los sobrenadantes como los precipitados obtenidos empleando una proporción de acetona 5 y 8 veces superior a la de digesto de rhEPO y APO-C3. Tal y como se muestra en la

**Tabla 5.1.** se obtuvieron mayores porcentajes de precipitación para los O-glicopéptidos de ambas proteínas al aumentar la cantidad de acetona adicionada, pero los porcentajes de precipitación para las glicofomas del O-glicopéptido de la APO-C3 (O<sub>74</sub>) continuaron siendo bajos comparados con los del O-glicopéptido de la rhEPO (O<sub>126</sub>).

**Tabla 5.1.-** Porcentaje de precipitación obtenido para las glicofomas de los glicopéptidos O<sub>126</sub> de la rhEPO y O<sub>74</sub> de la APO-C3 usando un exceso de cinco y ocho veces de acetona.

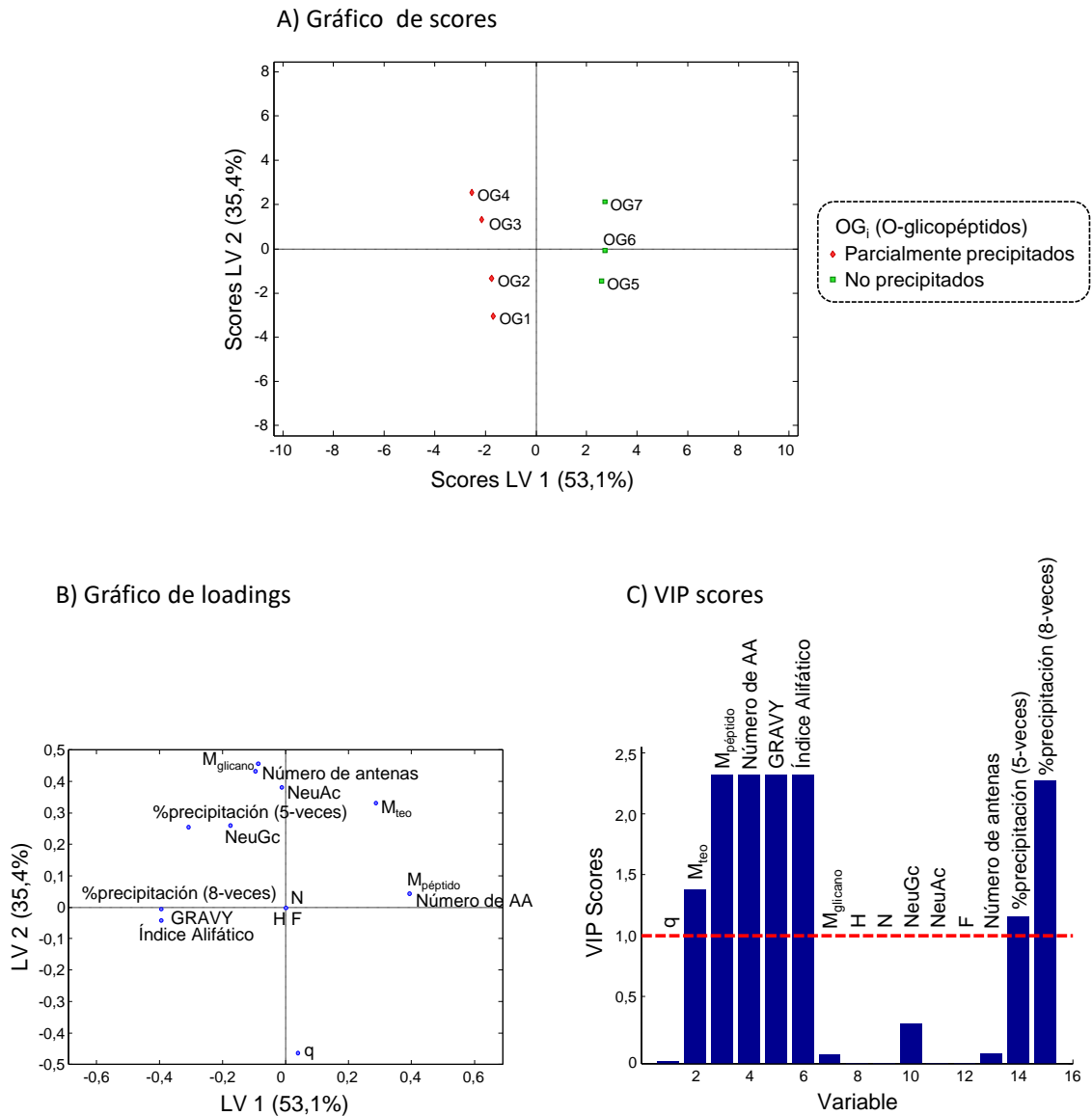
| Secuencia peptídica                          | Glicano   | Símbolo <sup>a</sup> | % precipitación |         |
|--|-----------|----------------------|-----------------|---------|
|  |           |                      | 5-veces         | 8-veces |
| EAISPPDAASAAPLR<br>(O <sub>126</sub> )       | H1N1      | OG1                  | 0               | 71,5    |
|  | H1N1S1    | OG2                  | 7,5             | 59,8    |
|  | H1N1S2    | OG3                  | 13,4            | 77,2    |
|  | H1N1S1S1* | OG4                  | 14,2            | 75,3    |
| FSEFWDLDPVTRTS<br>AVAA<br>(O <sub>74</sub> ) | H1N1      | OG5                  | 0               | 0       |
|  | H1N1S1    | OG6                  | 0               | 1,9     |
|  | H1N1S2    | OG7                  | 0               | 2,1     |

Punto de glicosilación marcado en rojo.

<sup>a</sup> OG<sub>i</sub>: O-glicopéptidos.

\* Un ácido siálico (H1N1S1S1) es NeuGc en lugar de NeuAc.

Con el fin de encontrar la razón de estas diferencias entre O-glicopéptidos, se construyó un modelo de PCA y, seguidamente, uno de PLS-DA únicamente con los O-glicopéptidos de la rhEPO y de la APO-C3, incluyendo todos los parámetros físico-químicos y propiedades citados anteriormente, así como los porcentajes de precipitación obtenidos con un exceso de 5 y 8 veces de acetona. Los resultados obtenidos en los gráficos de scores (A), loadings (B) y VIP scores (C) al aplicar el modelo de PLS-DA se muestran en la **Figura 5.2**. Tanto el gráfico de loadings como el de VIP scores (**Figura 5.2B y C**) mostraron que los parámetros relacionados con la hidrofobicidad de las cadenas peptídicas (GRAVY e índice alifático), así como el tamaño del péptido ( $M_{\text{péptido}}$  y número de AA) eran esenciales para explicar la diferente tendencia a precipitar de los O-glicopéptidos. Puesto que ambos glicopéptidos presentan una hidrofobicidad similar, se concluyó que la menor tendencia a la precipitación de las glicofomas del O-glicopéptido de la APO-C3 era debida a su mayor masa y longitud de la cadena peptídica con respecto al O-glicopéptido de la rhEPO.

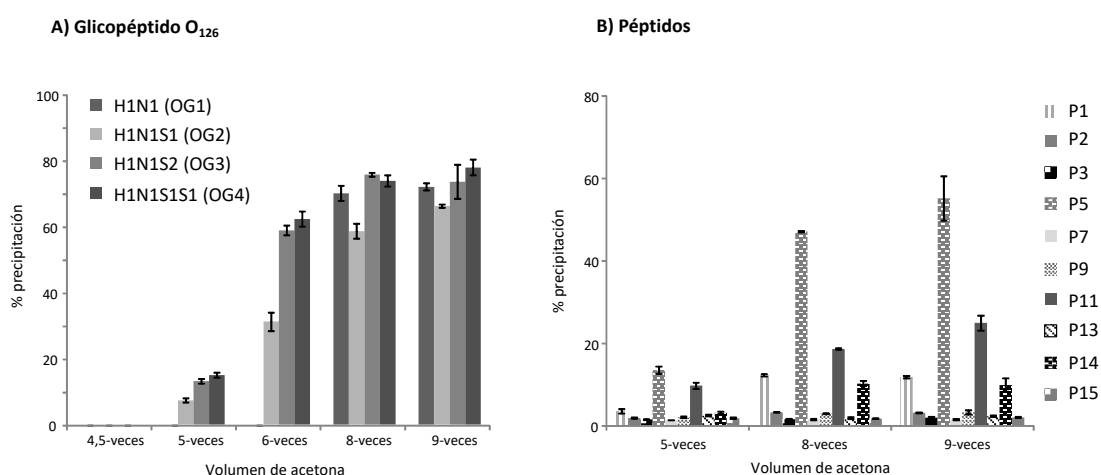


**Figura 5.2.-** (A) Gráfico de scores, (B) gráfico de loadings y (C) VIP scores del modelo de PLS-DA aplicado al porcentaje de precipitación obtenido, añadiendo un exceso de 5 y 8 veces de acetona, y a los diferentes parámetros físico-químicos y propiedades de los O-glicopéptidos obtenidos de los digestos de rhEPO y APO-C3 (ver artículo 2.1).

### 5.1.1.3. Optimización de la precipitación del glicopéptido O<sub>126</sub> de la rhEPO

Los resultados previos demuestran que la cantidad de acetona se puede optimizar para conseguir la precipitación selectiva de un determinado O-glicopéptido de interés. Dada la importancia de la rhEPO como biofármaco y agente dopante, en este estudio se seleccionó el O-glicopéptido de la rhEPO como modelo (O<sub>126</sub>), con el objeto de optimizar la cantidad de

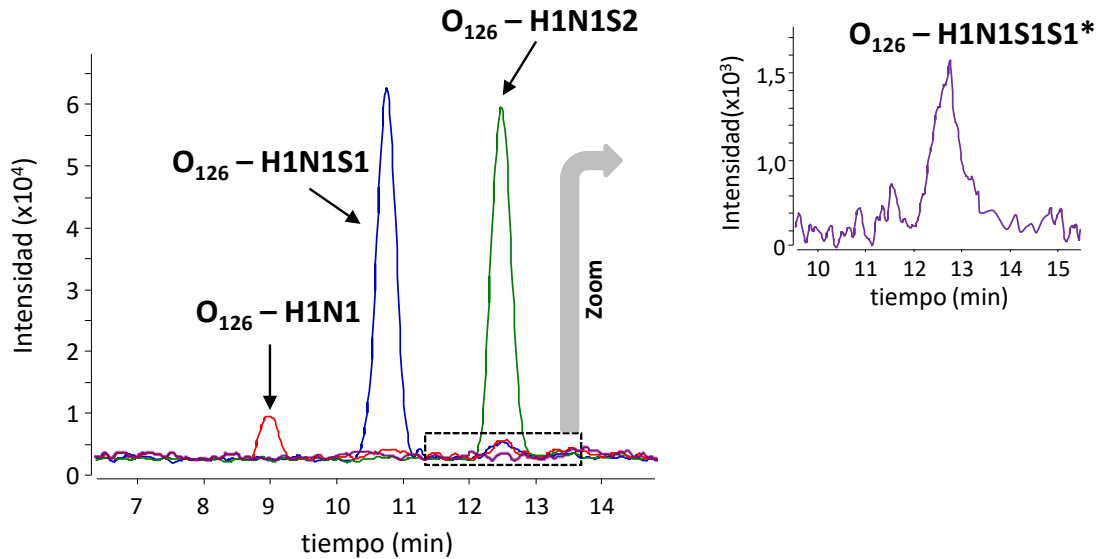
acetona necesaria para precipitar las glicofomas del O<sub>126</sub> evitando la coprecipitación de los péptidos del digesto. Con este propósito, se añadieron cinco cantidades diferentes de acetona (excesos de 4,5, 5, 6, 8 y 9 veces de volumen de acetona) a diferentes digestos trípticos de rhEPO y se analizaron por CE-MS los sobrenadantes y los precipitados. La **Figura 5.3** muestra los gráficos de barras del porcentaje de precipitación obtenido en función del volumen de acetona añadido para las glicofomas del O<sub>126</sub> (**Figura 5.3A**) y para algunos péptidos representativos del digesto de la rhEPO (**Figura 5.3B**).



**Figura 5.3.-** Gráfico de barras con el porcentaje de precipitación obtenido añadiendo diferentes proporciones de acetona para (A) las glicofomas del glicopéptido O<sub>126</sub> y (B) algunos péptidos representativos de la rhEPO.

Como se puede observar, las glicofomas del O<sub>126</sub> empiezan a precipitar con un exceso de acetona de 5 veces, primero aquellas más glicosiladas y con mayor número de ácidos siálicos (H1N1S1, H1N1S2 y H1N1S1S1). Por el contrario, la glicofoma no sialilada (H1N1) no inicia su precipitación hasta que se alcanza un exceso de acetona de 8 veces. Los péptidos, en cambio, se mantienen mayoritariamente en disolución, aunque los más hidrofílicos (P1, P5, P11, P14) comienzan a precipitar a partir de un exceso de 5 veces de acetona, aumentando su porcentaje de precipitación al incrementar el volumen de acetona añadido (**Figura 5.3B**). A fin de maximizar la recuperación de todas las glicofomas del O-glicopéptido de la rhEPO, se

seleccionó un exceso de 8 veces de acetona ya que se obtenían porcentajes de precipitación elevados (~70%) evitando en gran medida la coprecipitación de la mayoría de péptidos del digesto. La **Figura 5.4** muestra, a modo de ejemplo, los electroferogramas de iones extraídos (EIEs) de las glicofomas más abundantes del  $O_{126}$  en la fracción precipitada empleando la proporción de acetona seleccionada.



**Figura 5.4.-** EIEs de las glicofomas más abundantes del glicopéptido  $O_{126}$  de la rhEPO obtenidas en la fracción precipitada usando un exceso de 8 veces de acetona. Se muestra una ampliación del EIE del H1N1S1S1 dada su baja abundancia. (\*) Uno de los ácidos siálicos es NeuGc en lugar de NeuAc.

### 5.1.2. Purificación de glicopéptidos mediante $TiO_2$ -SPE-CE-MS

Las partículas de  $TiO_2$  han sido usadas para la purificación *off-line* de fosfopéptidos y de otros compuestos fosforilados [113,114,207]. De manera menos común, el uso de estas partículas también ha sido descrito para la purificación de glicopéptidos sialilados [116]. Sin embargo, los métodos de purificación *off-line* son laboriosos, aumentan de manera significativa el tiempo de análisis y no permiten la automatización. Una de las mejores alternativas a los métodos *off-line* consiste en acoplar la extracción en fase sólida en línea a la CE (SPE-CE). La preconcentración por SPE-CE se basa en la interacción reversible de los analitos de la muestra con una fase

estacionaria que se encuentra en un microcartucho de extracción o preconcentrador, cerca de la entrada del capilar de separación. Esta fase estacionaria permite retener los analitos presentes en un volumen elevado de muestra (~100  $\mu\text{L}$ ), que posteriormente se eluyen directamente dentro del capilar de separación con un pequeño volumen de una disolución adecuada (~50 nL). De esta manera, se consigue purificar el analito de interés y aumentar su concentración, disminuyendo consecuentemente los límites de detección (LODs). En nuestro grupo de investigación se han establecido diversos métodos de purificación mediante SPE-CE empleando fases estacionarias muy diversas para el análisis de una gran variedad de moléculas pequeñas, péptidos y proteínas [208–210].

En esta tesis doctoral, se evaluó el uso de partículas de  $\text{TiO}_2$  como sorbente en SPE-CE-MS para purificar selectivamente los glicopéptidos obtenidos de digestos enzimáticos de glicoproteínas. El O-glicopéptido de la rhEPO fue utilizado de nuevo como modelo para optimizar la metodología, aunque también se evaluó el método establecido para el análisis de los N-glicopéptidos de esta glicoproteína. Finalmente y para comprobar la validez de la metodología desarrollada, se aplicó a la purificación de glicopéptidos de otras glicoproteínas, en concreto al O-glicopéptido de la APO-C3 y a los N-glicopéptidos de la bAGP.

#### **5.1.2.1. Construcción del preconcentrador**

El preconcentrador con partículas de  $\text{TiO}_2$  utilizado en esta tesis (**artículo 2.2**) no se encuentra disponible comercialmente, sino que ha sido preparado en el laboratorio. El cuerpo del preconcentrador consiste en un pequeño fragmento de capilar de sílice fundida de 0,7 cm de longitud ( $L_T$ ), 250  $\mu\text{m}$  (id) y 365  $\mu\text{m}$  (od), relleno de sorbente. En este caso el sorbente empleado fueron partículas de  $\text{TiO}_2$  obtenidas de un kit comercial diseñado para la purificación de fosfopéptidos (*Pierce Magnetic  $\text{TiO}_2$  Phosphopeptide Enrichment Kit*, Thermo Scientific). Para evitar la pérdida de las partículas empaquetadas, se colocan unas fritas de algodón en

cada extremo del microcartucho. El proceso de construcción del microcartucho, supervisado a través de una lupa binocular, se describe brevemente a continuación (**Figura 5.5**).

- El capilar de separación, previamente activado mediante lavado con NaOH y agua, se corta en dos fragmentos, el del capilar de entrada de 7,5 cm, y el fragmento de salida de 64,5 cm. El corte del capilar se realiza con un cortador de capilar específico para que sea lo más limpio posible, y así minimizar los volúmenes muertos y pérdidas a través de las juntas.
- Se introduce una frita de algodón en uno de los extremos del capilar de 0,7 cm que será el cuerpo del microcartucho (**Figura 5.5 A**) y se une al fragmento de entrada del capilar de separación con un tubo Tygon™ de diámetro adecuado (**Figura 5.5 B**).
- Se aspiran las partículas de fase estacionaria hasta el interior del microcartucho mediante la aplicación de vacío (**Figura 5.5 B-C**).
- Se introduce una frita de algodón en el otro extremo del microcartucho (**Figura 5.5 D**) y se une al fragmento de salida del capilar de separación mediante otro tubo Tygon™ (**Figura 5.5 E**).
- Por último, antes de comenzar los análisis, se comprueba que el microcartucho empaquetado no ofrece demasiada restricción al flujo ni inestabilidad de la corriente eléctrica.



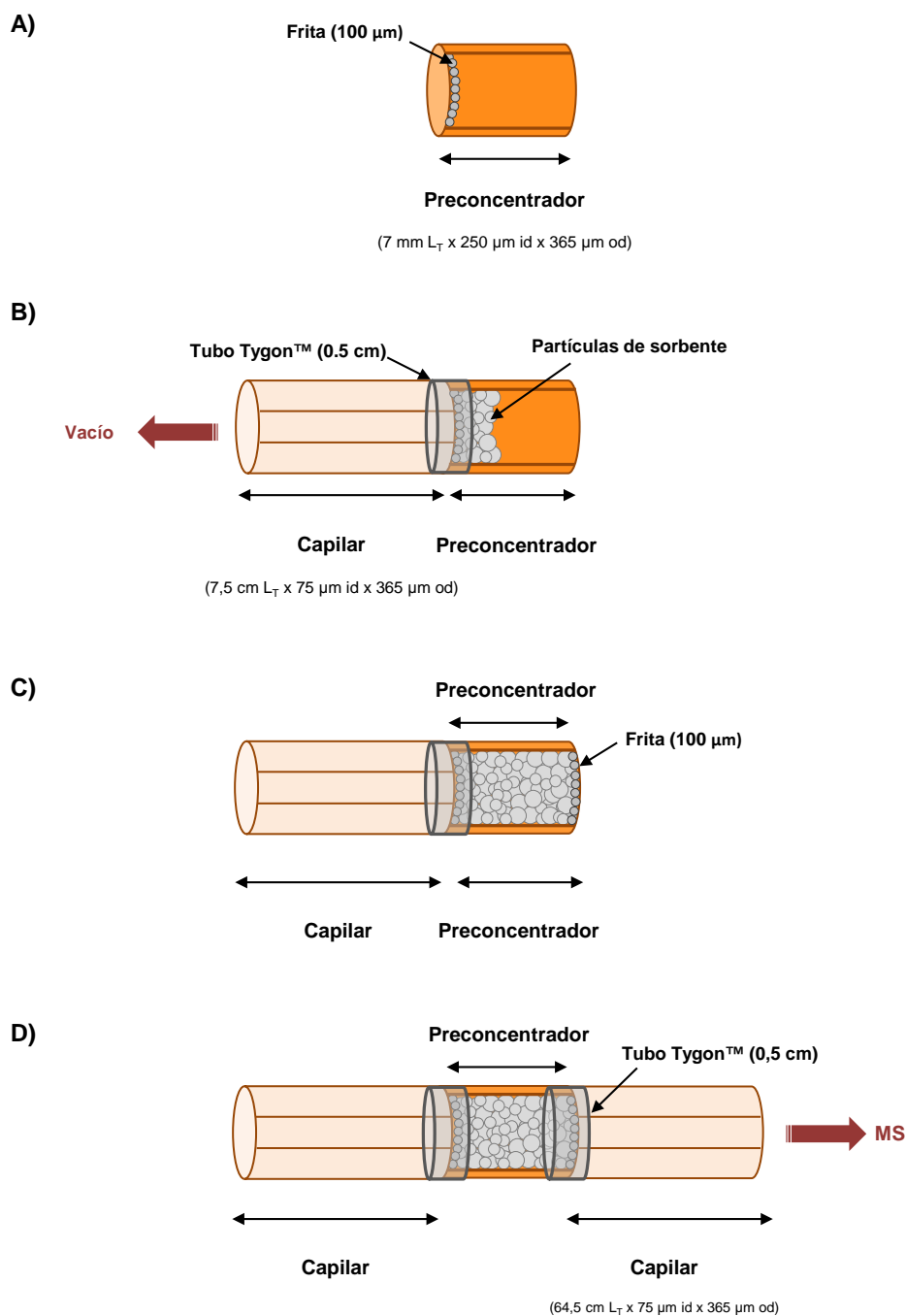


Figura 5.5.- Proceso de construcción del microcartucho para SPE-CE-MS.

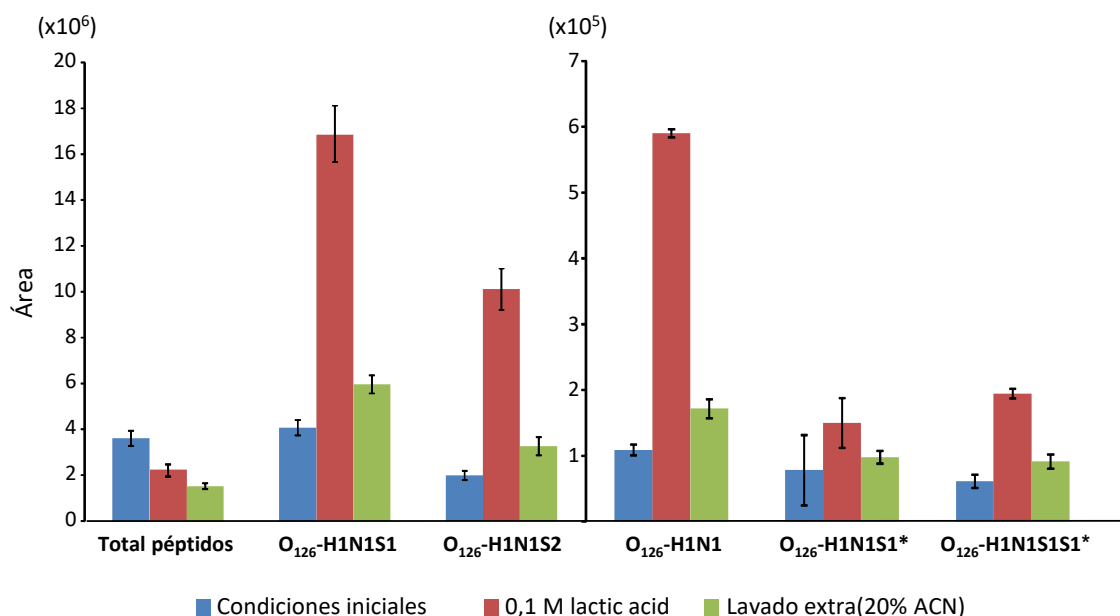
Una vez preparada la columna de  $\text{TiO}_2$ -SPE-CE-MS, se debe desarrollar y optimizar la metodología analítica, lo que constituye todo un reto, puesto que es necesario hacer compatibles las condiciones de la extracción con partículas  $\text{TiO}_2$  con las necesarias para conseguir una adecuada separación electroforética y una buena detección por MS.

### 5.1.2.2. Optimización de la purificación del O-glicopéptido de la rhEPO por TiO<sub>2</sub>-SPE-CE-MS

Las diferentes disoluciones suministradas con el kit comercial de TiO<sub>2</sub> eran de composición desconocida y resultaron incompatibles con el acoplamiento por MS al llevar a cabo algunos experimentos preliminares. Por ello, se establecieron unas condiciones iniciales a partir de las descritas por otros autores para el enriquecimiento *off-line* de fosfopéptidos y glicopéptidos con TiO<sub>2</sub> [115–117,207]. Dichas condiciones iniciales se detallan a continuación:

- Acondicionamiento de los microcartuchos con tampón de unión compuesto por 80% v/v de acetonitrilo (ACN) y 10% v/v de ácido fórmico (HFor).
- Carga de la muestra a 930 mbar durante 5 min en tampón de carga compuesto por 80% v/v de ACN y 2% v/v de HFor.
- Lavados con tampón de unión (1 min) y tampón de carga (1 min) seguido del llenado del capilar con BGE (50 mM ácido acético (HAc) y 50 mM HFor, pH 2,2) durante 2 min.
- Elución con 1 M de hidróxido de amonio (NH<sub>4</sub>OH).

El gráfico de barras de la **Figura 5.6** muestra el área total de los péptidos, así como de las diferentes glicofomas del O-glicopéptido de la rhEPO (O<sub>126</sub>) detectados en un digesto de 50 mg·L<sup>-1</sup> de rhEPO por TiO<sub>2</sub>-SPE-CE-MS empleando diferentes condiciones.



**Figura 5.6.-** Gráfico de barras mostrando el efecto en las áreas de las glicofórmulas del glicopéptido  $O_{126}$  y del total de péptidos de la rhEPO, detectados por  $TiO_2$ -SPE-CE-MS, al añadir 0,1 M de ácido láctico a los tampones de unión y carga así como una etapa de lavado extra con 20% de acetonitrilo (ACN). (\*) Uno (H1N1S1S1) o todos (H1N1S1) los ácidos siálicos son NeuGc en lugar de NeuAc.

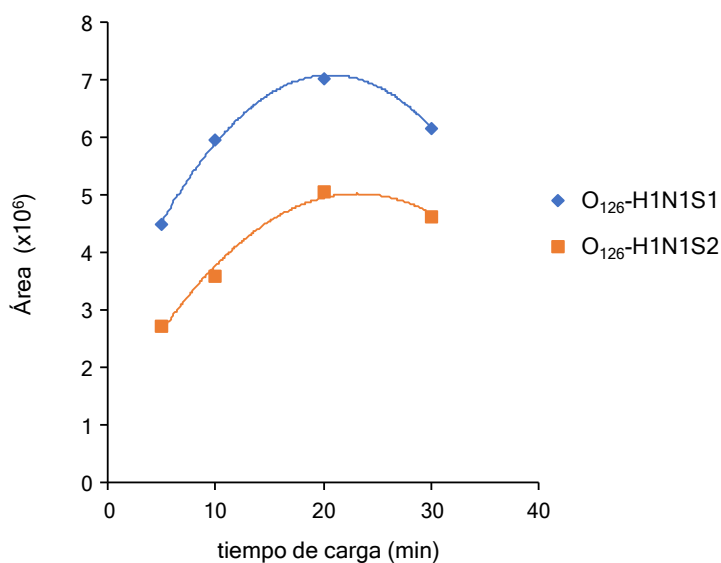
Usando las condiciones iniciales, puede observarse la existencia de retención inespecífica de los péptidos en el sorbente de  $TiO_2$ , aunque menor que la retención selectiva de las glicofórmulas del  $O_{126}$ . Algunos autores han descrito el uso de ácido láctico o glicólico en la purificación de fosfopéptidos [116,207] para mejorar la selectividad y reducir las uniones no específicas, por lo que se evaluó el uso de ácido láctico a una concentración 0,1 M en los tampones de elución y de carga. Usando estas condiciones, la retención de los péptidos disminuyó mientras que la intensidad de las glicofórmulas del  $O_{126}$  aumentó significativamente (ver **Figura 5.6**). Se descartó el uso de concentraciones superiores de ácido láctico ya que producían mucho ruido de fondo e inestabilidades de corriente durante la separación electroforética. Asimismo, con el propósito de reducir la retención inespecífica en el sorbente de los péptidos más hidrofílicos, se añadió una etapa de lavado extra con un porcentaje menor de ACN (20% v/v ACN con 2% v/v de HFor, 1 min) [115,116]. Sin embargo, mientras que el área total de los péptidos experimentó una reducción escasa (**Figura 5.6**), el área de las glicofórmulas

del O<sub>126</sub> disminuyó drásticamente. Por este motivo, se descartó esta etapa de lavado extra, y se emplearon los tampones de unión y carga con 0,1 M de ácido láctico para el resto de los experimentos. Cabe destacar que, a pesar de que las partículas de TiO<sub>2</sub> han sido descritas para llevar a cabo la retención únicamente de glicopéptidos sialilados, la glicoforma O<sub>126</sub>-H1N1 también se detectó usando estas condiciones (**Figura 5.6**), tal y como ya se había descrito para otros glicopéptidos no sialilados [117].

Seguidamente, se procedió a la optimización de la composición del eluyente partiendo de la disolución 1 M de NH<sub>4</sub>OH inicial. Algunos autores han propuesto el uso de concentraciones superiores de NH<sub>4</sub>OH para eluir glicopéptidos altamente sialilados en protocolos *off-line* [115,211], por lo que se analizó un digesto de 10 mg·L<sup>-1</sup> de rhEPO empleando como eluyente disoluciones de concentración 1 M, 3 M y 6 M de NH<sub>4</sub>OH. No obstante, se observó una disminución de las áreas de las glicoformas al aumentar la concentración de NH<sub>4</sub>OH, especialmente la glicoforma con más contenido de ácidos siálicos (O<sub>126</sub>-H1N1S2). También se probó una disolución de fosfato de amonio (pH 9) como eluyente esperando obtener mejores recuperaciones de los glicopéptidos dada la gran afinidad de los grupos fosfato por el sorbente de TiO<sub>2</sub>. Sin embargo, la señal de todas las glicoformas también disminuyó, llegando a no detectarse la glicoforma no sialilada (O<sub>126</sub>-H1N1). Finalmente, se evaluó el uso de un eluyente ácido (0,1% v/v de HFor) [117], pero los resultados obtenidos fueron peores que en medio básico. De esta manera, se confirmó que el uso de una disolución de 1 M de NH<sub>4</sub>OH era la más adecuada para el análisis de glicopéptidos por TiO<sub>2</sub>-SPE-CE-MS.

A continuación, se investigó el tiempo de carga de muestra, usando las condiciones óptimas establecidas, inyectando un digesto de 10 mg·L<sup>-1</sup> de rhEPO durante 5, 10, 20 y 30 min a 930 mbar. Tal y como se observa en el gráfico de la **Figura 5.7** para las glicoformas O<sub>126</sub>-H1N1S1 y O<sub>126</sub>-H1N1S2, las áreas de pico aumentaron progresivamente hasta alcanzar un máximo a los 20 min de introducción de muestra. No obstante, en estas condiciones se produjeron

inestabilidades en la corriente que afectaron a la repetitividad de la separación, por lo que se seleccionó un tiempo de carga de 10 min como óptimo.



**Figura 5.7.-** Evaluación del tiempo de carga de muestra obtenido para las glicofomas O<sub>126</sub>-H1N1S1 y O<sub>126</sub>-H1N1S2 de la rhEPO mediante TiO<sub>2</sub>-SPE-CE-MS (50 mg·L<sup>-1</sup> rhEPO).

### 5.1.2.3. Análisis de los glicopéptidos de la rhEPO por TiO<sub>2</sub>-SPE-CE-MS

Una vez optimizada la metodología de TiO<sub>2</sub>-SPE-CE-MS, ésta se evaluó para el análisis de todas las glicofomas de los glicopéptidos O<sub>126</sub> y N<sub>83</sub> de la rhEPO y se determinaron sus parámetros de calidad. En primer lugar, se estudió la repetitividad mediante análisis consecutivos de un digesto de 10 mg·L<sup>-1</sup> de rhEPO, obteniéndose desviaciones estándar relativas (n = 3) entre 9-11% y 6-11% para los tiempos de migración y las áreas de pico, respectivamente, para todas las glicofomas del O<sub>126</sub>. La vida útil del microcartucho se estableció en aproximadamente 10 análisis. Con respecto a la linealidad, el método fue lineal ( $R^2 > 0,99$ ) entre 0,5 y 50 mg·L<sup>-1</sup> de rhEPO para las glicofomas con ácido N-acetilneuramínico (NeuAc) y entre 10 y 50 mg·L<sup>-1</sup> de rhEPO para las glicofomas del O<sub>126</sub> con ácido N-glicolilneuramínico (NeuGc). La **Tabla 5.2** muestra las glicofomas detectadas por CE-MS y TiO<sub>2</sub>-SPE-CE-MS para los glicopéptidos O<sub>126</sub> i N<sub>83</sub>.

**Tabla 5.2.-** Péptidos y glicofomas de los glicopéptidos O<sub>126</sub> y N<sub>83</sub> detectadas mediante CE-MS y TiO<sub>2</sub>-SPE-CE-MS en el digesto de rhEPO con las correspondientes áreas obtenidas (n=3).

|                                   | Secuencia                        | M <sub>teo</sub> | CE-MS<br>(1000 mg·L <sup>-1</sup> ) |                                 | TiO <sub>2</sub> -SPE-CE-MS<br>(100 mg·L <sup>-1</sup> ) |
|-----------------------------------|----------------------------------|------------------|-------------------------------------|---------------------------------|--|
|                                   |                                  |                  | Error <sup>a</sup><br>(ppm)         | Área<br>(x10 <sup>6</sup> a.u.) | Área<br>(x10 <sup>6</sup> a.u.)                          |
| <b>Péptidos</b>                   | APPR                             | 439,2543         | 3,2                                 | 1,0                             | 0,01   |
|                                   | LICDSR-(Cys-IAA <sup>b</sup> )   | 762,3694         | 1,1                                 | 2,8                             | 0,3  |
|                                   | VLER                             | 515,3067         | 3,4                                 | 8,9                             | 0,05   |
|                                   | YLLEAK                           | 735,4167         | 4,2                                 | 8,3                             | 0,07   |
|                                   | VNFYAWK                          | 926,4650         | 1,2                                 | 4,1                             | 0,009  |
|                                   | MEVGQQAVEVWQGLALLSEAVLR          | 2525,3312        | 5,3                                 | 0,3                             | 0,02   |
|                                   | AVSGLR                           | 601,3548         | 1,9                                 | 4,6                             | 0,01   |
|                                   | SLTLLR                           | 802,4913         | 7,3                                 | 19,2                            | 0,01   |
|                                   | ALGAQK                           | 586,3439         | 1,7                                 | 3,6                             | 0,1  |
|                                   | TITADTFR                         | 923,4712         | 2,4                                 | 5,4                             | 0,03   |
|                                   | VYSNFLR                          | 897,4709         | 1,4                                 | 3,8                             | 0,2  |
|                                   | LYTGEACR-(Cys-IAA <sup>b</sup> ) | 968,4386         | 1,1                                 | 3,7                             | 0,2  |
|                                   | TGDR                             | 447,2077         | 5,8                                 | 0,05                            | 0,05   |
| <b>Glicofomas O<sub>126</sub></b> | -H1N1                            | 1829,8895        | 5,4                                 | 0,5                             | 0,5  |
|                                   | -H1N1S1                          | 2120,9849        | 4,5                                 | 13,7                            | 10,7   |
|                                   | -H1N1S2                          | 2412,0803        | 4,9                                 | 4,8                             | 6,2  |
|                                   | -H1N1S1 <sup>c</sup>             | 2136,9798        | 5,7                                 | 0,2                             | 0,1  |
|                                   | -H1N1S1S1 <sup>c</sup>           | 2428,0752        | 3,0                                 | 0,1                             | 0,1  |
| <b>Glicofomas N<sub>83</sub></b>  | -H6N5S2F1                        | 5074,1962        | 9,1                                 | 0,06                            | 0,1  |
|                                   | -H6N5S3F1                        | 5365,2919        | 5,7                                 | 0,2                             | 0,3  |
|                                   | -H7N6S2F1                        | 5439,3279        | 5,1                                 | 0,08                            | 0,09   |
|                                   | -H7N6S3F1                        | 5730,4234        | 3,7                                 | 0,6                             | 0,5  |
|                                   | -H7N6S4F1                        | 6021,5188        | 1,9                                 | 1,5                             | 1,1  |

|           |           |     |      |      |
|-----------|-----------|-----|------|------|
| -H8N7S2F1 | 5804,4602 | 1,1 | 0,09 | 0,09 |
| -H8N7S3F1 | 6095,5556 | 2,2 | 0,8  | 0,5  |
| -H8N7S4F1 | 6386,6510 | 3,5 | 1,9  | 1,4  |
| -H9N8S3F1 | 6460,6878 | 6,5 | 0,3  | 0,3  |
| -H9N8S4F1 | 6751,7832 | 4,6 | 0,7  | 0,5  |

Los péptidos detectados con menos de 4 amino ácidos no han sido considerados (ie. R, K, LFR, GK,LK).

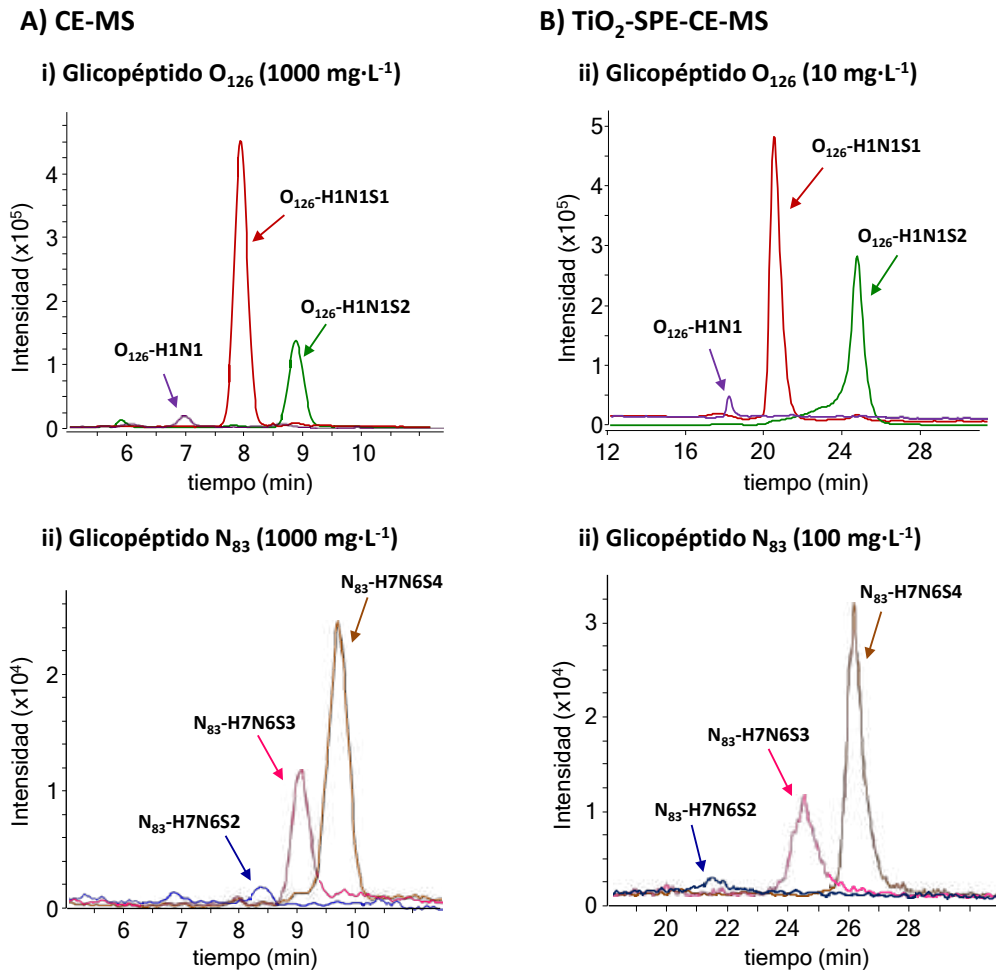
<sup>a</sup> Error calculado en ppm siguiendo la ecuación:  $(|M_{exp} - M_{teo}| / M_{teo}) \times 10^6$  (exp = experimental and teo = teórica).

<sup>b</sup> IAA se refiere a iodoacetamida.

<sup>c</sup> Un (H1N1S1S1) o todos (H1N1S1) los ácidos siálicos son NeuGc en lugar de NeuAc.

Mientras que el área de los péptidos disminuye drásticamente a una concentración de digesto 10 veces menor, en general, el área de los glicopéptidos se mantiene o aumenta. En el caso del  $O_{126}$ , el límite de detección (LOD) en  $TiO_2$ -SPE-CE-MS se estableció en  $0,25 \text{ mg}\cdot\text{L}^{-1}$  de rhEPO para detectar todas la glicofomas con NeuAc y en  $10 \text{ mg}\cdot\text{L}^{-1}$  de rhEPO para aquellas glicofomas con NeuGc. Estos resultados suponen un aumento de sensibilidad de 100 veces respecto a CE-MS, donde se establecieron LODs de 25 y  $100 \text{ mg}\cdot\text{L}^{-1}$  de rhEPO, respectivamente. La disminución de los LODs tiene especial interés en el caso de las glicofomas con NeuGc de la rhEPO dado que son particularmente importantes en el control de calidad de este biofármaco y en el control antidopaje. Asimismo, tal y como se muestra en la **Tabla 5.2**, el método por  $TiO_2$ -SPE-CE-MS permitió también la detección de todas las glicofomas del  $N_{83}$  detectadas por CE-MS, obteniendo un aumento de sensibilidad de 10 veces, sin ninguna reoptimización, a pesar de que dicha metodología fue desarrollada y optimizada para el glicopéptido  $O_{126}$ . La disminución de los LODs no fue tan notable como en el caso del glicopéptido  $O_{126}$ , dado que las glicofomas del  $N_{83}$  son más difíciles de ionizar en modo positivo al ser, en general, más grandes y presentar una carga negativa superior. A modo de ejemplo, en la **Figura 5.8** se muestran los EIEs de las glicofomas mayoritarias del  $O_{126}$  con NeuAc y las glicofomas tetraantenarias del  $N_{83}$  detectadas por CE-MS y  $TiO_2$ -SPE-CE-MS. Como puede observarse, en ningún caso se vio comprometida la separación entre las glicofomas con diferente número de

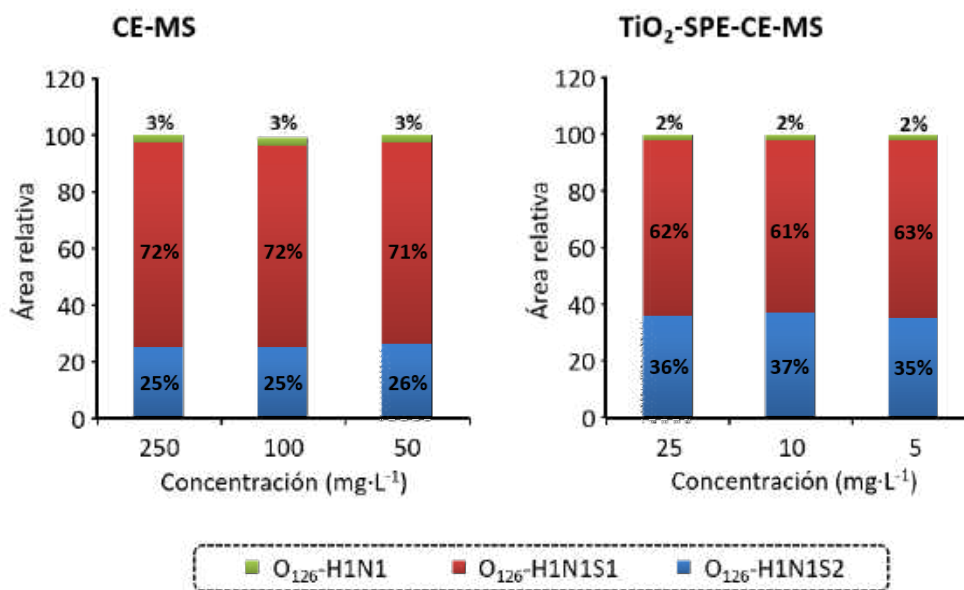
ácidos siálicos, a pesar de aplicar presión durante la separación electroforética por TiO<sub>2</sub>-SPE-CE-MS.



**Figura 5.8.-** Análisis del digesto de rhEPO mediante (A) CE-MS y (B) TiO<sub>2</sub>-SPE-CE-MS. EIEs de las glicofomas más abundantes (i) del glicopéptido O<sub>126</sub> y (ii) de las glicofomas más relevantes del glicopéptido N<sub>83</sub> de la rhEPO.

Finalmente, en este estudio se evaluó la capacidad del método TiO<sub>2</sub>-SPE-CE-MS para proporcionar perfiles inalterados de las glicofomas de los glicopéptidos de la rhEPO, investigando la posible interacción preferencial del sorbente TiO<sub>2</sub> hacia aquellas glicofomas más sialiladas. Para ello, se analizaron digestos de rhEPO a diferentes concentraciones por CE-MS y TiO<sub>2</sub>-SPE-CE-MS y se compararon las áreas relativas de las glicofomas mayoritarias, tal y como se muestra en la **Figura 5.9**.



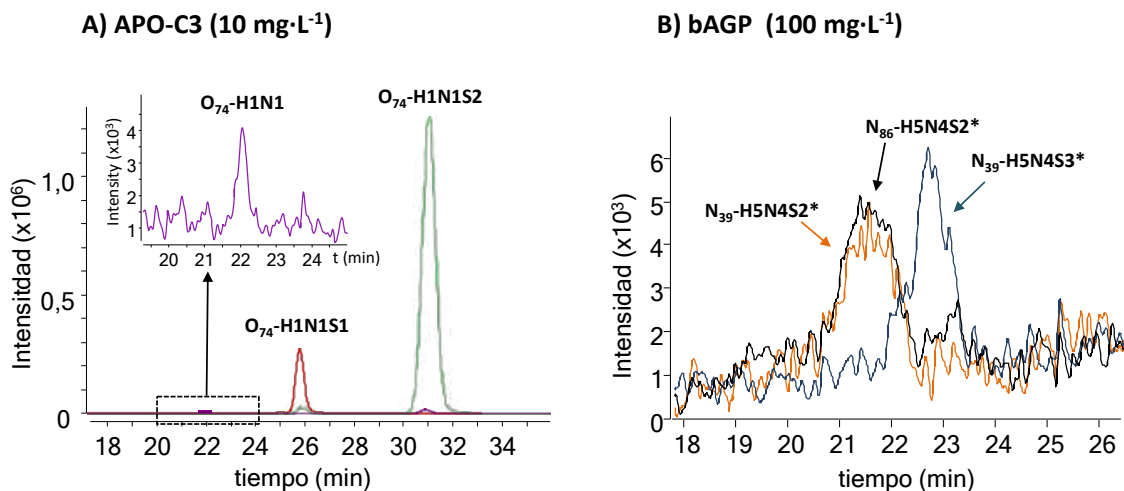


**Figura 5.9.-** Gráfico de barras mostrando las áreas relativas de las glicofomas más abundantes del glicopéptido O<sub>126</sub> a diferentes concentraciones del digesto de rhEPO obtenidas mediante CE-MS y TiO<sub>2</sub>-SPE-CE-MS. El área relativa se calculó como el área de pico de cada glicofoma dividida entre la suma de las áreas de pico de todas las glicofomas detectadas del O<sub>126</sub>.

Como puede observarse, el porcentaje de área relativa obtenido para las glicofomas con más ácidos siálicos (O<sub>126</sub>-H1N1S2) fué un poco superior utilizando TiO<sub>2</sub>-SPE-CE-MS (~36%) respecto a CE-MS (~25%), y a la inversa para las glicofomas menos sialiladas, demostrando que el sorbente de TiO<sub>2</sub> presenta una cierta preferencia por aquellas glicofomas con más ácidos siálicos. No obstante, en el caso del glicopéptido N<sub>83</sub> no se observaron diferencias significativas en este sentido entre CE-MS y TiO<sub>2</sub>-SPE-CE-MS. El hecho de que la selectividad del sorbente hacia las estructuras más sialiladas se produzca sólo en el caso del O<sub>126</sub>, podría ser atribuida a que éste, a diferencia del N<sub>83</sub>, presenta glicanos muy pequeños principalmente compuestos por ácidos siálicos. De este modo, estos ácidos siálicos determinan la interacción del glicopéptido O<sub>126</sub> con las partículas de TiO<sub>2</sub>.

### 5.1.2.4. Análisis de los glicopéptidos de otras glicoproteínas por TiO<sub>2</sub>-SPE-CE-MS

Con el fin de evaluar el potencial del método desarrollado para la preconcentración de glicopéptidos mediante TiO<sub>2</sub>-SPE-CE-MS, se analizaron también los digestos trípticos de la APO-C3 y la bAGP por CE-MS y TiO<sub>2</sub>-SPE-CE-MS. Para estas glicoproteínas los resultados fueron similares a los obtenidos con la rhEPO, detectándose por TiO<sub>2</sub>-SPE-CE-MS todas las glicofomas de los glicopéptidos previamente detectadas por CE-MS. Los factores de preconcentración oscilaron entre 100 veces para el O-glicopéptido de la APO-C3 a 10 veces para los N-glicopéptidos de la bAGP. La **Figura 5.10** muestra, a modo de ejemplo, los EIEs de las glicofomas del O<sub>74</sub> de la APO-C3 y de algunas glicofomas de los glicopéptidos N<sub>39</sub> y N<sub>86</sub> obtenidos por TiO<sub>2</sub>-SPE-CE-MS.



**Figura 5.10.-** EIEs de las glicofomas del glicopéptido O<sub>74</sub> de la APO-C3 (10 mg·L<sup>-1</sup>) y de las glicofomas N<sub>39</sub>-H5N4S2, N<sub>39</sub>-H5N4S3 y N<sub>86</sub>-H5N4S2 de las bAGP (100 mg·L<sup>-1</sup>) detectadas mediante TiO<sub>2</sub>-SPE-CE-MS. (\*) Todos los ácidos siálicos son NeuGc en lugar de NeuAc.

### 5.1.3. Comparación de los métodos de purificación de glicopéptidos

En esta tesis doctoral, se han establecido dos métodos para la purificación de glicopéptidos de digestos enzimáticos. Ambas metodologías permiten separar los glicopéptidos, presentes en un digesto, de los péptidos que podrían suprimir su ionización por MS. Sin embargo, el método

de TiO<sub>2</sub>-SPE-CE-MS resulta más eficaz para llevar a cabo el *clean-up* del digesto. En este sentido, si nos centramos en el digesto de la rHEPO, se puede observar que el área de todos los péptidos disminuye sustancialmente empleando TiO<sub>2</sub>-SPE-CE-MS (**Tabla 5.2**) mientras que, en el método de precipitación con acetona, los péptidos más hidrofílicos también precipitan en gran medida con la proporción de acetona seleccionada para la purificación de los O-glicopéptidos (**Figura 5.3B**). Por otro lado, aunque con ambos métodos se detectan las glicofomas no sialiladas de los glicopéptidos estudiados, el método TiO<sub>2</sub>-SPE-CE-MS permite detectar un número superior de glicofomas. Estos resultados ponen de manifiesto que, a pesar de que el método por TiO<sub>2</sub>-SPE-CE-MS es más laborioso, éste proporciona una purificación mucho más selectiva de los glicopéptidos y una disminución de los LODs, obteniendo factores de preconcentración de hasta 100 veces en comparación con CE-MS. No obstante, la selección del método de purificación debe ser realizada en función del propósito del estudio y de la concentración de glicopéptidos en la muestra. De este modo, en los casos en que la concentración de glicopéptidos no sea limitante y se precise de un método rápido para su purificación, se emplearía la precipitación con acetona. Por el contrario, cuando se quieran detectar glicopéptidos poco abundantes en el digesto sería más adecuado emplear el método por TiO<sub>2</sub>-SPE-CE-MS.

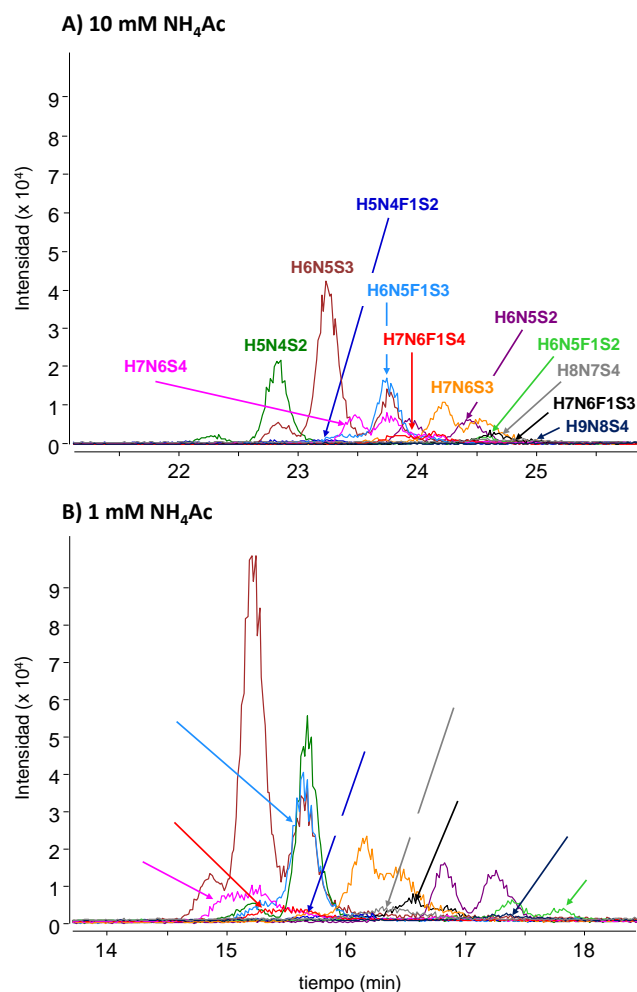
## 5.2. Separación y caracterización de glicanos y sus correspondientes isómeros

En los últimos años, el estudio de los glicanos y sus correspondientes isómeros ha suscitado gran interés en biomedicina debido a que ciertos tipos de enlace o determinadas estructuras de glicano están relacionadas con la aparición, el desarrollo y la progresión de algunas patologías importantes como el cáncer, por lo que podrían actuar como biomarcadores para su diagnóstico y seguimiento [32,35,95,212]. En este sentido, la separación de los diferentes isómeros de un glicano y su caracterización exhaustiva representa un reto analítico, enfocado a la detección de alteraciones en isómeros que se diferencian en el tipo de enlace de los ácidos siálicos o de las fucosas, en lugar del simple análisis de los niveles totales de sialilación o fucosilación de los glicanos.

Por este motivo, en esta tesis doctoral se ha llevado a cabo el desarrollo de una metodología analítica para separar e identificar glicanos y sus correspondientes isómeros mediante cromatografía de líquidos capilar de interacción hidrofílica zwitteriónica acoplada a la espectrometría de masas (CapZIC-HILIC-MS) (**artículo 3.1**). Además, se han desarrollado diversos métodos para caracterizar las estructuras de los glicanos y los tipos de enlace de los ácidos siálicos y fucosas que estos presentan. Se han utilizado tres estrategias de análisis diferentes: la digestión con enzimas específicas (exoglicosidasas) previa al análisis por CapZIC-HILIC-MS (**artículo 3.1**), la detección utilizando espectrometría de masas en tándem (**artículo 3.2**), y la espectrometría de masas de movilidad iónica. Para todos estos estudios se ha utilizado como glicoproteína modelo la alfa-1-glicoproteína ácida humana (hAGP) dada su relación con algunas enfermedades inflamatorias y ciertos tipos de cáncer [46,49,53], además del amplio abanico de glicanos que presenta, desde estructuras sialofucosiladas biantenarias hasta tetraantenarias.

### 5.2.1. Optimización de las condiciones cromatográficas por CapZIC-HILIC-MS

En un trabajo previo de nuestro grupo de investigación se desarrolló un método por CapZIC-HILIC-MS, en modo negativo, para el análisis de los glicanos de la hAGP derivatizados con [ $^{12}\text{C}_6$ ]-anilina, utilizando una fase móvil ACN:10mM acetato de amonio ( $\text{NH}_4\text{Ac}$ ), pH 6,4 [54]. Este método permitía separar los diferentes isómeros de un glicano, pero no presentaba suficiente sensibilidad para detectar glicanos minoritarios o con porcentajes de glicosilación pequeños. Por esta razón, en esta tesis doctoral se ha llevado a cabo una reoptimización de la metodología, evaluando la influencia del pH y de la fuerza iónica de la fase móvil, para mejorar la detección y la separación de los glicanos y sus isómeros. En relación al pH, se evaluaron diferentes condiciones (pH 3, 5, 7 y 8) pero no se consiguió una mejora en la separación de los glicanos y además, la intensidad de su señal disminuyó considerablemente. El efecto de la fuerza iónica también fue evaluado utilizando fases móviles acuosas de concentración 10, 5 y 1 mM de  $\text{NH}_4\text{Ac}$ . En la **Figura 5.11** se muestran los cromatogramas de iones extraídos (EICs) de algunos de los *N*-glicanos detectados derivatizados con anilina, utilizando como fase móvil acuosa (A) 10 mM  $\text{NH}_4\text{Ac}$  y (B) 1 mM  $\text{NH}_4\text{Ac}$ . Como se puede observar, aunque el orden de elución se ve alterado en algunos casos (por ejemplo, en el H5N4S2), no se observan cambios significativos en la separación de los glicanos. En cambio, la intensidad de la señal aumenta considerablemente al disminuir la concentración de sal, obteniéndose los mejores resultados con la fase móvil acuosa de 1 mM  $\text{NH}_4\text{Ac}$  (**Figura 5.11B**). De este modo, se estableció como fase móvil óptima ACN:1 mM  $\text{NH}_4\text{Ac}$ , pH 6,4. Finalmente, empleando dicha fase móvil, se optimizó el gradiente para obtener la mejor separación posible entre isómeros.



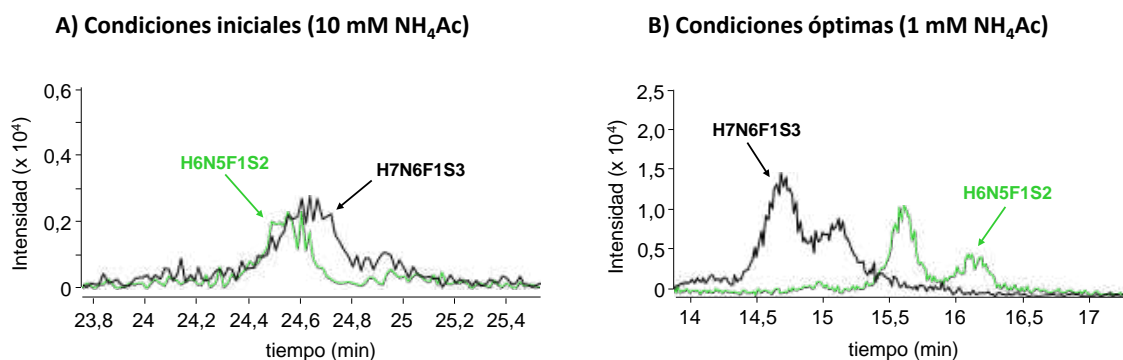
**Figura 5.11.-** EICs de los N-glicanos más abundantes obtenidos de una muestra patrón de hAGP y derivatizados con [<sup>12</sup>C<sub>6</sub>]AN utilizando el gradiente cromatográfico descrito en [31] y una fase móvil acuosa de 10 mM de NH<sub>4</sub>Ac (A) y 1 mM de NH<sub>4</sub>Ac (B).

Utilizando estas condiciones cromatográficas, se detectaron, en una muestra de hAGP patrón, los glicanos fucosilados y no fucosilados con estructuras ramificadas bi-, tri- y tetraantenarias previamente descritos en los trabajos de nuestro grupo de investigación [54]. A su vez, gracias al aumento de sensibilidad obtenido con la optimización de la fase móvil, se detectaron estructuras minoritarias menos sialiladas como el H5N4S1, H7N6S1 y H7N6S2, así como glicanos tetraantenarios con una y dos unidades extra de N-acetillactosamina (LacNAc), algunos de ellos también fucosilados. A modo de resumen, todos los glicanos de la hAGP detectados se muestran en la **Tabla 5.3**.

**Tabla 5.3.** N-glicanos de la hAGP derivatizados con [<sup>12</sup>C<sub>6</sub>]AN y detectados mediante CapZIC-HILIC-MS en modo negativo (concentración de los glicanos: 50 pmol/μL).

| Glicano-[ <sup>12</sup> C <sub>6</sub> ]AN | t <sub>R</sub> (min) | M <sub>teo</sub> | M <sub>exp</sub> | Error (ppm) |
|--|----------------------|------------------|------------------|-------------|
| H5N4S1                                     | Isómero 1 17,3       | 2008,7267        | 2008,7298        | 1,5         |
| H5N4S2                                     | Isómero 1 12,9       | 2299,8232        | 2299,8288        | 2,4         |
|  | Isómero 2 13,5       | 2299,8232        | 2299,8218        | 0,6         |
| H5N4S2F1                                   | Isómero 1 14,0       | 2445,8810        | 2445,8844        | 1,4         |
|  | Isómero 2 14,2       | 2445,8810        | 2445,8884        | 3,0         |
| H6N5S2                                     | Isómero 1 14,8       | 2664,9554        | 2664,9590        | 1,4         |
|  | Isómero 2 15,4       | 2664,9554        | 2664,9539        | 0,6         |
| H6N5S2F1                                   | Isómero 1 15,5       | 2811,0132        | 2810,9946        | 6,6         |
|  | Isómero 2 16,0       | 2811,0132        | 2810,9850        | 10,0        |
| H6N5S3                                     | Isómero 1 12,6       | 2956,0513        | 2956,0484        | 1,0         |
|  | Isómero 2 13,0       | 2956,0513        | 2956,0463        | 1,7         |
|  | Isómero 3 13,5       | 2956,0513        | 2956,0469        | 1,5         |
| H6N5S3F1                                   | Isómero 1 13,1       | 3102,1093        | 3102,0830        | 8,5         |
|  | Isómero 2 13,5       | 3102,1093        | 3102,1025        | 2,2         |
| H6N5S3F2                                   | Isómero 1 13,9       | 3248,1656        | 3248,1578        | 2,4         |
| H7N6S1                                     | Isómero 1 22,1       | 2738,9911        | 2738,9606        | 11,1        |
| H7N6S2                                     | Isómero 1 17,0       | 3030,0865        | 3030,0824        | 1,3         |
|  | Isómero 2 17,5       | 3030,0865        | 3030,0818        | 1,5         |
| H7N6S2F1                                   | Isómero 1 17,6       | 3176,1444        | 3176,1245        | 6,3         |
|  | Isómero 2 18,1       | 3176,1444        | 3176,1386        | 1,8         |
| H7N6S3                                     | Isómero 1 13,5       | 3321,1837        | 3321,1616        | 6,6         |
|  | Isómero 2 14,1       | 3321,1837        | 3321,1778        | 1,8         |
|  | Isómero 3 14,4       | 3321,1837        | 3321,1772        | 1,9         |
|  | Isómero 4 14,9       | 3321,1837        | 3321,1664        | 5,2         |
| H7N6S3F1                                   | Isómero 1 13,9       | 3467,2414        | 3467,2259        | 4,5         |
|  | Isómero 2 14,6       | 3467,2414        | 3467,2289        | 3,6         |
|  | Isómero 3 14,9       | 3467,2414        | 3467,2181        | 6,7         |
|  | Isómero 4 15,3       | 3467,2414        | 3467,2226        | 5,4         |
| H7N6S3F2                                   | Isómero 1 15,0       | 3613,2978        | 3613,2707        | 7,5         |
|  | Isómero 2 15,4       | 3613,2978        | 3613,2509        | 13,0        |
| H7N6S4                                     | Isómero 1 12,8       | 3612,2791        | 3612,2714        | 2,1         |
|  | Isómero 2 13,1       | 3612,2791        | 3612,2582        | 5,8         |
|  | Isómero 3 13,5       | 3612,2791        | 3612,2681        | 3,0         |
| H7N6S4F1                                   | Isómero 1 13,1       | 3758,3368        | 3758,3177        | 5,1         |
|  | Isómero 2 13,5       | 3758,3368        | 3758,3147        | 5,9         |
| H7N6S4F2                                   | Isómero 1 13,7       | 3904,3932        | 3904,3727        | 5,2         |
| H7N6S4F3                                   | Isómero 1 14,2       | 4050,4511        | 4050,3887        | 15,4        |
| H8N7S3                                     | Isómero 1 16,2       | 3686,3141        | 3686,3096        | 1,2         |
| H8N7S4                                     | Isómero 1 14,6       | 3977,4096        | 3977,4023        | 1,8         |
| H8N7S3F1                                   | Isómero 1 16,5       | 3832,3720        | 3832,3859        | 3,6         |
| H8N7S4F1                                   | Isómero 1 14,9       | 4123,4675        | 4123,4318        | 8,7         |
| H8N7S4F2                                   | Isómero 1 15,3       | 4269,5254        | 4269,5423        | 4,0         |
| H8N7S4F3                                   | Isómero 1 15,5       | 4415,5833        | 4415,5567        | 6,0         |
| H9N8S3                                     | Isómero 1 17,3       | 4051,4463        | 4051,4378        | 2,1         |
| H9N8S4                                     | Isómero 1 15,8       | 4342,5417        | 4342,5251        | 3,8         |
| H9N8S4F1                                   | Isómero 1 16,2       | 4488,5997        | 4488,6103        | 2,4         |

Además, la optimización del gradiente de elución permitió obtener una mejor resolución de los isómeros, especialmente de aquellos menos abundantes, como se aprecia en la **Figura 5.12**.



**Figura 5.12.-** EICs de los N-glicanos minoritarios H6N5F1S2 y H7N6F1S3 obtenidos de una muestra patrón de hAGP y derivatizados con [<sup>12</sup>C<sub>6</sub>]AN utilizando (A) el gradiente cromatográfico descrito en [31] y una fase móvil acuosa de 10 mM de NH<sub>4</sub>Ac y (B) el gradiente cromatográfico optimizado en esta tesis y una fase móvil acuosa de 1 mM de NH<sub>4</sub>Ac.

### 5.2.2. Caracterización de glicanos mediante digestión con exoglicosidasas

Una vez establecido el método de análisis para la separación y detección de los glicanos de la hAGP y sus correspondientes isómeros, se procedió a la caracterización de los isómeros mediante digestión con exoglicosidasas. Estas enzimas liberan los residuos terminales de los glicanos, tales como ácidos siálicos (sialidasas) o fucosas (fucosidasas), unidos mediante un tipo de enlace específico. Para poder evaluar los cambios producidos en los glicanos a causa de la digestión de la glicoproteína con una cierta exoglicosidasa, se empleó la estrategia de marcaje isotópico GRIL utilizando [<sup>12</sup>C<sub>6</sub>]-anilina ([<sup>12</sup>C<sub>6</sub>]AN) y [<sup>13</sup>C<sub>6</sub>]-anilina ([<sup>13</sup>C<sub>6</sub>]AN). Esta estrategia permitió asignar inequívocamente los diferentes tipos de enlace presentes en cada isómero mediante comparación directa de las áreas de los picos de los glicanos intactos con los correspondientes digeridos con una exoglicosidasa específica. De esta manera se evita la variabilidad experimental que puede tener lugar entre inyecciones, la cual podría afectar a los tiempos de retención así como a las áreas de los picos. En primer lugar, se evaluó la reproducibilidad del método con el fin de demostrar que dos muestras idénticas derivatizadas



con anilinas isotópicamente diferentes dan los mismos resultados. Con este propósito, siguiendo la estrategia GRIL planteada en la **Figura 1.10** de la introducción, tres muestras independientes de 25 µg de hAGP patrón se marcaron con [<sup>12</sup>C<sub>6</sub>]AN tras la liberación de los correspondientes glicanos mediante digestión con PNGasa F, y otras tres con [<sup>13</sup>C<sub>6</sub>]AN. Después de la derivatización y purificación de los glicanos, se prepararon mezclas equimolares de hAGP-glicanos-[<sup>12</sup>C<sub>6</sub>]AN y hAGP-glicanos-[<sup>13</sup>C<sub>6</sub>]AN y, las mezclas se analizaron por CapZIC-HILIC-MS por triplicado. En la **Tabla 5.4** se muestran las relaciones entre las correspondientes áreas obtenidas experimentalmente a partir del análisis de las tres mezclas 1:1 independientes, junto con sus valores de desviación estándar relativa (%RSDs), para siete glicanos de la hAGP tanto mayoritarios como minoritarios. Como puede observarse, la metodología establecida es muy fiable ya que se obtuvieron valores muy cercanos a la unidad. Además, los %RSD obtenidos fueron bajos (≤ 5%) demostrando que la metodología es muy reproducible y puede ser usada para comparar y cuantificar las áreas de los glicanos de dos muestras diferentes.

**Tabla 5.4.** Relación entre las áreas o ratio (área glicano derivatizado con [<sup>12</sup>C<sub>6</sub>]AN / área glicano derivatizado con [<sup>13</sup>C<sub>6</sub>]AN) obtenidas experimentalmente a partir del análisis de las tres mezclas 1:1 independientes, con los correspondientes valores de %RSDs, para siete glicanos de la hAGP tanto mayoritarios como minoritarios.

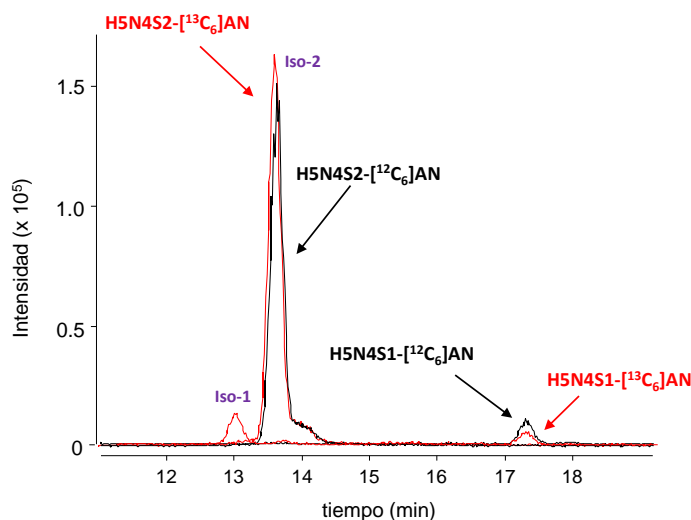
|          |           | Área glicanos-[ <sup>12</sup> C <sub>6</sub> ]AN | Área glicanos-[ <sup>13</sup> C <sub>6</sub> ]AN | Ratio Exp. (n=3) | %RSD |
|----------|-----------|--|--|------------------|------|
| H5N4S2   | Isómero 1 | 177535   | 184148   | 0,96             | 0,4  |
|          | Isómero 2 | 2176190  | 2294800  | 0,94             | 1,1  |
| H6N5S3   | Isómero 3 | 1363059  | 1403159  | 1,00             | 4,5  |
| H6N5S3F1 | Isómero 2 | 1695484  | 1884968  | 0,90             | 0,3  |
| H7N6S3F1 | Isómero 1 | 169434   | 186321   | 0,95             | 5,0  |
| H7N6S4   | Isómero 2 | 405132   | 476167   | 0,86             | 0,3  |
| H7N6F1S4 | Isómero 2 | 151613   | 172866   | 0,86             | 2,0  |

### 5.2.2.1. Caracterización de los enlaces de los ácidos siálicos

Con el objetivo de asignar el tipo de enlace de los ácidos siálicos terminales (SiA  $\alpha$ 2-3 o  $\alpha$ 2-6) de los isómeros de los glicanos de la hAGP, en esta tesis doctoral se utilizaron dos sialidasas: una sialidasa total, que permitió liberar todos los ácidos siálicos (SiA) presentes en un glicano (unidos mediante enlaces  $\alpha$ 2-3,  $\alpha$ 2-6 y  $\alpha$ 2-8), y una sialidasa específica para el enlace  $\alpha$ 2-3, en combinación con la estrategia de marcaje isotópico GRIL con anilina [ $^{12}\text{C}_6$ ]AN/[ $^{13}\text{C}_6$ ]AN. Las condiciones de digestión debieron de optimizarse para cada sialidasa, con el fin de obtener rendimientos adecuados en tiempos de digestión razonables (**artículo 3.1**).

Una vez establecidas las condiciones de digestión para las sialidasas, tres muestras independientes de 25  $\mu\text{g}$  de hAGP patrón fueron digeridas con PNGasa F y los glicanos liberados se derivatizaron con [ $^{13}\text{C}_6$ ]AN. Al mismo tiempo, otras tres muestras fueron digeridas con PNGasa F y sialidasa  $\alpha$ 2-3 y sus glicanos derivatizados con [ $^{12}\text{C}_6$ ]AN. Después de la derivatización y purificación de los glicanos, se prepararon mezclas equimolares de glicanos-[ $^{13}\text{C}_6$ ]AN y glicanos  $\alpha$ 2-3-desialilados-[ $^{12}\text{C}_6$ ]AN y se analizaron por CapZIC-HILIC-MS. En la **Figura 5.13** se muestran los resultados obtenidos para los glicanos biantenarios no fucosilados, que corresponden a la asignación más sencilla. Como se puede observar, después del tratamiento con sialidasa  $\alpha$ 2-3, el isómero 1 del H5N4S2 desaparece y el área de pico del H5N4S1 aumenta. De esto se deduce que uno de los ácidos siálicos del isómero 1 del H5N4S2 está unido mediante un enlace  $\alpha$ 2-3 y, por lo tanto, se elimina con la digestión. En cambio, el isómero 2 del H5N4S2 se mantiene en el mismo tiempo de retención y prácticamente con la misma área de pico (ver **Figura 5.13** y **Tabla 5.5**), revelando la inexistencia de SiA con uniones  $\alpha$ 2-3. En conclusión y por lo que se refiere al glicano H5N4S2, se puede afirmar que el isómero 1 contiene un SiA unido  $\alpha$ 2-6 y otro  $\alpha$ 2-3, mientras que el isómero 2 sólo presenta SiA  $\alpha$ 2-6. En relación al glicano H5N4S1, su único SiA presenta unión de tipo  $\alpha$ 2-6, dado que no se detectó

el glicano totalmente desialilado (H5N4) después del tratamiento con sialidasa  $\alpha$ 2-3 (Tabla 5.5).



**Figura 5.13.-** EICs de los N-glicanos biantenarios no fucosilados provenientes de la mezcla equimolar de los N-glicanos obtenidos de la digestión de una muestra de hAGP sólo con PNGasa F y derivatizados con  $[^{13}\text{C}_6]$ AN (EICs en rojo), y los obtenidos de la digestión de una muestra de hAGP con PNGasa F y sialidasa  $\alpha$ 2-3 y derivatizados con  $[^{12}\text{C}_6]$ AN (EICs en negro).

La asignación de los enlaces de los SiA del resto de glicanos, incluyendo los fucosilados, se llevó a cabo siguiendo la misma estrategia de comparación directa de las áreas de pico, tal y como se explica en detalle en el **artículo 3.1** de esta tesis doctoral. La asignación en el caso de los glicanos tetraantenarios fue más compleja dadas las múltiples combinaciones posibles de los enlaces de los SiA. Aun así, se llevó a cabo una caracterización tentativa de algunos isómeros de estos glicanos (**Tabla 5.5, artículo 3.1**). Para los glicanos menos abundantes de la hAGP, tales como H7N6S4F3, H9N8S4, H8N7S4F1 y H9N8S4F1, no se pudo llevar a cabo la asignación de los diferentes enlaces debido a la baja intensidad de señal obtenida por MS.

**Tabla 5.5.** Tiempos de retención y áreas obtenidas para los glicanos no fucosilados de la hAGP nativos (derivatizados con [<sup>13</sup>C<sub>6</sub>]AN) y desialilados con sialidasa α2-3 (derivatizados con [<sup>12</sup>C<sub>6</sub>]AN) mediante CapZIC-HILIC-MS. Asignación del enlace de los ácidos siálicos (SiA) para cada isómero.

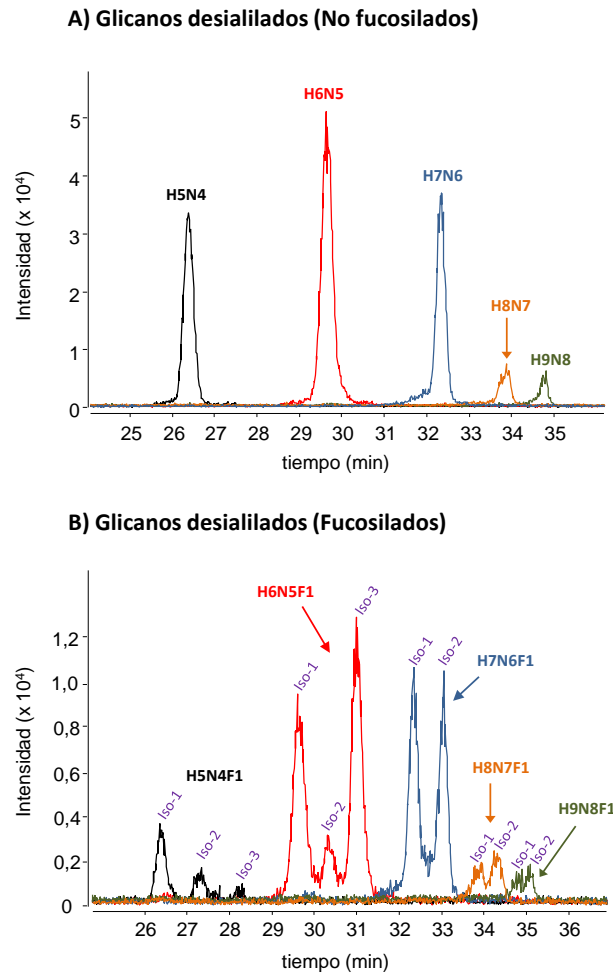
| Glicano       |                            | t <sub>R</sub><br>(min) | Área del glicano-<br>[ <sup>13</sup> C <sub>6</sub> ]AN (n=3) | Área del glicano-<br>[ <sup>12</sup> C <sub>6</sub> ]AN α2-3<br>desialilado (n=3) | Asignación enlace SiA                         |   |
|---------------|----------------------------|-------------------------|---|---|---|---|
| <b>H5N4S2</b> | Isómero 1                  | 13,6                    | 169825  | -   | 1 SiA α2-6<br>1 SiA α2-3                      |   |
|               | Isómero 2                  | 14,0                    | 2010769   | 1783084   | 2 SiA α2-6                                    |   |
| <b>H5N4S1</b> | Isómero 1                  | 17,3                    | 68941   | 135571  | 1 SiA α2-6                                    |   |
| <b>H5N4*</b>  | Isómero 1                  | 26,5                    | -   | -   | -   | - |
| <b>H6N5S3</b> | Isómero 1                  | 12,6                    | 396885  | -   | 2 SiA α2-6<br>1 SiA α2-3                      |   |
|               | Isómero 2                  | 13,1                    | 4156023   | -   | 2 SiA α2-6<br>1 SiA α2-3                      |   |
|               | Isómero 3                  | 13,7                    | 1536871   | 1412485   | 3 SiA α2-6                                    |   |
| <b>H6N5S2</b> | Isómero 1                  | 14,9                    | 292825  | -   | 1 SiA α2-6<br>1 SiA α2-3                      |   |
|               | Isómero 2                  | 15,4                    | 386282  | 2277812   | 2 SiA α2-6                                    |   |
|               | Isómero 3 <sup>&amp;</sup> | 15,9                    | -   | 225692  | -   | - |
| <b>H6N5S1</b> | Isómero 1 <sup>&amp;</sup> | 19,6                    | -   | 314765  | -   | - |
| <b>H6N5*</b>  | Isómero 1                  | 29,8                    | -   | -   | -   | - |
| <b>H7N6S4</b> | Isómero 1                  | 12,9                    | 252257  | -   | 1 SiA α2-6<br>2 SiA α2-3<br>1 SiA desconocido |   |
|               | Isómero 2                  | 13,3                    | 317537  | -   | 1 SiA α2-6<br>2 SiA α2-3<br>1 SiA desconocido |   |
|               | Isómero 3                  | 13,5                    | 126903  | -   | 2 SiA α2-6<br>2 SiA α2-3                      |   |
| <b>H7N6S3</b> | Isómero 1                  | 13,6                    | 101293  | -   | 1 SiA α2-6<br>2 SiA α2-3                      |   |

|               |                            |      |        |         |   |  |
|---------------|----------------------------|------|--------|---------|---|--|
|               | Isómero 2                  | 14,2 | 883588 | -       | 1 SiA $\alpha$ 2-6<br>1 SiA $\alpha$ 2-3<br>1 SiA desconocido |  |
|               | Isómero 3                  | 14,6 | 670502 | -       | 1 SiA $\alpha$ 2-6<br>1 SiA $\alpha$ 2-3<br>1 SiA desconocido |  |
|               | Isómero 4                  | 15,1 | 214102 | 397994  | 3 SiA $\alpha$ 2-6  |  |
| <b>H7N6S2</b> | Isómero 1                  | 16,9 | 195765 | -       | 1 SiA $\alpha$ 2-6<br>1 SiA $\alpha$ 2-3                      |  |
|               | Isómero 2                  | 17,4 | 66762  | 1004143 | 2 SiA $\alpha$ 2-6  |  |
| <b>H7N6S1</b> | Isómero 1                  | 22,1 | 43075  | 648373  | 1 SiA $\alpha$ 2-6  |  |
|               | Isómero 2 <sup>&amp;</sup> | 22,7 | -      | 76451   | -   |  |
| <b>H7N6*</b>  | Isómero 1                  | 32,4 | -      | -       | -   |  |
| <b>H8N7S4</b> | Isómero 1                  | 14,6 | 425915 | -       | 1 SiA $\alpha$ 2-6<br>2 SiA $\alpha$ 2-3<br>1 SiA desconocido |  |
| <b>H8N7S3</b> | Isómero 1                  | 16,2 | 248701 | 77459   | 1 SiA $\alpha$ 2-6<br>1 SiA $\alpha$ 2-3<br>1 SiA desconocido |  |
| <b>H8N7S2</b> | Isómero 1 <sup>&amp;</sup> | 19,6 | -      | 112626  | -   |  |
| <b>H8N7S1</b> | Isómero 1 <sup>&amp;</sup> | 24,7 | -      | 32151   | -   |  |
| <b>H8N7*</b>  | Isómero 1                  | 33,9 | -      | -       | -   |  |

\* Glicano no presente en la glicoproteína nativa. Detectado sólo después de la digestión con sialidasa total.

<sup>&</sup> Isómero no presente en la glicoproteína nativa. Detectado sólo después de la digestión con sialidasa  $\alpha$ 2-3.

Por otro lado, se llevó a cabo la digestión de los glicanos de la hAGP con sialidasa total. Los EICs obtenidos para todos los glicanos desialilados no fucosilados así como para los que presentan una fucosa se muestran en la **Figura 5.14**. Como se puede observar, sólo en el caso de los glicanos fucosilados se obtuvo más de un pico cromatográfico para cada glicano desialilado, poniendo de manifiesto la presencia de isómeros que podrían diferir en el tipo de enlace de las fucosas (**Figura 5.14B**).



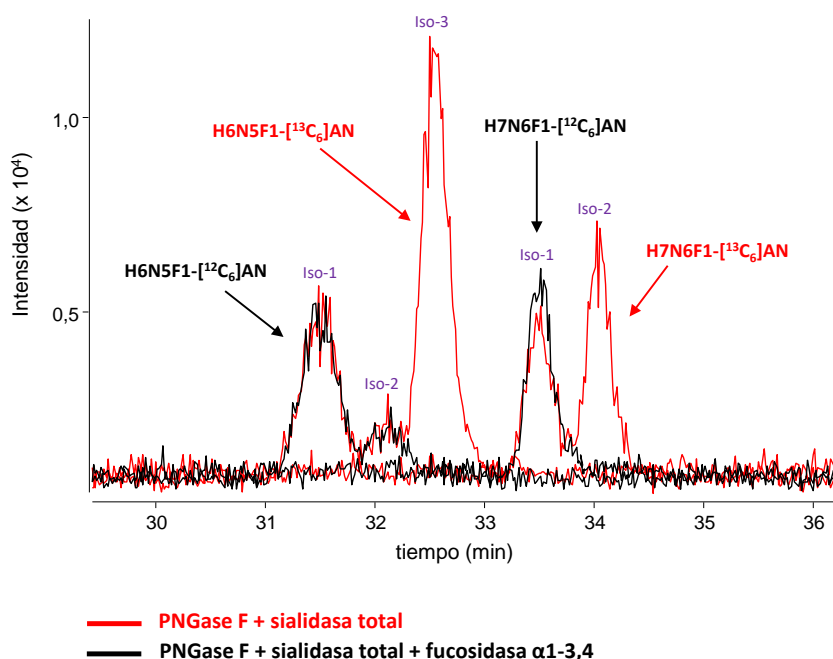
**Figura 5.14.-** EICs de algunos glicanos desialilados no fucosilados (A) y fucosilados (B) obtenidos de la digestión de una muestra patrón de hAGP con PNGasa F y sialidasa total.

**5.2.2.2. Caracterización de los enlaces de las fucosas**

Las fucosas (Fuc) pueden presentar diferentes tipos de enlace (Fuc core  $\alpha$ 1-6, Fuc antena  $\alpha$ 1-3/4 y Fuc  $\alpha$ 1-2 unida a una galactosa), tal y como se explicó en la introducción de esta tesis doctoral (**Figura 1.2**). En el caso de la hAGP, aunque únicamente se ha descrito la presencia de Fuc antena  $\alpha$ 1-3 y Fuc core  $\alpha$ 1-6 en el glicano H5N4S2F1 [55], los resultados obtenidos de la digestión con sialidasa total confirmaron la necesidad de hacer una caracterización exhaustiva de los enlaces de las fucosas en los glicanos de esta proteína. La presencia de Fuc antena  $\alpha$ 1-4 se descartó dado que la hAGP sólo presenta estructuras de *Lewis* tipo II (**Figura 1.5**). Con el fin de asignar los isómeros debidos a diferencias en el enlace de las fucosas, se utilizó como

exoglicosidasa específica una fucosidasa  $\alpha$ 1-3,4, aplicando una estrategia de marcaje isotópico GRIL con anilina [ $^{12}\text{C}_6$ ]AN/[ $^{13}\text{C}_6$ ]AN.

Primero, se evaluaron diferentes condiciones de digestión con fucosidasa  $\alpha$ 1-3,4, confirmándose la necesidad de una digestión conjunta con sialidasa total para mejorar la accesibilidad de la fucosidasa y liberar las fucosas enlazadas  $\alpha$ 1-3 de los glicanos de la hAGP. Una vez seleccionadas dichas condiciones, se llevó a cabo la asignación de los enlaces de las fucosas. Con este propósito, tres muestras independientes de 10  $\mu\text{g}$  de hAGP patrón se digirieron con PNGasa F y sialidasa total y sus glicanos se derivatizaron con [ $^{13}\text{C}_6$ ]AN, mientras que otras tres fueron digeridas con PNGasa F, sialidasa total y fucosidasa  $\alpha$ 1-3,4 y sus glicanos derivatizados con [ $^{12}\text{C}_6$ ]AN. Siguiendo la metodología GRIL, se prepararon mezclas equimolares de estas muestras y se analizaron por CapZIC-HILIC-MS. Los resultados obtenidos para los glicanos desialilados tri- y tetraantenarios con una fucosa se muestran en la **Figura 5.15**.



**Figura 5.15.-** EICs de los N-glicanos tri- y tetraantenarios con una fucosa provenientes de la mezcla equimolar de los N-glicanos obtenidos de la digestión de una muestra de hAGP con PNGasa F y sialidasa total y derivatizados con [ $^{13}\text{C}_6$ ]AN (EICs en rojo), y los obtenidos de la digestión de una muestra de hAGP con PNGasa F, sialidasa total y fucosidasa  $\alpha$ 1-3,4 y derivatizados con [ $^{12}\text{C}_6$ ]AN (EICs en negro).

Como se puede observar en la figura, en ambos casos desaparecen ciertos isómeros después del tratamiento con fucosidasa  $\alpha$ 1-3,4, indicando que sólo estos isómeros presentan enlaces de Fuc antena  $\alpha$ 1-3. En el caso de los isómeros cuyas áreas de pico permanecen intactas, las fucosas podrían estar enlazadas  $\alpha$ 1-6 al core o  $\alpha$ 1-2 a una galactosa (Gal). Para diferenciar ambas uniones, se llevó a cabo una digestión adicional con una galactosidasa  $\beta$ 1-4. La detección del mismo número de isómeros que los obtenidos tras la digestión con fucosidasa  $\alpha$ 1-3/4, confirmó que ninguno de los glicanos con una fucosa presentaba Fuc  $\alpha$ 1-2 unida a Gal. En la **Tabla 5.6** se muestra la asignación realizada para las fucosas de los glicanos desialilados fucosilados de la hAGP que se pudieron caracterizar y el procedimiento para la obtención de esta caracterización se detalla en el **artículo 3.1** de esta tesis doctoral.

Tal y como ocurrió en la asignación de los ácidos siálicos (sección 5.2.2.1), la caracterización de los enlaces de las fucosas de glicanos minoritarios fue más difícil de llevar a cabo. De este modo, la asignación de los glicanos con 1 o 2 unidades de LacNAc (H8N7F1, H8N7F2, H8N7F3 y H9N8F1) no pudo realizarse. En el caso de los glicanos con 2 o 3 fucosas, sólo se pudo demostrar que todos ellos presentaban al menos una Fuc antena  $\alpha$ 1-3, excepto el isómero 1 del H7N6F2.

Así, los resultados obtenidos demostraron que, aunque la mayoría de glicanos de la hAGP presentan Fuc antena  $\alpha$ 1-3, tanto los glicanos bi- tri- como tetraantenarios con una fucosa también presentan algún isómero con Fuc core  $\alpha$ 1-6, al contrario de lo que se había descrito para la hAGP [55]. La diferencia en la estructura de los dos isómeros asignados como Fuc core  $\alpha$ 1-6 en los glicanos H5N4F1 y H6N5F1 (**Tabla 5.6**) podría ser debida al enlace de la fucosa a una u otra de las dos N-acetilglucosaminas (GlcNAc) presentes en el core.



**Tabla 5.6.** Tiempos de retención y áreas obtenidas para los glicanos de la hAGP desialilados (derivatizados con [ $^{13}\text{C}_6$ ]AN) y los desialilados digeridos también con fucosidasa  $\alpha$ 1-3,4 (derivatizados con [ $^{12}\text{C}_6$ ]AN) mediante CapZIC-HILIC-MS. Assignación del enlace de las fucosas (Fuc) para cada isómero.

| Glicano       |           | $t_R$ (min) | Área glicano- $^{13}\text{C}_6$ ]AN desialilado | Área glicano- $^{12}\text{C}_6$ ]AN desialilado + $\alpha$ 1-3,4 defucosilado | Assignación enlace Fuc                 |
|---------------|-----------|-------------|---|---|--|
| <b>H5N4</b>   | Isómero 1 | 28,2        | 392811  | 348791  | -                                      |
| <b>H5N4F1</b> | Isómero 1 | 28,3        | 31920   | 29711   | 1 Fuc core                             |
|               | Isómero 2 | 29,0        | 20008   | 17474   | 1 Fuc core                             |
|               | Isómero 3 | 30,2        | 12472   | -   | 1 Fuc antena $\alpha$ 1-3              |
| <b>H6N5</b>   | Isómero 1 | 31,5        | 644160  | 666457  | -                                      |
| <b>H6N5F1</b> | Isómero 1 | 31,5        | 92394   | 95158   | 1 Fuc core                             |
|               | Isómero 2 | 32,1        | 19997   | 25852   | 1 Fuc core                             |
|               | Isómero 3 | 32,5        | 183321  | -   | 1 Fuc antena $\alpha$ 1-3              |
| <b>H6N5F2</b> | Isómero 1 | 32,6        | 48452   | -   | Al menos 1 Fuc antena $\alpha$ 1-3     |
| <b>H7N6</b>   | Isómero 1 | 33,5        | 298959  | 345043  | -                                      |
| <b>H7N6F1</b> | Isómero 1 | 33,5        | 64200   | 76963   | 1 Fuc core                             |
|               | Isómero 2 | 34,0        | 78593   | -   | 1 Fuc antena $\alpha$ 1-3              |
| <b>H7N6F2</b> | Isómero 1 | 33,5        | 24799   | 24713   | 2 Fuc core or $\alpha$ 1-2 unida a Gal |
|               | Isómero 2 | 34,1        | 20435   | -   | Al menos 1 Fuc antena $\alpha$ 1-3     |
|               | Isómero 3 | 34,4        | 22952   | -   | Al menos 1 Fuc antena $\alpha$ 1-3     |
| <b>H7N6F3</b> | Isómero 1 | 34,1        | 7247  | -   | Al menos 1 Fuc antena $\alpha$ 1-3     |
|               | Isómero 2 | 34,4        | 8236  | -   | Al menos 1 Fuc antena $\alpha$ 1-3     |

### 5.2.3. Caracterización de glicanos por espectrometría de masas en tándem

El establecimiento de metodologías por espectrometría de masas en tándem (MS/MS) para la caracterización estructural de glicanos ha suscitado gran interés en los últimos años [213–218]. Aunque la digestión con exoglicosidasas es de gran utilidad en la caracterización de los glicanos de una proteína, esta metodología presenta ciertas limitaciones al no disponer comercialmente de enzimas para todos los tipos de enlace, además de que en ocasiones su especificidad no puede ser completamente garantizada. Para solventar estos inconvenientes, algunos autores han propuesto la espectrometría de masas en tándem como alternativa para identificar los isómeros de un glicano a partir de sus espectros de masas, los cuales contienen

iones fragmento característicos de ciertas estructuras o tipos de enlace que pueden ser usados como iones diagnóstico [214,216–218].

En esta tesis doctoral, se ha establecido un método por CapZIC-HILIC-MS/MS en modo negativo utilizando la hAGP como glicoproteína modelo. El método de separación por CapZIC-HILIC y la asignación previa con exoglicosidasas permitió determinar qué iones fragmento descritos en la bibliografía tenían realmente un valor diagnóstico para caracterizar de manera fiable los enlaces de los ácidos siálicos y de las fucosas. Se utilizaron dos espectrómetros de masas híbridos: un QTrap 6500 (AB Sciex®) y un LTQ Orbitrap Velos (Thermo Scientific®). Aunque a nivel de número de iones fragmento se obtuvieron resultados similares, se decidió llevar a cabo la caracterización de los glicanos empleando el LTQ Orbitrap dada la posibilidad de obtener una masa exacta de los iones fragmento, su mejor resolución y su mayor velocidad de escaneo. Además, si lo comparamos con la fragmentación obtenida por CID, el uso de la celda HCD del LTQ Orbitrap permitió detectar de más iones producto con un mayor valor diagnóstico, como los iones fragmento *cross-ring*. En la **Tabla 5.7** se recogen los iones fragmento considerados en la bibliografía como diagnóstico de características estructurales específicas, tales como el tipo de enlace de los ácidos siálicos para glicanos intactos o de las fucosas para glicanos desialilados.

**Tabla 5.7.** Resumen de los iones fragmento diagnóstico descritos previamente en la bibliografía para la caracterización del enlace de los ácidos siálicos (SiA) y de las fucosas (Fuc) de N-glicanos mediante espectrometría de masas en tándem en modo negativo.

|   | Nomenclatura  | m/z , m/z – H <sub>2</sub> O | Fragmento diagnóstico de: |
|---|---|------------------------------|---------------------------|
| Iones diagnóstico del enlace de los SiA | <b>Iones fragmento para N-glicanos intactos</b>   |                              |                           |
|   | <sup>0,4</sup> A <sub>2</sub> -CO <sub>2</sub>  | 306,1189                     | SiA α2-6                  |
|   | <sup>0,4</sup> A <sub>2</sub> or <sup>1,3</sup> A <sub>2</sub> or <sup>2,4</sup> A <sub>2</sub> | 350,1087                     | SiA α2-3                  |
|   | B <sub>3</sub> /Y <sub>6</sub> or <sup>3,5</sup> A <sub>2</sub>                                 | 364,1243                     | SiA α2-6                  |

|   |  |                    |  |
|---|--|--------------------|--|
|   | B <sub>2</sub> -CO <sub>2</sub>  | 408,1506           | SiA α2-3                                 |
|   | C <sub>2</sub>   | 470,1510           | SiA α2-6                                 |
| Iones diagnóstico del enlace de las Fuc | <b>Iones fragmento para N-glicanos desialilados</b>  |                    |  |
|   | Y <sub>1</sub>   | 297,1212, 279,1106 | Ausencia de Fuc core                     |
|   | C <sub>1</sub>   | 325,1134           | Fuc unida a Gal <sup>c</sup>             |
|   | Y <sub>1</sub>   | 461,1898, 443,1791 | Fuc core <sup>c</sup>                    |
|   | B <sub>2</sub>   | 528,1928, 510,1823 | Fuc antena <sup>c</sup>                  |
|   | <sup>0,4</sup> A <sub>4</sub> /Y <sub>5</sub> or <sup>1,3</sup> A <sub>4</sub> /Y <sub>5</sub> <sup>a</sup> or<br><sup>0,4</sup> A <sub>3</sub> or <sup>1,3</sup> A <sub>3</sub> | 570,2034           | Fuc antena <sup>c</sup>                  |
|   | Y <sub>2</sub>   | 646,2585, 628,2479 | Fuc core <sup>c</sup>                    |
|   | <sup>1,5</sup> X <sub>2</sub>  | 674,2534           | Fuc core <sup>c</sup>                    |
|   | <sup>0,2</sup> X <sub>2</sub>  | 688,2690           | Fuc core <sup>c</sup>                    |
|   | <sup>0,4</sup> A <sub>4</sub> /Y <sub>5</sub> or <sup>1,3</sup> A <sub>4</sub> /Y <sub>5</sub> <sup>a</sup>  | 789,2777, 771,2671 | Ausencia de Fuc en la antena ramificada  |
|   | <sup>0,4</sup> A <sub>4</sub> /Y <sub>5</sub> or <sup>1,3</sup> A <sub>4</sub> /Y <sub>5</sub> <sup>a</sup>  | 935,3356           | Fuc en la antena ramificada <sup>c</sup> |
|   | <sup>1,5</sup> A <sub>3</sub> or <sup>1,5</sup> A <sub>5</sub> /Y <sub>3</sub> <sup>b</sup>  | 1009,3723          | Fuc antena <sup>c</sup>                  |

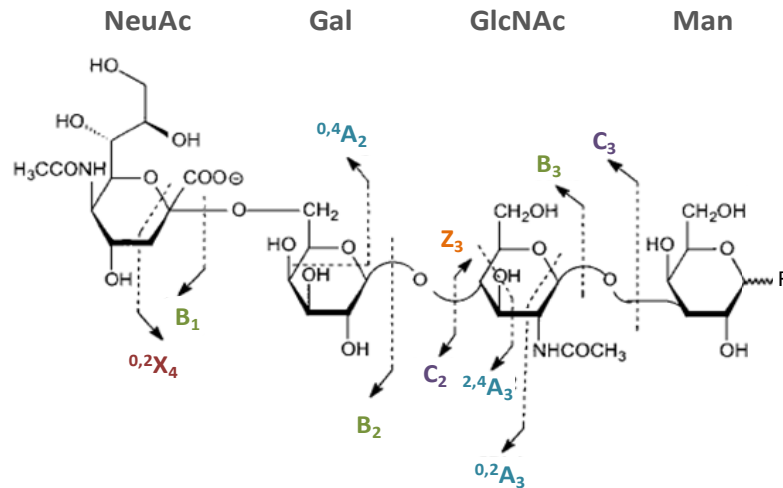
<sup>a</sup> El fragmento <sup>1,3</sup>A<sub>4</sub> se corresponde con la misma masa exacta que el fragmento <sup>2,4</sup>A<sub>4</sub>.

<sup>b</sup> Única estructura posible en el caso de glicanos biantenarios.

<sup>c</sup> La masa exacta de este ion fragmento incluye la masa de una unidad de fucosa.

Además, se incluyen las diferentes posibilidades de ion fragmento que dan lugar a la misma masa exacta, determinados con el programa SimGlycan 5 (Premier Biosoft®). La nomenclatura empleada para la denominar los iones producto en esta tabla y a lo largo de toda esta sección, fue la propuesta por Domon y Costello [6] y posteriormente modificada por Harvey *et. al.* [219]. Siguiendo esta nomenclatura, las letras A<sub>i</sub>, B<sub>i</sub> y C<sub>i</sub> se usan para designar los fragmentos que contienen el monosacárido terminal de las antenas, mientras que las letras X<sub>j</sub>, Y<sub>j</sub> y Z<sub>j</sub> representan los iones que aun contienen el monosacárido del extremo reductor. Los subíndices indican la posición en relación al monosacárido considerado terminal en cada caso y los superíndices indican los carbonos donde tiene lugar la fragmentación dentro del anillo. Los

fragmentos del tipo  $A_i/X_j$  no contienen ningún monosacárido terminal y se refieren al fragmento de glicano comprendido entre los iones  $A_i$  y  $X_j$ . A modo de ejemplo, la **Figura 5.16** muestra la ilustración de una antena de glicano con la representación de algunos iones fragmento usando esta nomenclatura.



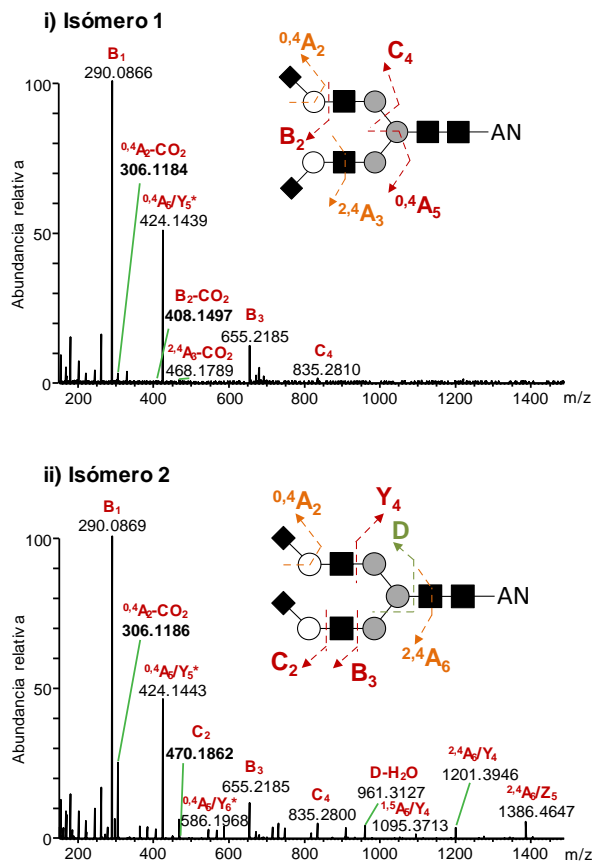
**Figura 5.16.** Posible estructura de la antena de un N-glicano (NeuAc( $\alpha$ 2-6)-Gal-GlcNAc-Man) con la representación de algunos fragmentos típicos empleando la nomenclatura propuesta por Domon and Costello [39] y posteriormente modificada por Harvey *et. al.* [40].

### 5.2.3.1. Caracterización de los enlaces de los ácidos siálicos

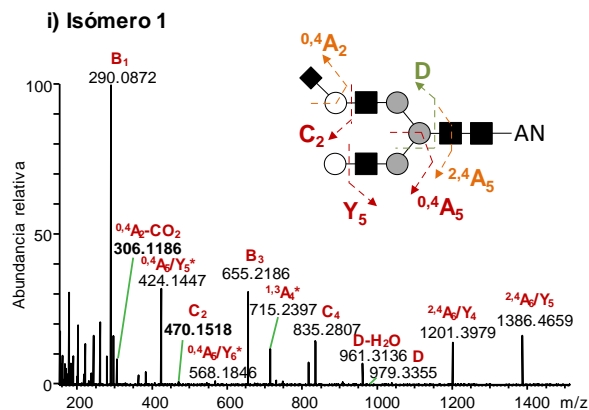
Para poder confirmar o descartar los iones fragmento descritos en la bibliografía como diagnóstico de cada tipo de enlace de ácido siálico, se analizaron por CapZIC-HILIC-MS/MS los glicanos intactos (de muestras de hAGP digeridas con PNGasa F). Los iones 306,1, 408,1 y 470,1, propuestos por Wheeler *et al.* para asignar el enlace de los SiA [218], fueron detectados también en nuestro caso y seleccionados como iones fragmento diagnóstico. Por el contrario, en nuestro estudio no se consideraron los iones 350,1 y 364,1, descritos por Michael *et al.* [217], dado que diversas estructuras daban lugar a estas masas demostrando su falta de especificidad (ver **Tabla 5.7**). Los iones diagnóstico observados en los espectros de masas en tándem para los glicanos bi- y triantenarios estudiados se muestran en la **Tabla 5.8**, mientras

que en la **Figura 5.17** se presentan, a modo de ejemplo, los espectros de MS/MS obtenidos para los glicanos biantenarios (el H5N4S1 y los dos isómeros del H5N4S2).

### A) H5N4S2



### B) H5N4S1



**Figura 5.17.-** (A) Espectros de MS/MS de los isómeros 1 (i) y 2 (ii) del glicano intacto H5N4S2 en una muestra patrón de hAGP. (B) Espectro de MS/MS del glicano H5N4S1 obtenido de una muestra patrón de hAGP después de su digestión con sialidasa  $\alpha$ 2-3. (\*) Este ión fragmento puede corresponder a otras posibles estructuras detalladas en el artículo 3.2 de esta tesis doctoral.

En el caso del H5N4S1 y del isómero 2 del H5N4S2 (**Figura 5.17B** y **Figura 5.17A(ii)**) se observan dos iones a  $m/z$  306,1 y 470,1, descritos como característicos de SiA enlazados  $\alpha$ 2-6. En ambos casos se constató la ausencia del ion 408,1, característico de SiA  $\alpha$ 2-3, por lo que se deduce que para estos dos glicanos todos los SiA se encuentran enlazados  $\alpha$ 2-6. Además, se puede observar que el isómero 2 del H5N4S2 presenta una mayor intensidad relativa del ion 306,1 respecto al H5N4S1, lo que sugiere una proporción más grande de SiA  $\alpha$ 2-6 como corresponde

a la presencia de dos SiA. Respecto al isómero 1 del glicano H5N4S2, se detectaron los iones 306,1 y 408,1 (ver **Tabla 5.8** y **Figura 5.17A(i)**), aunque a una intensidad inferior debido a su coelución con el glicano H6N5S3 que es mucho más abundante, lo que confirmó la presencia de un SiA de cada tipo en este isómero. Estos resultados coinciden con la asignación previa llevada a cabo mediante digestión con exoglicosidasas, confirmando el valor diagnóstico de estos tres iones fragmento (306,1, 408,1 y 470,1) para asignar los enlaces de los SiA.

Adicionalmente, y para obtener información sobre la localización de los diferentes SiA en las antenas del glicano, se analizaron mediante CapZIC-HILIC-MS/MS los glicanos desialilados  $\alpha$ 2-3, obtenidos de muestras de hAGP digeridas con PNGasa F y sialidasa  $\alpha$ 2-3. La presencia de los iones D y D-H<sub>2</sub>O en el glicano H5N4S1 (m/z 979,3 y 961,3), característicos de la presencia de una antena  $\alpha$ 1-6 (6-antena) con el SiA terminal, reveló que su SiA  $\alpha$ 2-6 se localiza en la 6-antena (ver **Tabla 5.8** y **Figura 5.17B**). Consecuentemente, estos resultados permitieron saber que el isómero 1 del H5N4S2, que tras la digestión se convierte en H5N4S1 (apartado 5.2.2.1), presenta el SiA  $\alpha$ 2-3 en la antena  $\alpha$ 1-3 (3-antena) y el SiA  $\alpha$ 2-6 en la 6-antena. En este caso no se pudieron detectar los esperados iones D y D-H<sub>2</sub>O a m/z 979,3 y 961,3, posiblemente debido a la coelución de este isómero 1 con el glicano H6N5S3, como se ha comentado anteriormente.

**Tabla 5.8.** Iones fragmento observados en el espectro de masas en tándem de los isómeros de los glicanos bi- y triantenarios estudiados de la hAGP y usados para la caracterización del enlace y posición de los ácidos siálicos. Se incluye la nomenclatura, asignación y estructura de cada ion fragmento.

| Glicano            | Isómero                       | Fragmentos observados (m/z, m/z - H <sub>2</sub> O) | Nomenclatura                                   | Diagnóstico                        | Estructura |
|--------------------|-------------------------------|---|--|------------------------------------|------------|
| H5N4S1             | Isómero 1                     | 306,1186  | <sup>0,4</sup> A <sub>2</sub> -CO <sub>2</sub> | SiA α2-6                           |            |
|                    |                               | 470,1518  | C <sub>2</sub>                                 | SiA α2-6                           |            |
|                    |                               | 979,3355, 961,3136                                  | D  | 6-antena sialilada y no ramificada |            |
| H5N4S2             | Isómero 1                     | 306,1184  | <sup>0,4</sup> A <sub>2</sub> -CO <sub>2</sub> | SiA α2-6                           |            |
|                    |                               | 408,1497  | B <sub>2</sub> -CO <sub>2</sub>                | SiA α2-3                           |            |
|                    |                               | -   | -  | -                                  |            |
| H5N4S2             | Isómero 2                     | 306,1186  | <sup>0,4</sup> A <sub>2</sub> -CO <sub>2</sub> | SiA α2-6                           |            |
|                    |                               | 470,1862  | C <sub>2</sub>                                 | SiA α2-6                           |            |
|                    |                               | 979,3333, 961,3127                                  | D  | 6-antena sialilada y no ramificada |            |
| H6N5S2             | Isómero 2                     | 306,1186  | <sup>0,4</sup> A <sub>2</sub> -CO <sub>2</sub> | SiA α2-6                           |            |
|                    |                               | 470,1498  | C <sub>2</sub>                                 | SiA α2-6                           |            |
|                    |                               | 775,2702  | <sup>0,2</sup> A <sub>4</sub>                  | Ramificació 4 sialilada            |            |
| H6N5S2             | Isómero 3 <sup>b</sup>        | 979,3119, 961,3134                                  | D  | 6-antena sialilada y no ramificada |            |
|                    |                               | 306,1186  | <sup>0,4</sup> A <sub>2</sub> -CO <sub>2</sub> | SiA α2-6                           |            |
|                    |                               | 470,1505  | C <sub>2</sub>                                 | SiA α2-6                           |            |
| 979,3166, 961,3119 | D                             | 6-antena sialilada y no ramificada                  |  |                                    |            |
| H6N5S3             | Isómero 1                     | 306,1186  | <sup>0,4</sup> A <sub>2</sub> -CO <sub>2</sub> | SiA α2-6                           |            |
|                    |                               | 408,1492  | B <sub>2</sub> -CO <sub>2</sub>                | SiA α2-3                           |            |
|                    |                               | 775,2560  | <sup>0,2</sup> A <sub>4</sub>                  | Ramificació 4 sialilada            |            |
| H6N5S3             | Isómero 2                     | n.d., 961,3204                                      | D  | 6-antena sialilada y no ramificada |            |
|                    |                               | 306,1187  | <sup>0,4</sup> A <sub>2</sub> -CO <sub>2</sub> | SiA α2-6                           |            |
|                    |                               | 408,1546  | B <sub>2</sub> -CO <sub>2</sub>                | SiA α2-3                           |            |
| 775,2590           | <sup>0,2</sup> A <sub>4</sub> | Ramificació 4 sialilada                             |  |                                    |            |
| H6N5S3             | Isómero 3                     | n.d., 961,3251                                      | D  | 6-antena sialilada y no ramificada |            |
|                    |                               | 306,1187  | <sup>0,4</sup> A <sub>2</sub> -CO <sub>2</sub> | SiA α2-6                           |            |
|                    |                               | 470,1503  | C <sub>2</sub>                                 | SiA α2-6                           |            |
| 775,2588           | <sup>0,2</sup> A <sub>4</sub> | Ramificació 4 sialilada                             |  |                                    |            |
| H6N5S3             | Isómero 3                     | n.d., 961,3140                                      | D  | 6-antena sialilada y no ramificada |            |

<sup>a</sup> Isómero no estudiado.

<sup>b</sup> Isómero no presente en la glicoproteína nativa. Detectado sólo después de la digestión con sialidasa α2-3.

(n.d.) Ion fragmento no detectado.

La caracterización de los glicanos triantenarios se llevó a cabo siguiendo la misma estrategia que con los biantenarios. Los resultados obtenidos confirmaron, también con estos glicanos, la asignación previa realizada con exoglicosidasas y dieron validez a los iones fragmento seleccionados, tal y como se explica en detalle en el **artículo 3.2**. Cabe destacar que mientras el ion 306,1 se detectó en todos los isómeros que presentaban algún SiA unido  $\alpha$ 2-6, el ion 470,1 sólo se observó cuando todos los SiA del isómero en cuestión estaban enlazados  $\alpha$ 2-6 (**Tabla 5.8**).

Con el fin de mejorar la caracterización del glicano H6N5S3 asignando la posición de cada SiA en las antenas, se analizaron también por MS/MS los isómeros 2 y 3 del H6N5S2, obtenidos después de la digestión con sialidasa  $\alpha$ 2-3. Al eliminarse los SiA con enlace  $\alpha$ 2-3 tras la digestión, los isómeros 1 y 2 del H6N5S3 se convierten en los isómeros 3 y 2 del H6N5S2, respectivamente (**artículo 3.2**). La detección de los iones D a  $m/z$  979,3 y 961,3 en ambos isómeros, desveló la presencia de SiA en la 6-antena, igual que en el caso de los biantenarios. Por tanto se pudo concluir que todos los isómeros del glicano H6N5S3 poseen un único SiA  $\alpha$ 2-6 en la 6-antena no ramificada. Asimismo, la identificación del fragmento a  $m/z$  775,3 en el isómero 2 del H6N5S2 y su ausencia en el isómero 3 del mismo glicano permitió localizar el SiA  $\alpha$ 2-6 de la 3-antena en la ramificación 2 ( $\beta$ 1-2, **Figura 1.2**) en el caso del isómero 2 y en la ramificación 4 ( $\beta$ 1-4, **Figura 1.2**) para el isómero 3 (**Tabla 5.8**). De este modo, se pudo determinar la posición de todos los SiA en los isómeros 1 y 2 del glicano H6N5S3.

La caracterización de los glicanos tetraantenarios por MS/MS fue de nuevo muy complicada dada su baja abundancia y su mayor dificultad para ionizarse. Sin embargo, se llevó a cabo una estimación de la composición de SiA  $\alpha$ 2-6 en los cuatro isómeros del glicano H7N6S3, teniendo en cuenta la intensidad relativa del ion 306,1, tal y como se explica en el **artículo 3.2** de esta tesis. De esta manera, se identificó el isómero 2 con dos SiA  $\alpha$ 2-3 y uno  $\alpha$ 2-6 y el isómero 3 a la inversa, mientras que se asignó el isómero 4 con todos los SiA  $\alpha$ 2-6 y el isómero 1 con todos

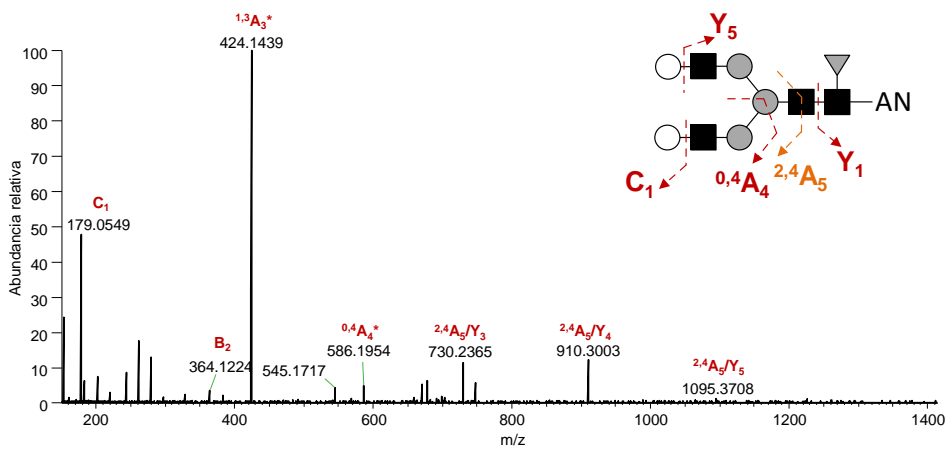


los SiA  $\alpha$ 2-3. Por otro lado, la caracterización de los SiA de todos los glicanos realizada tanto en el estudio con exoglicosidasas como por MS/MS, demostró la existencia de una estrecha relación entre el contenido de SiA  $\alpha$ 2-6 y el orden de elución de los isómeros en la columna ZIC-HILIC, observándose que aquellos isómeros con mayor proporción de SiA  $\alpha$ 2-6 se eluyen más tarde.

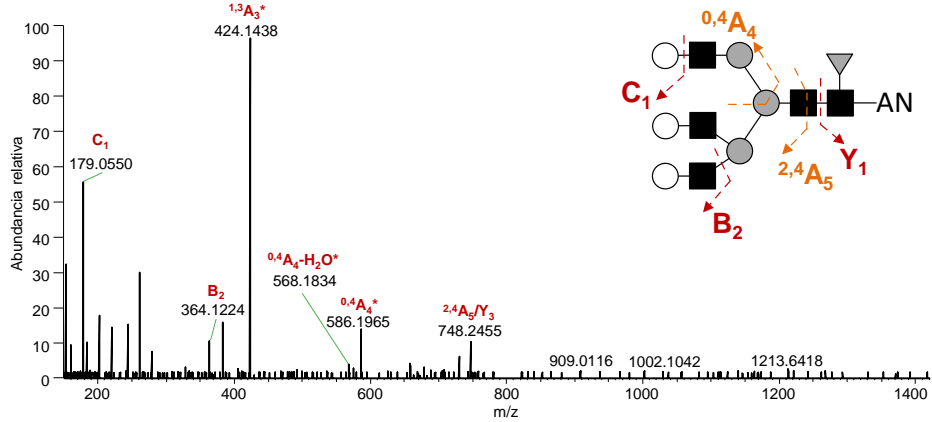
### 5.2.3.3. Caracterización de los enlaces de las fucosas

La caracterización de los enlaces de las fucosas de los glicanos de la hAGP también se llevó a cabo por espectrometría de masas en tándem. Habitualmente, los iones diagnóstico de los enlaces de las fucosas son detectados a muy baja intensidad, o directamente no se observan en los espectros de MS/MS, dado que las fucosas son muy lábiles y se pierden con facilidad durante la fragmentación en la celda de colisión (especialmente la Fuc antena  $\alpha$ 1-3) [214]. Por este motivo, el establecimiento de metodología analítica por MS/MS para identificar nuevos iones producto que permitan diferenciar sin ambigüedades entre una fucosa unida al core o a la antena, resulta de especial interés. En esta tesis, para verificar el valor diagnóstico de ciertos iones descritos en la bibliografía (ver **Tabla 5.7**), se analizaron los glicanos desialilados de la hAGP por CapZIC-HILIC-MS/MS (**artículo 3.2**). Se seleccionaron los glicanos fucosilados H5N4F1 (isómeros 1 y 2) y H6N5F1 (isómeros 1, 2 y 3), dado que estos glicanos presentan una abundancia suficiente para identificar los iones fragmento con potencial valor diagnóstico de cada tipo de enlace de las fucosas. Además, los isómeros 1 y 2 de estos glicanos habían sido perfectamente caracterizados con exoglicosidasas como Fuc core y el isómero 3 del H6N5F1 como Fuc antena  $\alpha$ 1-3. En la **Figura 5.18** se muestran los espectros de MS/MS y en la **Tabla 5.9** los iones fragmento detectados para los glicanos seleccionados.

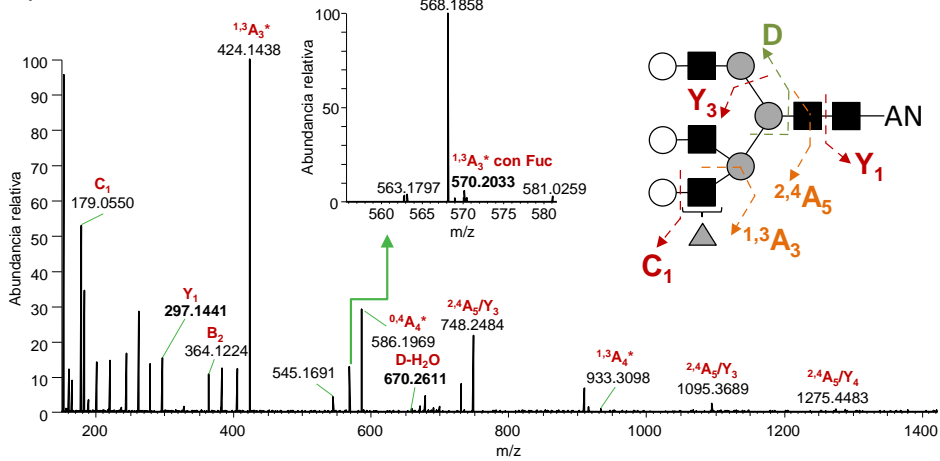
A) H5N4F1 – Isómero 1 y 2



B) H6N5F1 – Isómero 1 y 2

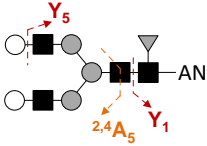
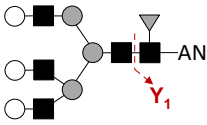
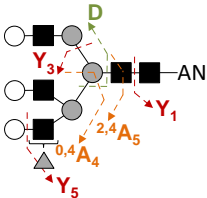


C) H6N5F1 – Isómero 3



**Figura 5.18.-** Espectros de MS/MS de los glicanos H5N4F1 (isómeros 1 y 2 (A)) y H6N5F1 (isómeros 1 y 2 (B) e isómero 3 (C)) en una muestra patrón de hAGP después de la digestión con sialidasa total. (\*) Este ión fragmento puede corresponder a otras posibles estructuras detalladas en el artículo 3.2 de esta tesis doctoral.

**Tabla 5.9.** Iones fragmento observados en el espectro de masas en tándem de los isómeros de los glicanos de la hAGP estudiados y usados para la caracterización del enlace de las fucosas. Se incluye la nomenclatura, asignación y estructura de cada ion fragmento.

| Glicano | Isómero       | Fragmentos observados (m/z, m/z - H <sub>2</sub> O) | Nomenclatura                                    | Diagnóstico            | Estructura   |
|---------|---------------|---|---|------------------------|--|
| H5N4F1  | Isómero 1 y 2 | 461,2251, 443,2035                                  | Y <sub>1</sub>                                  | Fuc core               |   |
|         | Isómero 1 y 2 | 443,2040 <sup>#</sup>                               | Y <sub>1</sub>                                  | Fuc core               |   |
| H6N5F1  | Isómero 3     | 297,1441  | Y <sub>1</sub>                                  | Ausencia Fuc core      |  |
|         |               | 570,2033  | <sup>0,4</sup> A <sub>4</sub> /Y <sub>5</sub> * | Fuc antena             |  |
|         |               | 670,2611  | D   | 6-antena no ramificada |  |

\* Este fragmento puede corresponder a otras estructuras posibles detalladas en el artículo 3.2 de esta tesis doctoral.

<sup>#</sup> Detectado solo en la mitad de los análisis.

Algunos fragmentos diagnóstico de Fuc core como el ion 443,2 (Y<sub>1</sub>, ver **Tabla 5.9**), se detectaron en los isómeros 1 y 2 de los glicanos H5N4F1 y H6N5F1 pero a muy baja intensidad y no en todos los análisis, demostrando la falta de fiabilidad de este ion. En cambio, la ausencia del ion 297,1 en los espectros de MS/MS de estos isómeros (**Figura 5.18A y B**), reveló que podían presentar la fucosa unida al core. Este ión corresponde al fragmento Y<sub>1</sub> pero sin la fucosa enlazada y su abundancia en los espectros de los glicanos no fucosilados es muy elevada. Además, se detectaron otros iones fragmento a m/z 1095,3 que podrían sugerir la ausencia de Fuc antena en los isómeros 1 y 2 del glicano biantenarico (**Figura 5.18A y Tabla 5.9**).

En relación al isómero 3 del H6N5F1 (**Figura 5.18C y Tabla 5.9**), se detectó el ion a m/z 297,1 con prácticamente la misma intensidad relativa que en el espectro de MS/MS del

correspondiente glicano no fucosilado (**artículo 3.2**), demostrando que la fucosa no estaba enlazada al core. Asimismo, se confirmó la existencia de Fuc antena en este isómero por la presencia del fragmento 570,2, a pesar de su baja abundancia (**Figura 5.18C**). Además, la ausencia del ion 325,1 descrito en la bibliografía (**Tabla 5.7**) reveló que la fucosa se encontraba unida al GlcNAc de la antena y no a la Gal, tal y como se describió en el estudio previo con exoglicosidasas. Por otro lado, la presencia del fragmento D a  $m/z$  670,1, correspondiente a la 6-antena no ramificada, puso de manifiesto que la fucosa del isómero 3 del H6N5F1 se encontraba unida al GlcNAc de una de las ramificaciones de la 3-antena. Teniendo en cuenta los resultados obtenidos en los diferentes isómeros estudiados, se corroboró la dificultad para detectar fragmentos diagnóstico que permitan caracterizar el tipo de enlace de las fucosas. De este modo, de entre todos los fragmentos descritos en la bibliografía para este fin (**Tabla 5.7**), sólo se pudo confirmar el valor diagnóstico del ión 297,1 para identificar la ausencia de Fuc core así como del ión a  $m/z$  570,2 para asignar los isómeros con Fuc antena  $\alpha$ 1-3.

#### **5.2.4. Caracterización de glicanos por espectrometría de masas de movilidad iónica (IM-MS)**

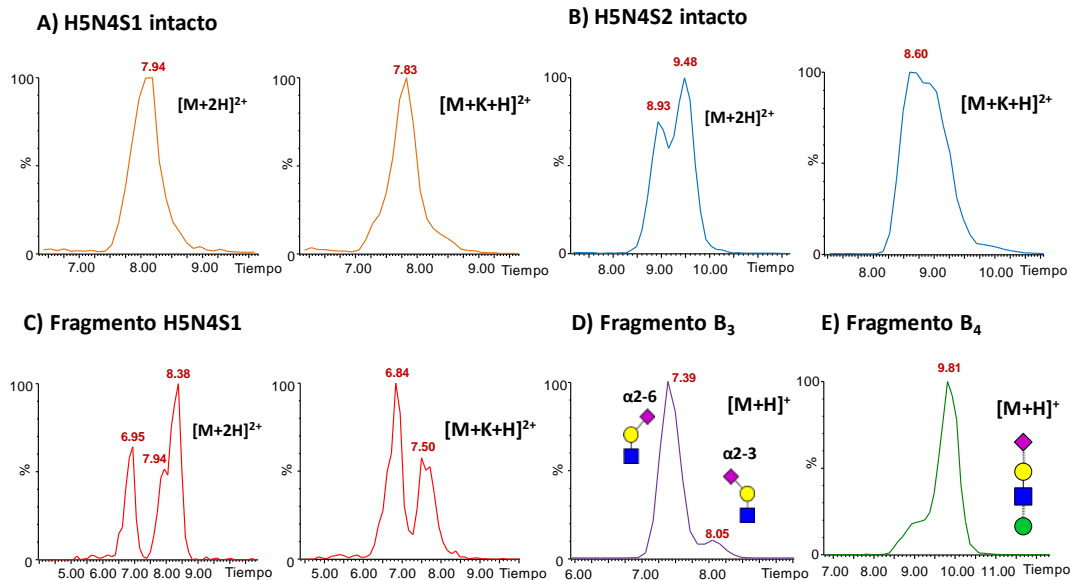
En esta tesis se ha demostrado que la cromatografía de líquidos capilar acoplada a la espectrometría de masas (CapLC-MS) puede resultar muy útil para separar y caracterizar glicanos y sus correspondientes isómeros, empleando fases estacionarias específicas como las ZIC-HILIC. Sin embargo, el establecimiento de estas metodologías suele ser complejo y laborioso además de requerir tiempos de análisis bastante largos.

La espectrometría de movilidad iónica (IMS) ha suscitado gran interés en los últimos años como alternativa para el análisis de glicoconjugados gracias a su capacidad para separar directamente compuestos isoméricos [184,186,192–194,220]. Tal y como se comentó en la introducción, esta técnica de separación en fase gas permite separar compuestos en base a su forma y tamaño además de por su carga, lo cual, junto con la detección por MS (IM-MS), la

convierte en una metodología de alto potencial para el análisis de glicanos y sus isómeros. En este sentido, se han establecido diversos métodos por IM-MS para la separación de isómeros de glicanos y glicopéptidos utilizando *travelling wave ion mobility spectrometry* (TWIMS), algunos de ellos centrándose en la diferenciación de los enlaces  $\alpha$ 2-3 y  $\alpha$ 2-6 de los SiA [186,193,221]. Sin embargo, mientras que para glicanos isoméricos muy pequeños se han obtenido *drift times* que permiten su diferenciación [186,221], sólo se ha conseguido una separación parcial en el caso de estructuras biantenarias más complejas [193]. Recientemente, fue descrito el uso del *high-field asymmetric waveform ion mobility* (FAIMS) para separar isómeros [183]. Aunque en este caso se consiguieron separar estructuras biantenarias isoméricas, las cuales presentaban diferentes tipos de enlace de SiA, estas correspondían a dos glicanos sintéticos con una estructura perfectamente conocida.

Dada la necesidad de establecer metodologías por IM-MS que permitan caracterizar glicanos más ramificados en muestras más complejas, como los glicanos obtenidos de la digestión enzimática de una glicoproteína, en esta tesis doctoral también se evaluó la eficacia del TWIMS para estudiar los isómeros de los glicanos de la hAGP. En primer lugar, fue necesario cambiar el método de digestión con PNGasaF ya que el surfactante NP-40, utilizado en los estudios anteriores de esta tesis (**artículos 3.1 y 3.2**), provocaba supresión iónica disminuyendo la intensidad de los glicanos y dificultando su detección. De este modo, la hAGP se trató con ditiotreitól (DTT) e iodoacetamida (IAA) antes de la digestión con PNGasa F, con el objeto de reducir y alquilar los puentes disulfuro. Los glicanos obtenidos se analizaron tras su purificación, mediante nanoESI-IM-MS con infusión directa, tanto en modo negativo como en positivo. Los resultados en modo negativo no fueron satisfactorios ya que no se consiguió una separación significativa entre isómeros, ni a partir del análisis del glicano intacto ni de sus fragmentos obtenidos por MS/MS, tal y como ya se observó para los glicanos de la hAGP en un estudio previo [193]. Por el contrario, en modo positivo, tras optimizar la altura y velocidad de la onda de la celda de movilidad (WH y WV respectivamente), se obtuvo más de un *drift time*

para ciertos glicanos demostrando la presencia de estructuras isoméricas. En la **Figura 5.19A, B** y **C** se muestran los resultados para los glicanos H5N4S1 y H5N4S2 intactos y para el fragmento H5N4S1, obtenido a partir de la fragmentación en la celda de movilidad iónica del glicano H5N4S2, respectivamente.

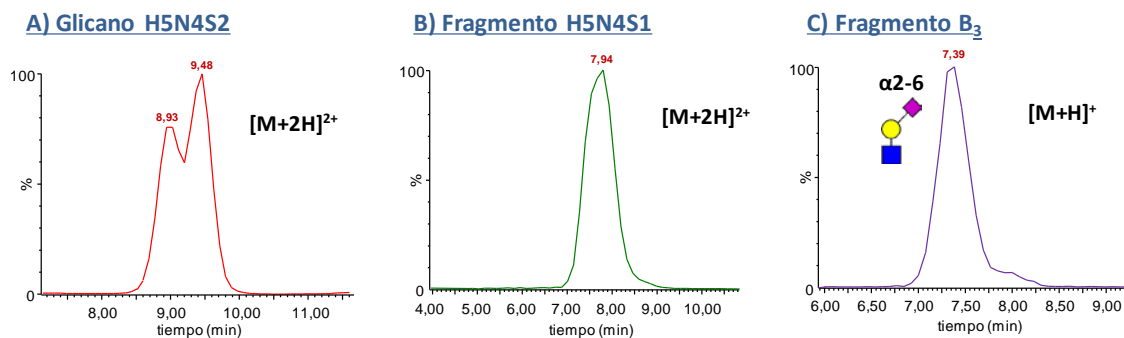


**Figura 5.19.-** ATDs de los aductos mayoritarios detectados para los glicanos intactos H5N4S1 (A) y H5N4S2(B), en una muestra patrón de hAGP, así como para los fragmentos H5N4S1 (C), B<sub>3</sub> (D) y B<sub>4</sub> (E) del glicano H5N4S2.

En los tres casos se detectaron dos aductos abundantes ( $[M+2H]^{2+}$  y  $[M+K+H]^{2+}$ ) que presentaban diferentes distribuciones del tiempo de llegada (*arrival time distribution* – ATD). De este modo, mientras que para el glicano intacto H5N4S1 se observó un único *drift time* en los dos aductos, para el glicano H5N4S2 se observaron claramente dos *drift times* en el caso del aducto  $[M+2H]^{2+}$ . Estos resultados demostraron, por un lado, que la formación de aductos afecta a la separación por IM-MS, tal y como ya ha sido descrito por otros autores [222,223]. Por otro lado, la observación de un único *drift time* para el glicano H5N4S1 y de dos para el H5N4S2 concordaría con las conclusiones alcanzadas en los capítulos anteriores acerca del número de isómeros de cada uno de estos glicanos (**Tabla 5.5**). En cambio, en el caso del H5N4S1, resultado de la fragmentación del H5N4S2 (**Figura 5.19C**), se obtuvo más de un *drift*

*time* en los dos aductos detectados, demostrando también la presencia de compuestos isoméricos, a diferencia de lo que ocurre con el glicano intacto H5N4S1. De nuevo se observó la mejor separación en el ATD del aducto  $[M+2H]^{2+}$ , y consecuentemente, este aducto fue el escogido para continuar el estudio. También se examinaron los ATD de otros fragmentos del glicano H5N4S2, tal y como se muestra en las **Figuras 5.19C** y **D** correspondientes a los fragmentos B<sub>3</sub> (H1N1S1) y B<sub>4</sub> (H2N1S1), respectivamente, observándose dos *drift times* en ambos casos. Para el fragmento B<sub>3</sub> se obtuvo una mejor separación, tal y como ya habían descrito otros autores [186,193], y además, la única diferencia posible entre estructuras es la debida al ácido siálico, por lo que se seleccionó este fragmento como el más adecuado para diferenciar SiA  $\alpha$ 2-6 y  $\alpha$ 2-3. La asignación de enlaces de SiA para los isómeros detectados en el ATD del fragmento B<sub>3</sub> se llevó a cabo según los resultados de estudios anteriores [186,193], donde el primer isómero con *drift time* menor (y por tanto sección transversal de colisión menor, CCS) correspondería al SiA  $\alpha$ 2-6 y el segundo al SiA  $\alpha$ 2-3 (ver **Figura 5.19D**).

Con el fin de comprobar si las diferencias en los *drift times* detectados en el ATD del glicano H5N4S2 y de sus fragmentos pueden atribuirse realmente a diferencias en el enlace de los SiA, se llevó a cabo una digestión de los glicanos de la hAGP con sialidasa  $\alpha$ 2-3 y se analizaron por nanoESI-IM-MS. Tal y como se muestra en la **Figura 5.20A**, después de la digestión con sialidasa  $\alpha$ 2-3 se siguen obteniendo dos *drift times* para el glicano H5N4S2, lo que demuestra que las diferencias entre las estructuras de los compuestos detectados no son debidas a los tipos de enlace de sus SiA. Por el contrario, con el fragmento H5N4S1, sólo se obtuvo un *drift time* (**Figura 5.20B**), a diferencia de lo observado con el mismo fragmento del glicano antes de la digestión con sialidasa  $\alpha$ 2-3 (**Figura 5.19C**).



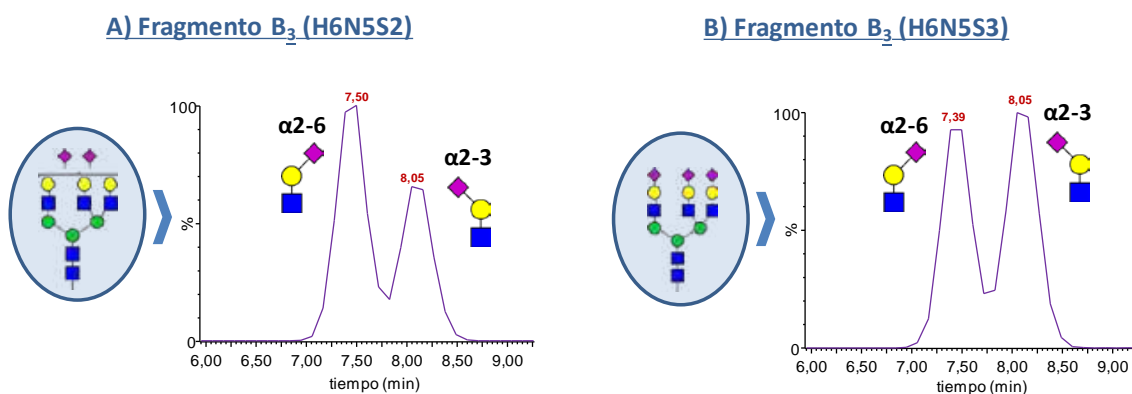
**Figura 5.20.-** ATDs del glicano intacto H5N4S2 (A) así como de sus fragmentos H5N4S1 (B) y B<sub>3</sub> (C) obtenidos de una muestra patrón de hAGP después de su digestión con sialidasa  $\alpha 2-3$ .

Si se comparan los ATD obtenidos para el glicano intacto H5N4S1 (**Figura 5.19A**) y para el fragmento H5N4S1 tras la digestión (**Figura 5.20B**), se puede observar que el valor del *drift time* coincide en ambos compuestos con un valor de 7,94. Estos resultados demostraron que ambos compuestos presentan sólo un isómero con el SiA unido  $\alpha 2-6$ , tal y como se describió en los estudios anteriores por Cap-ZIC-HILIC-MS (**artículos 3.1 y 3.2**). De este modo, se pudo confirmar que los isómeros del fragmento H5N4S1, son debidos a diferencias en el tipo de enlace de su SiA. Asimismo, tal y como muestra la **Figura 5.20C**, para el fragmento B<sub>3</sub> después del tratamiento con sialidasa  $\alpha 2-3$  se obtuvo también sólo un *drift time* con un valor de 7,39, que concuerda con al primer *drift time* del fragmento sin digerir, previamente asignado como SiA  $\alpha 2-6$  (**Figura 5.19D**). Con estos resultados se concluyó que en el caso de glicanos con un solo SiA, como el H5N4S1, o de estructuras más pequeñas, como el fragmento B<sub>3</sub>, la IM-MS permite la separación de isómeros que presentan diferentes tipos de enlace de SiA. Por el contrario, se descartó que la presencia de diferentes *drift times* en el ATD de estructuras complejas con varios SiA, como el H5N4S2, fuera debida a diferencias en los enlaces de los SiA, contrariamente a lo descrito en otros estudios [193].

Así pues, la caracterización de los enlaces de los SiA de los glicanos mediante IM-MS, debe abordarse a partir del estudio de sus fragmentos. En esta tesis se estudiaron los ATD del fragmento B<sub>3</sub> en diversos glicanos de la hAGP, incluyendo por primera vez estructuras



triantenarias, y se compararon las abundancias relativas de los isómeros del fragmento B<sub>3</sub> con los estudios previos por CapZIC-HILIC-MS realizados en esta tesis doctoral. Como se vio en la **Figura 5.19D**, el ATD del fragmento B<sub>3</sub> para el glicano biantenarico H5N4S2, mostró dos *drift times* correspondientes a SiA  $\alpha$ 2-6 y  $\alpha$ 2-3, siendo mucho más abundante el SiA  $\alpha$ 2-6 (*drift time* 7,39). Estas abundancias relativas podrían estar en concordancia con la asignación previa de dicho glicano donde se observaron dos isómeros, el primero muy poco abundante con un SiA  $\alpha$ 2-3 y un SiA  $\alpha$ 2-6 y el segundo muy abundante con los dos SiA  $\alpha$ 2-6 (ver **Tablas 5.5 y 5.8**). Del mismo modo, se estudiaron los ATD del fragmento B<sub>3</sub> de los glicanos triantenaricos de la hAGP (H6N5S2 y H6N5S3). Como se puede observar en la **Figura 5.21**, en ambos casos se observaron también dos *drift times* correspondientes a SiA enlazado  $\alpha$ 2-6 y  $\alpha$ 2-3, respectivamente.

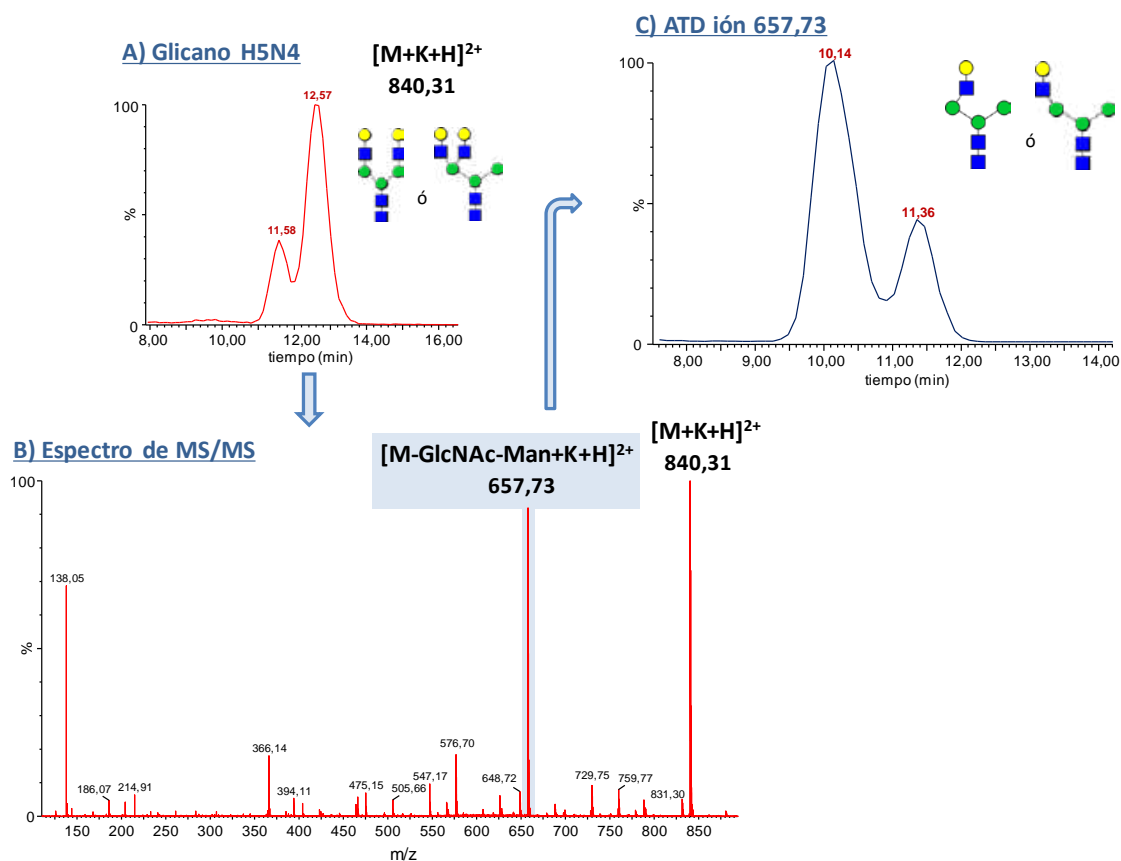


**Figura 5.21.-** ATDs del fragmento B<sub>3</sub> de los glicanos triantenaricos H6N5S2 (A) y H6N5S3 (B) en una muestra patrón de hAGP.

En el caso del H6N5S2, aunque la intensidad relativa del *drift time* SiA  $\alpha$ 2-6 sigue siendo superior, la abundancia del *drift time* SiA  $\alpha$ 2-3 también es bastante elevada. En la caracterización previa por CapZIC-HILIC-MS se obtuvieron dos isómeros de la misma abundancia (**Tablas 5.5 y 5.6**), el primero asignado con un SiA  $\alpha$ 2-3 y un  $\alpha$ 2-6 y el segundo con los dos SiA  $\alpha$ 2-6. De este modo, la abundancia relativa de los *drift times* del fragmento B<sub>3</sub> no coincidiría del todo con dicha caracterización dado que, en este caso, se debería observar una abundancia relativa del 75% para el *drift time* con SiA  $\alpha$ 2-6 y del 25% para el *drift time* con SiA

$\alpha$ 2-3. Para el glicano H6N5S3 (**Figura 5.21B**), se obtuvo una abundancia relativa similar para los dos *drift times* en el ATD del fragmento B<sub>3</sub>. Estos resultados estarían de nuevo en discordancia con los obtenidos en la caracterización previa, donde claramente se observó una mayor proporción de SiA unido  $\alpha$ 2-6 en dicho glicano. De este modo, en términos generales, se demostró que la abundancia relativa de los *drift times* obtenidos en el ATD del fragmento B<sub>3</sub> no puede ser utilizada para estimar la proporción de SiA enlazados  $\alpha$ 2-6 y  $\alpha$ 2-3 en un glicano. Sin embargo, el estudio del ATD de dicho fragmento podría ser usado para comparar muestras control y patológicas y así evaluar, si existe, la sobreexpresión de un tipo concreto de enlace de SiA en ciertas patologías como el cáncer. Esta metodología por nanoESI-IM-MS, empleando el fragmento B<sub>3</sub> obtenido de la fragmentación del glicano de interés, en un futuro debería ser evaluada con fines diagnósticos dados sus cortos tiempos de análisis.

Finalmente, para investigar las diferencias estructurales responsables de la obtención de dos *drift times* en el ATD del glicano H5N4S2, se realizó también una digestión con sialidasa total de los glicanos de la hAGP y se analizaron por nanoESI-IM-MS. Los resultados obtenidos para el glicano biantenarico desialilado (H5N4) se muestran en la **Figura 5.22**. En este caso se estudió el aducto  $[M+K+H]^{2+}$  dado que era el más abundante. Tal y como puede observarse en el ATD del H5N4, también se obtuvieron dos *drift times* para el glicano desialilado confirmando que las diferencias observadas no eran debidas al tipo de enlace de los SiA. Se llevó a cabo la fragmentación de dicho glicano con el fin de buscar fragmentos producto que desvelaran alguna característica estructural que diferenciara ambos isómeros. El espectro de MS/MS obtenido para el glicano H5N4 se presenta en la **Figura 5.22B**, observándose el ión molecular y un fragmento bastante intenso de  $m/z$  657,73, que correspondería a la pérdida de una Man + GlcNAc (**Figura 5.22C**). La presencia de dos *drift times* también en el ATD del fragmento a  $m/z$  657.73 puso de manifiesto que las diferencias entre estos isómeros son debidas a tipos de enlace diferentes entre la Man y el GlcNAc, dado que es el único factor diferenciador que puede presentar esta estructura.



**Figura 5.22.-** ATDs del glicano H5N4 (A), proveniente de una muestra patrón de hAGP después de su digestión con sialidasa total, y del ion 657,73 (C), observado en el espectro de MS/MS del glicano H5N4 (B).

De este modo, y teniendo en cuenta que el glicano desialilado H5N4 tiene que dar lugar a dos enlaces GlcNAc-Man diferentes, se propusieron las posibles estructuras presentadas en la **Figura 5.22A** para el glicano H5N4. Los resultados obtenidos desvelaron que en estructuras de glicano complejas, la IM-MS es capaz de detectar diferencias entre isómeros que difieren en el enlace de las antenas. No obstante, con IM-MS se requerirá de estudios de fragmentación complementarios como los realizados con el fragmento B<sub>3</sub> o bien de tratamientos con exoglicosidasas específicas para identificar los isómeros debidos únicamente a diferentes tipos de enlace de los SiA.

### 5.2.5. Comparación de los métodos de caracterización de glicanos establecidos

Las tres metodologías establecidas para la caracterización de glicanos y sus isómeros en esta tesis doctoral nos han proporcionado mucha información acerca de los glicanos de la hAGP y cada una de ellas nos ha permitido profundizar más en su identificación. Por este motivo, las tres técnicas deben ser consideradas complementarias. Aun así, cabe destacar que la caracterización mediante digestión con exoglicosidasas en combinación con la estrategia GRIL fue la que más glicanos permitió asignar, pudiendo caracterizar el enlace de los ácidos siálicos así como de las fucosas de la mayoría de glicanos de la hAGP. Además, dicha asignación sirvió de base para evaluar el potencial de los métodos por MS/MS y IM-MS. Por su parte, la metodología por MS/MS, nos proporcionó más información acerca de la posición de los ácidos siálicos y de las ramificaciones en las antenas. En cuanto al método por IM-MS, aunque fue el que menos información nueva nos ofreció de los glicanos de la hAGP, nos permitió establecer un método rápido y sencillo, a diferencia de las laboriosas digestiones con exoglicosidasas y los largos tiempo de análisis por CapZIC-HILIC-MS, para determinar los tipos de ácidos siálicos presentes en un glicano. Por todos estos motivos, si no es posible llevar a cabo una asignación completa utilizando los tres métodos, es necesario evaluar el tipo de metodología a usar dependiendo de la finalidad del estudio.

### 5.3. Estudio e identificación de glicanos biomarcadores

Las alteraciones en la glicosilación de las proteínas han sido descritas en diversas patologías como el cáncer y ciertos procesos inflamatorios [22,23,29,32,224]. La mayoría de las modificaciones detectadas en los glicanos han sido asociadas a un aumento o disminución de determinadas glicosiltransferasas, las cuales dan lugar a: niveles elevados de ácido siálico total (TSA), cambios de proporción de ácidos siálicos unidos  $\alpha$ 2-3 y  $\alpha$ 2-6, aumento de las ramificaciones, adición de unidades extra de N-acetilactosamina (poliLacNAc) así como sobreexpresión de ciertos epítomos fucosilados (**Figura 1.5**) [22,23,29,30,34,224]. Además, en el caso del cáncer, algunos autores han descrito modificaciones que podrían estar más relacionadas con fases iniciales de los tumores y otras con fases más avanzadas o de metástasis que permitirían la extravasación de las células tumorales [31–33,38]. Por este motivo, en los últimos años, la búsqueda de estructuras de glicano que puedan actuar como biomarcadoras de diversos procesos patológicos ha suscitado gran interés biomédico, tanto para ser usados en el diagnóstico precoz de la enfermedad como para controlar su progresión. Asimismo, se ha demostrado la importancia del análisis de los diferentes isómeros de los glicanos, dado que a veces es únicamente el isómero con un cierto tipo de enlace el que está relacionado con una patología y puede actuar como biomarcador de la misma.

En esta tesis doctoral, se han llevado a cabo dos trabajos diferentes (**artículo 4.1 y 4.2**) con el fin de identificar los glicanos de la hAGP y de la transferrina de ratón (mTf), así como sus diferentes isómeros, capaces de actuar como biomarcadores de cáncer de páncreas (PDAC) o artritis inducida por colágeno (CIA), respectivamente. Se han seleccionado estas dos glicoproteínas de fase aguda dado su potencial como biomarcadoras de estas enfermedades evidenciado en la literatura [44,54,55,225,226]. Para identificar inequívocamente las alteraciones en los glicanos de dichas proteínas en presencia de estas enfermedades, se han comparado los perfiles de los glicanos en muestras control y patológicas empleando la

estrategia GRIL de marcaje isotópico con  $[^{12}\text{C}_6]\text{AN}/[^{13}\text{C}_6]\text{AN}$ . Asimismo, dada la variabilidad existente entre muestras patológicas y el gran número de isómeros de glicanos identificados, en ambos estudios se ha llevado a cabo un análisis multivariante de datos para reducir la complejidad de la interpretación de los mismos e identificar correctamente los glicanos con mayor potencial biomarcador. Esta estrategia combinada nos ha permitido proponer isómeros de glicanos que podrían ser biomarcadores de estas patologías.

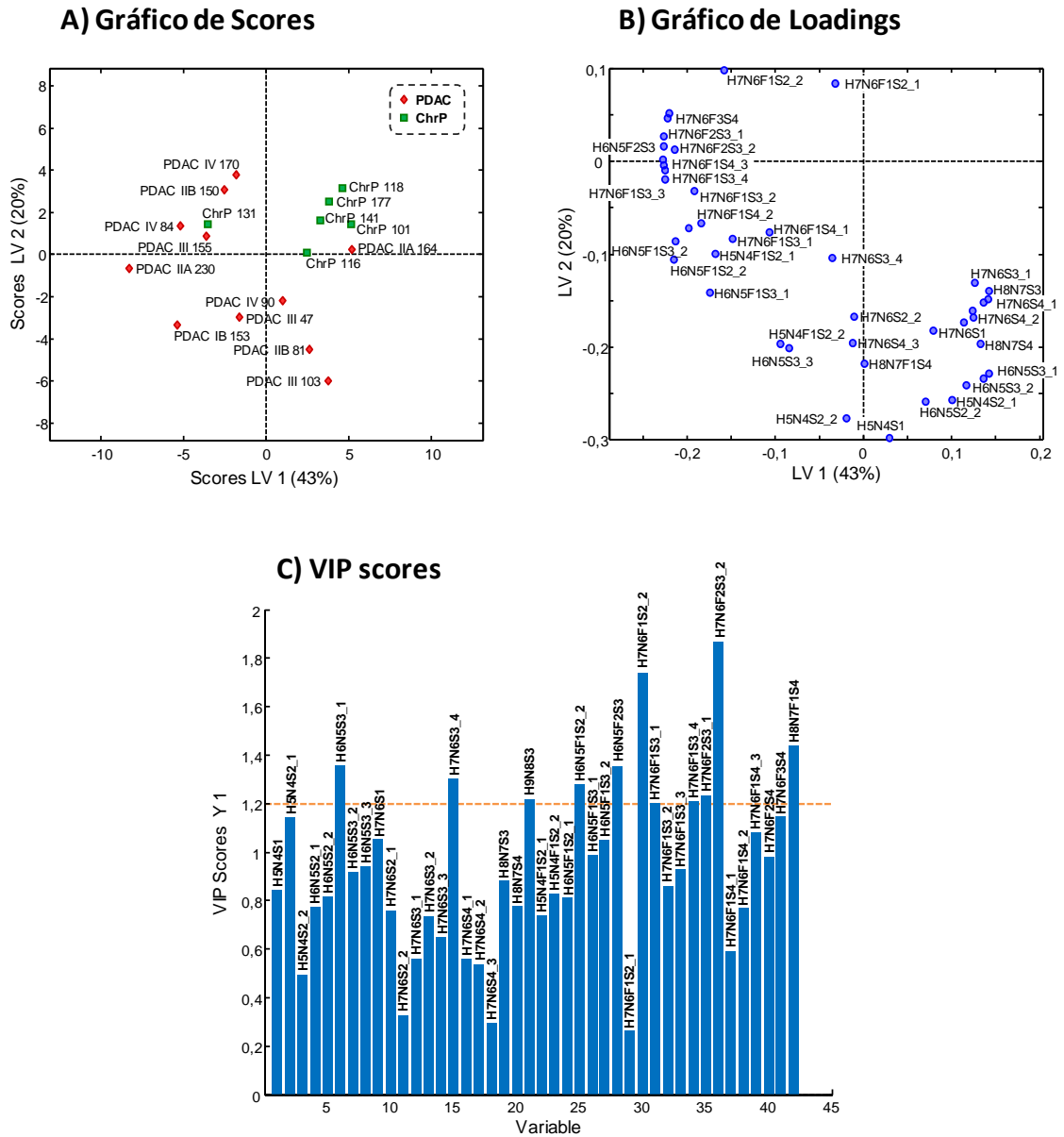
### **5.3.1. Análisis de los glicanos de la hAGP en muestras de cáncer de páncreas y pancreatitis crónica**

En trabajos previos de nuestro grupo de investigación, se analizaron los glicanos de la hAGP en muestras de suero de pacientes con cáncer de páncreas (PDAC) y pancreatitis crónica (ChrP) [54,55]. En estos estudios se detectó un aumento de ciertas estructuras sialofucosiladas que permitirían discriminar entre PDAC y ChrP, así como diferencias en la expresión de ciertos isómeros en estadios iniciales de PDAC [54,55]. Estos resultados, obtenidos con un número relativamente reducido de muestras patológicas, demostraron la necesidad de profundizar en el estudio y caracterización de los glicanos de la hAGP como posibles biomarcadores de PDAC.

#### **5.3.1.1. Análisis de los glicanos intactos**

Con el fin de encontrar isómeros de glicanos de la hAGP capaces de diferenciar PDAC de ChrP, se analizaron 6 muestras de suero de pacientes con ChrP, 12 de pacientes con PDAC de diferentes estadios de la enfermedad (estadios I, II, III y IV) y 4 muestras control de personas sanas (HC). Los detalles relacionados con la edad y género de cada una de las muestras de suero analizadas se presentan en el **artículo 4.1** de esta tesis doctoral. En primer lugar, la hAGP se purificó a partir de las muestras de suero mediante cromatografía de inmunofinidad (IAC), utilizando el proceso previamente descrito para el análisis de los glicanos de la hAGP [54,55], y se cuantificó la cantidad de proteína mediante espectrofotometría UV-Vis

(Nanodrop™). Después de la digestión de las muestras de hAGP con PNGasa F para conseguir la liberación de los glicanos y de la purificación de los mismos, se empleó la estrategia GRIL con el fin de realizar una cuantificación relativa fiable. Con este propósito, se prepararon mezclas equimolares de los glicanos de la hAGP de cada muestra patológica (derivatizados con [<sup>12</sup>C<sub>6</sub>]AN) y de los glicanos de la hAGP de una mezcla de muestras control (*pool* de HC, derivatizados con [<sup>13</sup>C<sub>6</sub>]AN) y se analizaron por CapZIC-HILIC-MS. El *pool* de HC se preparó mezclando la misma cantidad de hAGP purificada a partir de 4 muestras control diferentes (HC 44, HC 48, HC 51 y HC 200). Dada la gran cantidad de muestras y número de isómeros identificados (42 isómeros), se aplicaron métodos quimiométricos 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 HC). En primer lugar, se empleó un análisis por componentes principales (PCA) para explorar los datos y estudiar la existencia de tendencias o grupos entre las diferentes muestras, así como detectar la presencia de *outliers*. De este modo, tal y como se explicó en el **artículo 4.1**, la muestra PDAC IIB 86 se descartó para el siguiente análisis de datos multivariante dado su comportamiento anómalo en el modelo. A continuación, con el fin de mejorar la separación entre grupos e identificar qué isómeros podrían diferenciar PDAC y ChrP, se llevó a cabo un análisis discriminante por mínimos cuadrados (PLS-DA). En la **Figura 5.23** se muestran los gráficos de scores (A), loadings (B) y de VIP scores (C) obtenidos en el modelo de PLS-DA.



**Figura 5.23.-** A) Gráfico de scores, B) Gráfico de loadings y C) VIP scores del modelo de PLS-DA aplicado a las relaciones de áreas (área del isómero de glicano en la muestra patológica vs. área del mismo isómero en el *pool* de controles sanos) obtenidas para cada isómero de glicano intacto de la hAGP en todas las muestras de PDAC y ChrP analizadas.

Como se observa en el gráfico de scores, dos variables latentes (LV) permitieron discriminar entre los dos grupos de muestras (explicando un 63 y un 43% de la variancia de X e Y, respectivamente). Sólo las muestras PDAC IIA 164 y ChrP 131 presentaron un comportamiento anormal, tal y como ya se observó en un estudio previo [55], y fueron mal clasificadas por el



modelo. En el gráfico de loadings (**Figura 5.23B**) se observa que los isómeros de los glicanos fucosilados están más relacionados con PDAC, ya que aparecen en la misma zona del gráfico que la mayoría de muestras de PDAC en el gráfico de *scores*, correspondiente a valores de LV1 negativos (**Figura 5.23A**). Además, el gráfico de VIP scores (**Figura 5.23C**) permitió cuantificar la influencia de cada isómero en la separación entre muestras de PDAC y ChrP. Aunque habitualmente se consideran importantes todas las variables con VIP superior a 1, en este caso, dado que había muchos isómeros significativos, sólo se consideraron aquellos con VIP superior a 1,2 (7 de un total de 42), de los cuales 5 tenían al menos una unidad de fucosa y todos ellos habían sido caracterizados con todos o la mayoría de los SiA unidos  $\alpha$ 2-6. Con estos resultados, se concluyó que los isómeros de los glicanos de la hAGP fucosilados con la mayoría de los SiA unidos  $\alpha$ 2-6 son importantes para discriminar PDAC de ChrP.

#### 5.3.1.2. Análisis de los glicanos desialilados

Dado que en el estudio de los glicanos intactos se confirmó que la fucosilación juega un papel importante en la diferenciación entre PDAC y ChrP, se llevó a cabo el análisis de los glicanos de la hAGP de muestras de suero una vez desialilados con sialidasa total, con el fin de averiguar qué tipo de fucosilación estaba más relacionada con este tipo de cáncer. Las muestras analizadas y sus características se presentan en la tabla 1 del **artículo 4.1** de esta tesis doctoral. De nuevo se empleó la cuantificación relativa de los glicanos mediante GRIL. No obstante, en este caso las muestras control (HC) se analizaron como muestras independientes y todas las muestras patológicas (PDAC y ChrP) y control (HC) se cuantificaron de manera relativa respecto a un patrón de hAGP comercial purificado a partir de suero humano, que presenta un perfil de glicosilación significativamente igual al de las muestras control, tal y como ya se había demostrado en un estudio previo [54]. De este modo se dispone de un perfil de glicanos de referencia constante para esta glicoproteína, ya que no siempre es posible disponer de un pool adecuado de muestras HC. Con este propósito, siguiendo la estrategia GRIL planteada en la

**Figura 1.10** de la introducción, las muestras de hAGP purificadas de cada muestra patológica o control se marcaron con [ $^{12}\text{C}_6$ ]AN, tras la liberación de los correspondientes glicanos mediante digestión con PNGasa F y sialidasa total, y las muestras de hAGP patrón se marcaron con [ $^{13}\text{C}_6$ ]AN. Seguidamente, se prepararon mezclas equimolares de cada muestra patológica/control y de hAGP patrón, y se analizaron por CapZIC-HILIC-MS.

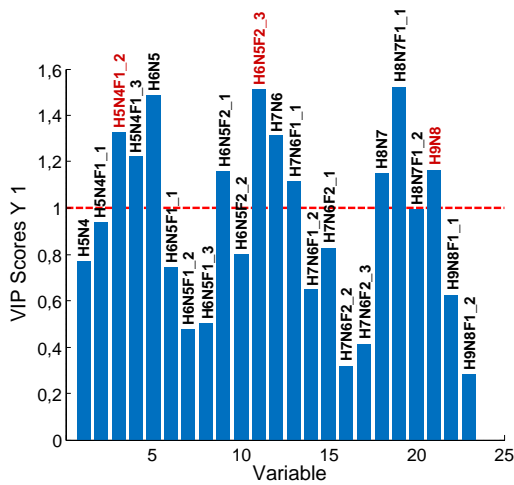
Nuevamente se emplearon herramientas quimiométricas para el análisis de los datos a partir de las relaciones de áreas obtenidas para cada isómero (área del glicano en la muestra patológica/control dividida entre el área del glicano en la muestra patrón). Primero, se aplicó un PCA pero no se detectó ningún *outlier* en el modelo, así que todas las muestras se tuvieron en cuenta para los siguientes tratamientos de datos. A continuación, se aplicaron dos modelos de PLS-DA, uno para estudiar la discriminación entre muestras de PDAC y HC y otro para diferenciar ChrP de HC. La separación obtenida entre los dos grupos de muestras fue buena en ambos casos, tal y como se detalla en el **artículo 4.1** de esta tesis doctoral. En la **Tabla 5.10** se muestran los isómeros identificados en el gráfico de VIP scores como significativos ( $\text{VIP} > 1$ ) en ambas diferenciaciones. Se puede observar que algunos isómeros son importantes tanto en PDAC como en ChrP, lo cual no es sorprendente dado que la sobreexpresión de ciertos isómeros se observó en ambas patologías, aunque generalmente en mayor grado en el caso de PDAC. Sin embargo, los glicanos H5N4F1\_3, H6N5F2\_1, H6N5F2\_2 y H8N7 (marcados en negrita en la **Tabla 5.10**) podrían ser interesantes ya que sólo fueron significativos en la diferenciación entre PDAC y HC.

**Tabla 5.10.-** Isómeros de glicano desialilados de la hAGP identificados por PLS-DA como significativos (VIP scores > 1) en las diferenciaciones de muestras de PDAC vs. controles sanos (HC) y de muestras de ChrP vs. HC.

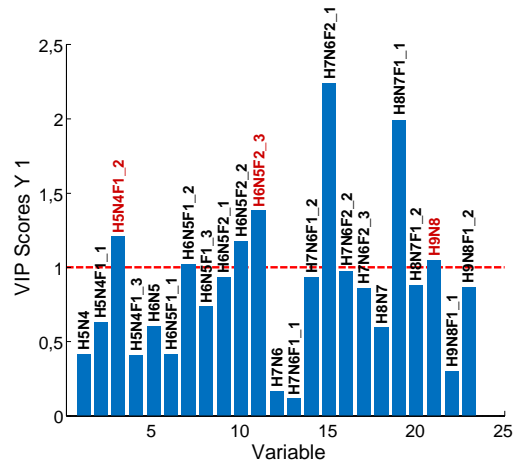
| PDAC vs. HC     | ChrP vs. HC |
|-----------------|-------------|
|                 | H5N4F1_1    |
| H5N4F1_2        | H5N4F1_2    |
| <b>H5N4F1_3</b> |             |
|                 | H6N5F1_2    |
| <b>H6N5F2_1</b> |             |
| <b>H6N5F2_2</b> |             |
| H6N5F2_3        | H6N5F2_3    |
| H7N6F1_2        | H7N6F1_2    |
| H7N6F2_2        | H7N6F2_2    |
| H7N6F2_3        | H7N6F2_3    |
| <b>H8N7</b>     |             |
|                 | H8N7F1_1    |
|                 | H8N7F1_2    |
|                 | H9N8F1_2    |

Por otro lado, con el objetivo de encontrar isómeros útiles para diferenciar entre diferentes estadios de PDAC, se construyeron 3 modelos de PLS-DA, comparando los pacientes de ChrP con los de PDAC a diferentes estadios de la enfermedad. Con este propósito, los pacientes de PDAC se dividieron en tres grupos diferentes: no avanzados (estadios I y II), localmente avanzados (estadio III) y con metástasis (estadio IV). Al contrario de lo ocurrido en el análisis de los glicanos intactos (apartado 5.3.1.1), en este caso el modelo clasificó correctamente las muestras PDAC IIA 164 y ChrP 131 en los correspondientes gráficos de scores (Figura 5, artículo 4.1). En la **Figura 5.24** se muestran los gráficos VIP scores de cada modelo de PLS-DA.

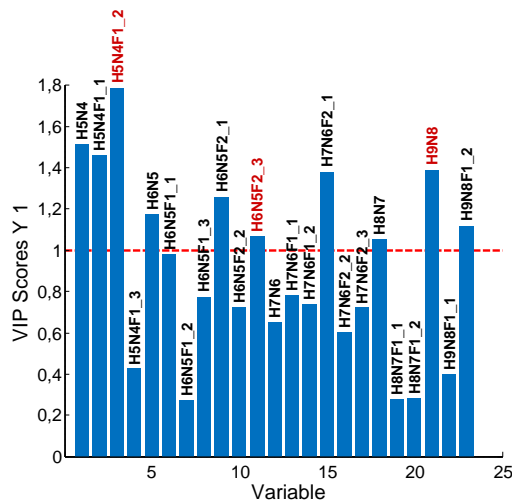
A) PDAC (Estadios I y II) vs. ChrP



B) PDAC (Estadio III) vs. ChrP



C) PDAC (Estadio IV) vs. ChrP



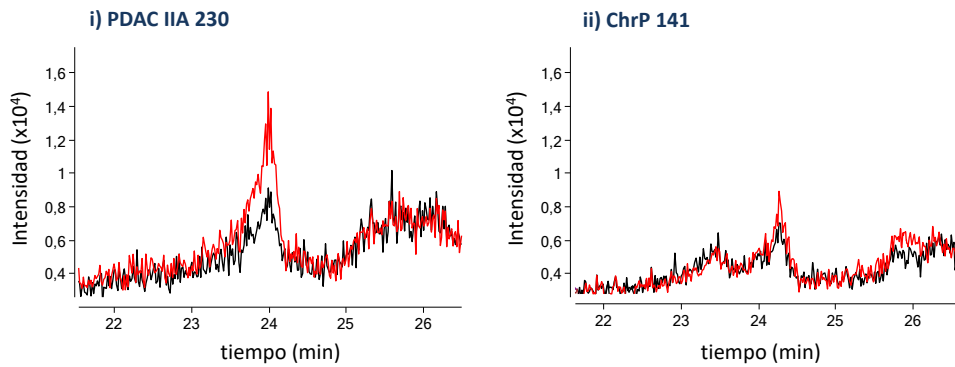
**Figura 5.24.-** VIP scores del modelo de PLS-DA aplicado a las relaciones de áreas (área del isómero de glicano en la muestra patológica vs. área del mismo isómero en la muestra patrón de hAGP) obtenidas en A) las muestras de PDAC en estadios I-II vs. las de ChrP, B) las muestras de PDAC en estadio III vs. las de ChrP y C) las muestras de PDAC en estadio IV vs. las de ChrP.

Como puede observarse, los glicanos considerados importantes en la diferenciación de PDAC y HC (H5N4F1\_3, H6N5F2\_1, H6N5F2\_2 y H8N7) no presentan VIP > 1 en todos los estadios de PDAC y, por tanto, sólo podrían actuar como biomarcadores de ciertos estadios de PDAC respecto a ChrP. Por otro lado, comparando los gráficos de VIP scores de los tres modelos, se puede observar que sólo 3 de un total de 23 isómeros desialilados de la hAGP son significativos

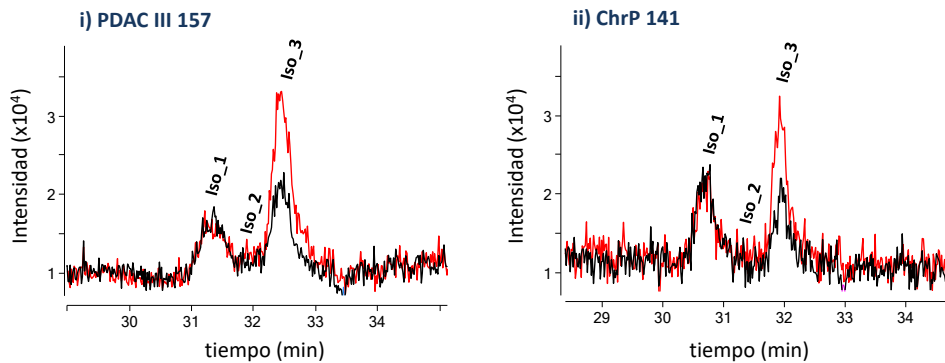
(VIP > 1) para distinguir PDAC de ChrP en todos los estadios de la patología (H5N4F1\_2, H6N5F2\_3 y H9N8, marcados en rojo en los tres gráficos de VIP scores). De estos tres glicanos, el H5N4F1\_2 presenta Fuc core y el H6N5F2\_3 presenta al menos una Fuc antena  $\alpha$ 1-3, lo que parece indicar que no hay un tipo de enlace de las fucosas específico de PDAC.

Finalmente, teniendo en cuenta los resultados obtenidos en el análisis tanto de los glicanos intactos como desialilados, se podría concluir que el glicano H6N5F2 (o H6N5F2S3 en el caso de los glicanos intactos) parece jugar un papel importante en la búsqueda de un nuevo biomarcador para el diagnóstico de PDAC ya que fue relevante en todos los modelos de PLS-DA. A modo de ejemplo, en la **Figura 5.25A** se muestran los EICs del glicano intacto H6N5F2S3 en las muestras patológicas PDAC IIA 230 y ChrP 141 (EICS en rojo) y en el pool de HC (EICs en negro). Del mismo modo, en la **Figura 5.25B** se presentan los EICs del correspondiente glicano desialilado (H6N5F2), en las muestras patológicas PDAC III 157 y ChrP 141 (EICs en rojo) así como en la muestra patrón de hAGP (EICs en negro). En relación al glicano intacto, se obtiene una sobreexpresión en PDAC con respecto al control sano y ChrP (**Figura 5.25A**). Además, tal y como puede observarse en el caso del glicano desialilado (**Figura 5.25B**), el isómero asignado con al menos una Fuc antena  $\alpha$ 1-3 (isómero 3, iso-3), aunque también se encuentra sobreexpresado en ChrP respecto al patrón, su sobreexpresión es mayor en el caso de PDAC. No obstante, en el futuro se debería analizar una cohorte superior de muestras de pacientes con PDAC y ChrP, así como controles sanos, para poder confirmar el potencial de este glicano como biomarcador.

**A) H6N5F2S3**



**B) H6N5F2**



**Figura 5.25.-** A) EICs obtenidos para el glicano H6N5S3 en las muestras patológicas PDAC IIA 230 (i) y ChrP 141 (ii). B) EICs obtenidos para el glicano H6N5F1 en las muestras patológicas PDAC IB 153 (i) y ChrP 141 (ii). EICs de los glicanos en las muestras patológicas se muestran en rojo y los EICs de los glicanos en el *pool* de controles sanos o en la muestra patrón de hAGP se muestran en negro.

**5.3.2. Análisis de los glicanos de la mTf en muestras de artritis inducida por colágeno**

Como se comentó en la introducción de esta tesis doctoral, la sialilación, fucosilación así como el grado de ramificación de los glicanos de la transferrina (Tf) se han visto alterados en presencia de diversas enfermedades inflamatorias, incluyendo la artritis reumatoide (RA) [69,227]. Por este motivo, y dado que a día de hoy no existe un único test para diagnosticar de la RA, diversos autores han abordado el estudio de la glicosilación de la transferrina humana (hTf) en muestras de pacientes con RA analizando las glicofomas intactas por enfoque isoeléctrico (IEF) [227], y más recientemente, por electroforesis capilar (CE) [69]. Sin embargo, los resultados reportados resultan contradictorios ya que mientras Feelders *et. al.* [227]

observaron un aumento de las glicoformas más sialiladas (S5 y S6) en RA, Gudowska *et. al.* [69] describieron un descenso de las sialoformas S3 y S5.

El estudio de enfermedades de origen humano en modelos de ratones se utiliza ampliamente como estudio preliminar a la investigación en humanos. En concreto, la artritis inducida por colágeno (CIA) en ratones se utiliza como alternativa para el estudio de la RA humana, debido al gran número de puntos en común que comparten ambas patologías, como la evolución de la enfermedad, las afectaciones en los tejidos de las articulaciones o la respuesta a los fármacos [65,228]. Por este motivo, este modelo se consideró adecuado como estudio previo que sirva de base para evaluar los cambios en la glicosilación de la Tf causados por la artritis [226]. En este sentido, en un trabajo anterior de nuestro grupo de investigación, se analizaron los glicopéptidos de la mTf de muestras de suero de ratones con CIA mediante CapLC-MS. En ese estudio se demostró que dichos glicopéptidos se ven modificados en presencia de CIA, observándose principalmente un aumento de la fucosilación y de las ramificaciones de los glicanos [225]. Estos resultados a nivel de glicopéptido, no proporcionan información sobre la posible alteración de los isómeros de los glicanos debido a la existencia de CIA. Por ello, en esta tesis doctoral, se llevó a cabo el análisis de los glicanos de la mTf para confirmar a nivel glicómico los resultados obtenidos en [225] y evaluar la relación de los diferentes isómeros con esta patología.

#### **5.3.2.1. Purificación de la mTf mediante cromatografía de inmunoafinidad**

Con el fin de analizar los glicanos de la mTf obtenidos de muestras de suero, fue necesario, en primer lugar, el establecimiento de una estrategia de purificación mediante cromatografía de inmunoafinidad (IAC), para capturar selectivamente la glicoproteína de interés evitando así la contribución de glicanos procedentes de otras glicoproteínas del suero. Con este propósito se empleó una columna de *sefarosa-CNBr* con un anticuerpo inmovilizado contra la hTf y el

proceso de purificación previamente usado para el análisis de los glicopéptidos de la hTf [24,68,225]. La **Tabla 5.11** muestra las proteínas identificadas por nanoLC-MS/MS en la fracción eluída por IAC de una muestra de suero de ratón control usando estas condiciones. Se han seleccionado aquellas que presentan valores de empAI superiores a 10% y que, por tanto, pueden considerarse abundantes en la muestra purificada por IAC.

**Tabla 5.11.-** Proteínas identificadas mediante nanoLC-MS/MS con valores de empAI superiores a 10%, en la muestra de ratón control purificada por cromatografía de inmunoafinidad (IAC), usando las condiciones de IAC iniciales y optimizadas.

| Nombre proteína <sup>a</sup>       | Identificador <sup>b</sup> | MW (kDa) | Puntos N-glicosilación | empAI <sup>c</sup> (%) |
|------------------------------------|----------------------------|----------|------------------------|------------------------|
| <b>Condiciones IAC iniciales</b>   |                            |          |                        |                        |
| Serum albumin                      | P07724                     | 68,6     | 0                      | 93,8                   |
| Serotransferrin                    | Q92111                     | 76,7     | 1                      | 34,3                   |
| Apolipoprotein A-I                 | Q00623                     | 30,6     | 0                      | 24,1                   |
| Vitamin D-binding protein          | P21614                     | 53,6     | 1                      | 23,2                   |
| Pregnancy zone protein             | Q61838                     | 165,7    | 11                     | 21,0                   |
| Alpha-1-antitrypsin 1-5            | Q00898                     | 45,9     | 3                      | 18,0                   |
| Serine protease inhibitor A3K      | P07759                     | 46,9     | 4                      | 16,3                   |
| Alpha-1-antitrypsin 1-3            | Q00896                     | 45,8     | 3                      | 14,5                   |
| Immunoglobulin kappa constant      | P01837                     | 11,8     | 0                      | 13,7                   |
| Alpha-1-antitrypsin 1-4            | Q00897                     | 46,0     | 3                      | 12,9                   |
| Alpha-1-antitrypsin 1-2            | P22599                     | 45,9     | 3                      | 11,6                   |
| Serine protease inhibitor A3M      | Q03734                     | 47,0     | 3                      | 10,1                   |
| <b>Condiciones IAC optimizadas</b> |                            |          |                        |                        |
| Serotransferrin                    | Q92111                     | 76,7     | 1                      | 55,2                   |
| Serum albumin                      | P07724                     | 68,6     | 0                      | 21,3                   |

<sup>a</sup>Nombre proteína de acuerdo a UniProt o NCBI.

<sup>b</sup>Identificador proteína en UniProtKB/Swiss-Prot o NCBI.

<sup>c</sup>*Exponentially Modified Protein Abundance Index* (empAI): cuantificación relativa de las proteínas en una mezcla en base a la cobertura de la proteína por las coincidencias en los péptidos en un resultado de búsqueda en base de datos.

En este caso, la mTf fue la segunda proteína más abundante con un empAI del 34,3% pero, dado que la mayoría de proteínas identificadas eran glicoproteínas, se decidió desarrollar un nuevo método de purificación más selectivo. Se consiguieron mejores resultados preparando



una columna IAC nueva, empleando un protocolo de unión del anticuerpo diferente, el cual se detalla en el **artículo 4.2**, y modificando algunas de las etapas del proceso de purificación: aumentando los volúmenes de lavado de la columna y de elución, y añadiendo 0,25 M de NaCl al tampón de unión para conseguir una mayor selectividad (**artículo 4.2**). Utilizando estas condiciones óptimas, en la fracción eluída sólo se identificaron la mTf y la albúmina mediante nanoLC-MS/MS, con valores de empAI superiores a 10% (ver **Tabla 5.11**). Dado que la albúmina de ratón es una proteína no glicosilada, se establecieron como óptimas estas condiciones para la purificación de la mTf en muestras de suero dado que no existe contribución de otras glicoproteínas al perfil de glicanos obtenido.

#### **5.3.2.2. Análisis de los glicanos de la mTf**

Con el objetivo de identificar modificaciones en la glicosilación de la mTf en ratones con CIA, se estableció en primer lugar un perfil de glicanos de referencia de dicha proteína. En la **Tabla 5.12** se muestran los glicanos y sus correspondientes isómeros detectados por CapZIC-HILIC-MS en una muestra de suero control comercial después de su purificación mediante IAC. La mayoría de glicanos corresponden a estructuras biantenarias, siendo el H5N4S2 el glicano más abundante. Todos los ácidos siálicos presentes en los glicanos de la mTf son ácidos N-glicolilneuramínicos (NeuGc), tal y como está descrito en la bibliografía [229]. Estos resultados están en concordancia con los resultados previos a nivel de glicopéptido [225].

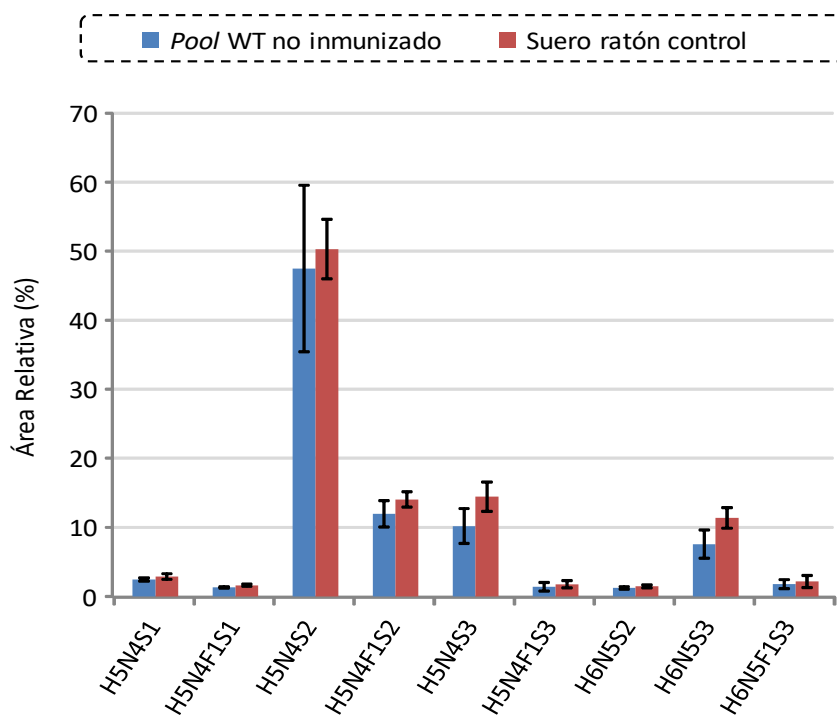
**Tabla 5.12.-** Glicanos detectados mediante CapZIC-HILIC-MS en la muestra de suero de ratón control después de su purificación por IAC.

| Glicano  | Isómero | t <sub>R</sub><br>(min) | Área relativa (%) <sup>a</sup> | M <sub>teo</sub> <sup>b</sup> | Error<br>(ppm) |
|----------|---------|-------------------------|--------------------------------|-------------------------------|----------------|
| H5N4S1   | 1       | 24,0                    | 2,41                           | 2024,7216                     | 8,4            |
| H5N4F1S1 | 1       | 24,8                    | 1,53                           | 2170,7796                     | 9,5            |
| H5N4S2   | 1       | 21,2                    | 9,99                           | 2331,8120                     | 3,1            |
|          | 2       | 21,9                    | 45,3                           | 2331,8120                     | 2,6            |
| H5N4F1S2 | 1       | 21,8                    | 1,84                           | 2477,8699                     | 5,1            |
|          | 2       | 22,5                    | 13,4                           | 2477,8699                     | 4,3            |
| H5N4S3   | 1       | 20,8                    | 1,61                           | 2638,9023                     | 6,3            |
|          | 2       | 21,6                    | 10,5                           | 2638,9023                     | 5,2            |
| H5N4F1S3 | 1       | 21,9                    | 1,21                           | 2784,9602                     | 10,9           |
| H6N5S2   | 1       | 23,3                    | 1,19                           | 2696,9441                     | 9,8            |
| H6N5S3   | 1       | 21,7                    | 0,49                           | 3004,0345                     | 7,3            |
|          | 2       | 22,3                    | 3,45                           | 3004,0345                     | 5,2            |
|          | 3       | 22,9                    | 5,90                           | 3004,0345                     | 5,6            |
| H6N5F1S3 | 1       | 24,2                    | 0,67                           | 3150,0924                     | 9,7            |
|          | 2       | 24,6                    | 0,55                           | 3150,0924                     | 10,3           |

<sup>a</sup>Área relativa calculada como el área de pico de cada isómero de glicano dividida entre la suma de las áreas de pico de todos los isómeros de glicano detectados.

<sup>b</sup>Masa teórica del glicano derivatizado con [<sup>12</sup>C<sub>6</sub>]-anilina.

A continuación, se comparó el perfil de glicanos de este suero con el obtenido al analizar un *pool* de sueros de ratón *wild-type* antes de inducirles CIA (WT). Como puede observarse en el gráfico de barras de la **Figura 5.26**, en las dos muestras se detectaron los mismos glicanos y con abundancias similares, confirmando que este suero comercial puede ser utilizado como referencia para el análisis de los glicanos en ratones no inmunizados (WT) y con CIA (WT-CIA).



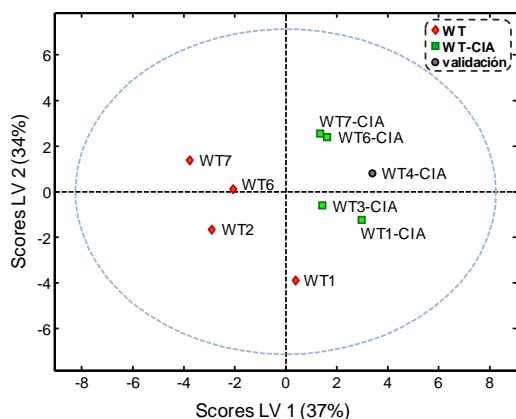
**Figura 5.26.-** Gráfico de barras del promedio de las áreas relativas de los glicanos obtenidos en el *pool* de muestras de ratón WT no inmunizadas y en el suero de ratón control purificados por IAC utilizando las condiciones óptimas. Las áreas relativas se calcularon como el área de pico de cada glicano dividido entre la suma de las áreas de pico de todos los glicanos detectados.

Seguidamente, se empleó la estrategia GRIL con  $[^{12}\text{C}_6]/[^{13}\text{C}_6]\text{AN}$ . Las muestras de mTf purificadas de cada muestra (WT o WT-CIA) se derivatizaron con  $[^{12}\text{C}_6]\text{AN}$ , tras la liberación de los correspondientes glicanos mediante digestión con PNGasa F, y las muestras de mTf purificadas del suero referencia se derivatizaron con  $[^{13}\text{C}_6]\text{AN}$ . Se prepararon mezclas equimolares de cada muestra WT o WT-CIA y de mTf referencia, y se analizaron por CapZIC-HILIC-MS. En este estudio se analizaron 4 muestras de suero de ratón WT (WT1, WT2, WT6 y WT7) y 4 WT con CIA (WT1-CIA, WT3-CIA, WT6-CIA y WT7-CIA), y se aplicó el mismo tratamiento quimiométrico utilizado con la hAGP, a partir de las relaciones de áreas de todos los isómeros de glicano detectados (área de cada isómero en la muestra WT dividido entre el área del mismo isómero detectado en la muestra de suero de referencia). En el modelo de PCA

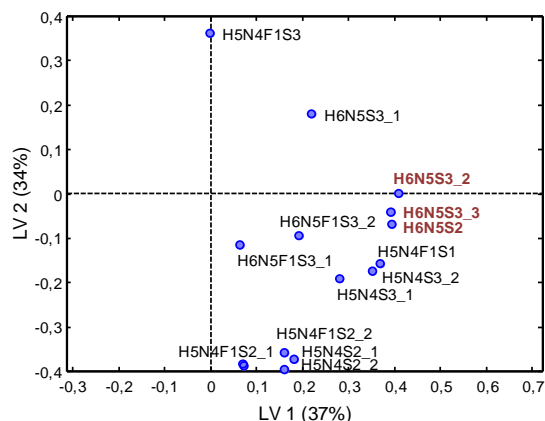
no se identificaron *outliers* y en el PLS-DA se usaron las mismas muestras clasificándolas en no inmunizadas (WT) o con CIA (WT-CIA). En la **Figura 5.27** se muestran los gráficos de scores (A), loadings (B) y VIP scores (C) obtenidos en este modelo. El gráfico de VIP scores muestra que sólo 5 de los 15 isómeros de glicano de la mTf son representativos (VIP>1) para distinguir entre ratones no inmunizados y con CIA (marcados en rojo en la **Figura 5.27C**). Tres de ellos corresponden a estructuras triantenarias (H6N5S2, H6N5S3\_2 y H6N5S3\_3), sólo uno presenta fucosa (H5N4F1S1) y la mayoría están altamente sialilados (H5N4S3\_2, H6N5S3\_2 y H6N5S3\_3). Estos resultados concuerdan con los obtenidos en el estudio previo a nivel de glicopéptidos [225], en el que se observó un aumento de las glicoformas triantenarias. A pesar de que no se ha llevado a cabo una caracterización previa de los isómeros de los glicanos de la mTf, éstos parecen contener un mayor número de SiA unidos  $\alpha$ 2-6 ya que se corresponden con los isómeros más retenidos en la columna ZIC-HILIC dentro de un mismo glicano (**artículos 3.1 y 3.2**).

Con el fin de validar el modelo de PLS-DA, se analizó una nueva muestra con CIA (WT4-CIA) y se procesaron los datos obtenidos siguiendo el mismo procedimiento explicado anteriormente. Como puede observarse en el gráfico de scores de la **Figura 5.27A**, la predicción de su clase fue excelente demostrando que el modelo puede diferenciar entre muestras de ratón no inmunizados y con CIA.

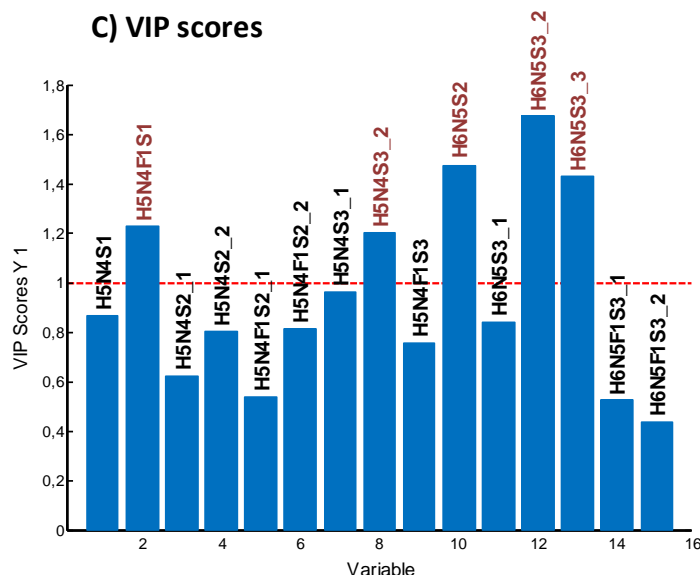
A) Gráfico de Scores



B) Gráfico de Loadings



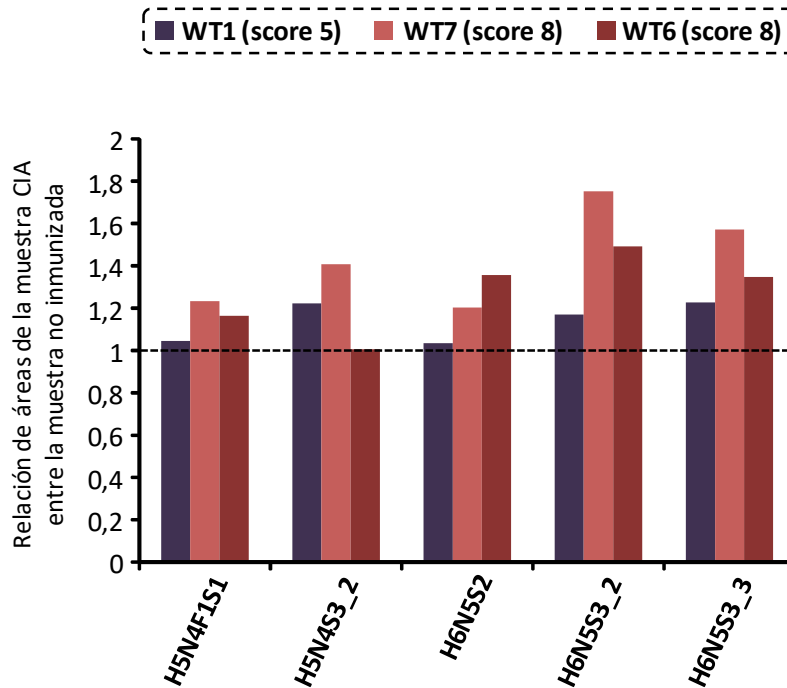
C) VIP scores



**Figura 5.27.-** (A) Gráfico de scores, (B) gráfico de loadings y (C) VIP scores del modelo de PLS-DA aplicado a las relaciones de áreas (área del isómero de glicano en la muestra no inmunizada o con CIA vs. área del mismo isómero en la muestra de suero de ratón de referencia obtenidos para cada isómero de glicano de la mTf en todas las muestras de ratón WT no inmunizadas y con CIA analizadas). Set de calibración: cuatro muestras no inmunizadas (WT1, WT2, WT6 y WT7) y cuatro con CIA (WT1-CIA, WT3-CIA, WT6-CIA y WT7-CIA). Muestra de validación: WT4-CIA.

Finalmente, dado que se analizaron algunas muestras de suero que provenían del mismo ratón antes y después de la inducción de CIA (WT1, WT6 y WT7), se decidió utilizar estas muestras para evaluar si la sobreexpresión de los isómeros, identificados por PLS-DA como importantes, aumentaba con el *score* clínico de la enfermedad. En **Figura 5.28** se muestra el gráfico de

barras de la relación de áreas para las muestras WT-CIA (WT-CIA / referencia) divididos entre la relación de áreas para las correspondientes muestras no inmunizadas (WT / referencia) en los 5 isómeros considerados más significativos por PLS-DA.



**Figura 5.28.-** Gráfico de barras de la relación de áreas obtenida para las muestras con CIA (Área CIA/Área referencia) dividida entre las relación de áreas obtenida para la correspondiente muestra no inmunizada (Área no inmunizada/Área referencia) en las muestras de ratón WT1 (score 5), WT6 (score 8) y WT7 (score 8).

Como se observa, todos los isómeros se encuentran más sobreexpresados en las muestras con un *score* clínico de 8 (WT6 y WT7) respecto a la muestra con un *score* clínico de 5 (WT5), a excepción del isómero H5N4S3\_2 en la muestra WT6. Estos resultados sugieren que la sobreexpresión de estos isómeros presenta relación con el *score* clínico de CIA, especialmente para los glicanos H6N5S2, H6N5S3\_2 y H6N5S3\_3. Por este motivo, aunque sería necesario analizar un número superior de muestras, estos isómeros podrían considerarse biomarcadores potenciales de CIA y utilizarse para evaluar la respuesta a nuevos tratamientos, así como para controlar la progresión de la patología.

### 5.3.3. Comparación de los resultados obtenidos en PDAC, ChrP y CIA

En los estudios realizados en esta tesis doctoral sobre las alteraciones de las estructuras de los glicanos y sus isómeros en procesos inflamatorios y cáncer, se han observado en líneas generales aumentos en la sialilación, la fucosilación y las ramificaciones, tal y como habían descrito anteriormente otros autores en estos tipos de patologías [22,23,29–31,34,38,43,224,230]. Sin embargo, es necesario abordar el estudio de la glicosilación de la glicoproteína de interés específicamente para cada patología para poder identificar candidatos biomarcadores.

En el caso de la CIA (proceso inflamatorio), se obtuvo un aumento de ciertas estructuras altamente sialiladas y ramificadas, considerándose el glicano H6N5S3 (isómeros 2 y 3) de la mTf posible biomarcador potencial de esta patología. En relación al PDAC, también se observó un aumento de las estructuras más sialiladas y ramificadas respecto a ChrP, pero solo para el isómero más retenido en la columna ZIC-HILIC dentro de un mismo glicano, es decir, aquel que presentaba más ácidos siálicos unidos  $\alpha$ 2-6. Por este motivo, el aumento de los ácidos siálicos  $\alpha$ 2-6 podría estar más relacionado con PDAC. También se observó un aumento de las estructuras fucosiladas en mayor medida que en ChrP, lo que demuestra que el aumento de la fucosilación guarda una relación más estrecha con la enfermedad PDAC. Teniendo en cuenta todos los resultados obtenidos, se ha seleccionado el glicano H6N5F2S3 de la hAGP como posible biomarcador de PDAC. Sin embargo, sería necesario llevar a cabo el análisis de más muestras patológicas en ambos estudios para poder confirmar el potencial de estas estructuras de glicano propuestas como biomarcadoras.

## Conclusions

---





The conclusions extracted from the work performed in this thesis are the following:

- ✓ Two purification methods (precipitation with acetone and TiO<sub>2</sub>-SPE-CE-MS) for the purification of glycopeptides from protein digests were developed.
  - **Precipitation with acetone** can be regarded as a simple and rapid method for the isolation of glycopeptides from protein digests.
    - The variables responsible of the different trend to precipitate of peptides, N- and O-glycopeptides in front of acetone were identified by **PLS-DA**.
    - It was demonstrated that the amount of acetone can be finely tuned for targeting the glycoforms of a certain O-glycopeptide.
    - An eight-fold volume of acetone was selected as the best conditions to precipitate the O<sub>126</sub>-glycopeptide of **recombinant human erythropoietin (rhEPO)**.
  - The on-line **TiO<sub>2</sub>-SPE-CE-MS** method allowed to selectively retain and enrich glycopeptides from protein digests.
    - The LODs obtained for glycopeptides were up to **100 times lower** than by conventional CE-MS.
    - It was demonstrated that the established method allowed preconcentrating glycopeptides with different compositions and regardless of the protein derived from.
    - Certain higher affinity of the TiO<sub>2</sub> sorbent for the more sialylated glycoforms, which would decrease with the size of the carbohydrate moiety, was observed but not in all the studied glycopeptides.
- ✓ Different strategies for the characterization of **alpha-1-acid glycoprotein (hAGP)** glycan isomers were developed.

- A **CapZIC-HILIC-MS** methodology has been optimized for the separation and characterization of glycan isomers.
  - Reliable assignments of sialic acid and fucose linkage-type isomers were achieved using **GRIL strategy** with [<sup>12</sup>C<sub>6</sub>]/[<sup>13</sup>C<sub>6</sub>]-aniline and specific **exoglycosidase digestions**.
  - Differences in the α2-3 and α2-6 sialic acid content were revealed that could be correlated with the changes in the abundance of certain isomers observed in previous works in pathological samples.
  - Characterization of fucose linkage-type isomers pointed out the presence of both α1-3 antennary and core fucosylation in most hAGP glycans.
  
- A **CapZIC-HILIC-MS/MS** methodology was established for the characterization of glycan isomers.
  - Structural feature characterizations as well as sialic acid and fucose linkage-type assignments of hAGP glycan isomers were performed.
  - Diagnostic ion fragments proposed in the literature were confirmed or discarded using both the exact mass of the daughter ions provided by the LTQ-Orbitrap mass spectrometer and the previous characterization of hAGP glycans by exoglycosidase digestions.
  - The established MS/MS method also enabled the location of α2-3 and α2-6 sialic acids on the antennas and the complete characterization of some tetraantennary glycans.
  - The ion fragment 297.1 was novelty established as an important diagnostic ion for fucose linkage-type assignment.
  - Robustness and future applicability of the MS/MS method for the study of control and pathological serum samples, was demonstrated.

- A **nanoESI-IM-MS** method has been developed and evaluated for the assignment of sialic acid linkage-types in hAGP glycan isomers.
  - Glycan isomer separation by IM-MS was only achieved in positive ion mode.
  - It was confirmed that **adduct** formation affects glycan isomer separation by IM-MS. The best separation for hAGP glycan isomers was obtained for  $[M+2H]^{2+}$  adduct.
  - It was demonstrated that the isomer separation obtained by IM-MS at the intact glycan level was not caused by differences in the sialic acid linkage-types.
  - Sialic acid linkage-type assignments were only successfully performed when using the glycan **fragment B<sub>3</sub>** (H1N1S1).
- ✓ The CapZIC-HILIC-MS methodology was successfully applied for the study of the alterations occurred in glycan isomers in relevant pathologies such as inflammatory processes and cancer. Reliable glycan relative quantification using the GRIL strategy with  $[^{12}C_6]/[^{13}C_6]$ -aniline was carried out in order to unequivocally identify variations in the glycoprofile of the studied protein. **PCA** and **PLS-DA** were used to properly identify the glycan isomer biomarker candidates of these diseases.
  - hAGP glycan isomers were observed to be altered in both **pancreatic ductal adenocarcinoma (PDAC)** and **chronic pancreatitis (ChrP)**.
    - Seven out forty-two hAGP intact glycan isomers, with high proportion of  $\alpha$ 2-6 sialic acids and most of them fucosylated, were considered the most meaningful to distinguish PDAC and ChrP patients.
    - An specific fucose linkage-type related only to PDAC, which would allow the differentiation from ChrP patients, was not identified in the analysis of desialylated hAGP glycans.

- H6N5F2S3 glycan could have a great biomarker potential as it was relevant in all PDAC and ChrP differentiations using PLS-DA for both intact and desialylated glycans.
- Alterations in **mouse transferrin (mTf)** glycan isomers were also identified in mice with **collagen-induced arthritis (CIA)**, an autoimmune inflammatory disease.
  - Purification of mTf from serum samples by IAC was firstly optimized to obtain a selective isolation of this glycoprotein.
  - Five out of fifteen mTf glycan isomers were identified by PLS-DA as important to distinguish CIA from control mice, most of them corresponding to highly branched and sialylated structures.
  - H6N5S2 and H6N5S3 seemed to be also related with the progression of CIA, being more overexpressed in samples with higher clinical scores.

The potential of this combined methodology in patho-glycomics to find novel **glycan-based biomarkers** was demonstrated.

## References

---



- [1] R.G. Spiro, Protein glycosylation: Nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds, *Glycobiology*. 12 (2002).  
<https://doi.org/10.1093/glycob/12.4.43R>.
- [2] S. Defaus, P. Gupta, D. Andreu, R. Gutiérrez-Gallego, Mammalian protein glycosylation - Structure versus function, *Analyst*. 139 (2014) 2944–2967.  
<https://doi.org/10.1039/c3an02245e>.
- [3] B.J. Arey, Chapter 12. The Role of Glycosylation in Receptor Signaling, in: *Glycosylation*, 2012: pp. 273–286. <https://doi.org/http://dx.doi.org/10.5772/57353>.
- [4] R. Fonseca-Maldonado, D.S. Vieira, J.S. Alponi, E. Bonneil, P. Thibault, R.J. Ward, Engineering the pattern of protein glycosylation modulates the thermostability of a GH11 xylanase, *J. Biol. Chem*. 288 (2013) 25522–25534.  
<https://doi.org/10.1074/jbc.M113.485953>.
- [5] N.G. Jayaprakash, A. Surolia, Role of glycosylation in nucleating protein folding and stability, *Biochem. J*. 474 (2017) 2333–2347. <https://doi.org/10.1042/BCJ20170111>.
- [6] B. Domon, C.E. Costello, A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS spectra of glycoconjugates, *Glycoconj. J*. 5 (1988) 397–409.  
<https://doi.org/10.1007/BF01049915>.
- [7] L. Krasnova, C.-H. Wong, *Understanding the Chemistry and Biology of Glycosylation with Glycan Synthesis*, 2016. <https://doi.org/10.1146/annurev-biochem-060614-034420>.
- [8] K.W. Moremen, M. Tiemeyer, A. V. Nairn, Vertebrate protein glycosylation: Diversity, synthesis and function, *Nat. Rev. Mol. Cell Biol*. 13 (2012) 448–462.  
<https://doi.org/10.1038/nrm3383>.
- [9] Consortium for functional glycomics. Functional glycomics gateway, (n.d.).  
<http://www.functinalglycomics.org/>.
- [10] A. Varki, N. Sharon, Chapter 1. Historical Background and Overview, in: A. Varki, R.D.



- Cummings, J.D. Esko, H.H. Freeze, P. Stanley, C.R. Bertozzi, G.W. Hart, M.E. Etzler (Eds.), Essentials Glycobiol., Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press, 2009.
- [11] A. Varki, H.H. Freeze, A.E. Manzi, Unit 12.1. Overview of Glycoconjugate Analysis, in: J.E. Coligan, B.M. Dunn, D.W. Speicher, P.T. Wingfield (Eds.), Curr. Protoc. Protein Sci., Wiley, 2009.
- [12] S.B. Lavery, C. Steentoft, A. Halim, Y. Narimatsu, H. Clausen, S.Y. Vakhrushev, Advances in mass spectrometry driven O-glycoproteomics, Biochim. Biophys. Acta. 1850 (2015) 33–42. <https://doi.org/10.1016/j.bbagen.2014.09.026>.
- [13] A. Varki, R. Cummings, J. Esko, H. Freeze, G. Hart, J. Marth, eds., Chapter 8. O-Glycans, in: Essentials Glycobiol., Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press, 1999.
- [14] E. Bieberich, Synthesis, processing, and function of N-glycans in N-glycoproteins, Adv Neurobiol. 9 (2014) 47–70. [https://doi.org/10.1007/978-1-4939-1154-7\\_3](https://doi.org/10.1007/978-1-4939-1154-7_3).
- [15] T.N. Stanley P, Schachter H, Chapter 8. N-Glycans, in: A. Varki, R. Cummings, J. Esko, H. Freeze, G. Hart, J. Marth (Eds.), Essentials Glycobiol., 2nd ed., Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press, 2009.
- [16] A. Varki, R. Cummings, J. Esko, H. Freeze, G. Hart, J. Marth, eds., Chapter 16. Structures Common to Different Types of Glycans., in: Essentials Glycobiol., Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press, 1999.
- [17] A. Varki, Sialic acids in human health and disease, Trends Mol Med. 14 (2009) 351–360. <https://doi.org/10.1016/j.molmed.2008.06.002>.
- [18] B. Byrne, G.G. Donohoe, R. O’Kennedy, Sialic acids: carbohydrate moieties that influence the biological and physical properties of biopharmaceutical proteins and living cells, Drug Discov. Today. 12 (2007) 319–326. <https://doi.org/10.1016/j.drudis.2007.02.010>.

- [19] N.M. Varki, A. Varki, Diversity in cell surface sialic acid presentations: Implications for biology and disease, *Lab. Investig.* 87 (2007) 851–857.  
<https://doi.org/10.1038/labinvest.3700656>.
- [20] A. Varki, R. Schauer, Chapter 14. Sialic Acids, in: A. Varki, R.D. Cummings, J.D. Esko, H.H. Freeze, P. Stanley, C.R. Bertozzi, G.W. Hart, M.E. Etzler (Eds.), *Essentials Glycobiol.*, Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press, 2009.
- [21] L. Veillon, Y. Huang, W. Peng, X. Dong, B.G. Cho, Y. Mechref, Characterization of Isomeric Glycan Structures by LC-MS/MS, *Electrophoresis*. 38 (2017) 2100–2114.  
<https://doi.org/10.1002/elps.201700042>.
- [22] S.S. Pinho, C.A. Reis, Glycosylation in cancer: mechanisms and clinical implications, *Nat Rev Cancer*. 15 (2015) 540–555. <https://doi.org/10.1038/nrc3982>.
- [23] O. Gornik, G. Lauc, Glycosylation of serum proteins in inflammatory diseases, *Dis. Markers*. 25 (2008) 267–278.
- [24] 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.  
<https://doi.org/10.1016/j.talanta.2016.07.055>.
- [25] D. Thompson, A. Milford-Ward, J.T. Whicher, The value of acute phase protein measurements in clinical practice, *Ann. Clin. Biochem.* 29 (1992) 123–131.  
<https://doi.org/10.1177/000456329202900201>.
- [26] S. Jain, V. Gautam, S. Naseem, Acute-phase proteins: As diagnostic tool, *J Pharm Bioallied Sci.* 3 (2011) 118–127. <https://doi.org/10.4103/0975-7406.76489>.
- [27] P. Zhang, S. Woen, T. Wang, B. Liao, S. Zhao, C. Chen, Y. Yang, Z. Song, M.R. Wormald, C. Yu, P.M. Rudd, Challenges of glycosylation analysis and control: An integrated approach to producing optimal and consistent therapeutic drugs, *Drug Discov. Today*. 21 (2016) 740–765. <https://doi.org/10.1016/j.drudis.2016.01.006>.

- [28] R. Jefferis, Review of Glycosylation Engineering of Biopharmaceuticals: Methods and Protocols, *Landes Biosci.* 5 (2013) 638–640. <https://doi.org/10.4161/mabs.25631>.
- [29] S.R. Stowell, T. Ju, R.D. Cummings, Protein Glycosylation in Cancer, *Annu. Rev. Pathol. Mech. Dis.* 10 (2015) 473–510. <https://doi.org/10.1146/annurev-pathol-012414-040438>.
- [30] B.N. Vajaria, P.S. Patel, Glycosylation: a hallmark of cancer?, *Glycoconj. J.* 34 (2017) 147–156. <https://doi.org/10.1007/s10719-016-9755-2>.
- [31] L. Oliveira-Ferrer, K. Legler, K. Milde-Langosch, Role of protein glycosylation in cancer metastasis, *Semin. Cancer Biol.* 44 (2017) 141–152. <https://doi.org/10.1016/j.semcancer.2017.03.002>.
- [32] J.G. Rodrigues, M. Balmaña, J.A. Macedo, J. Poças, Â. Fernandes, J.C.M. de-Freitas-Junior, S.S. Pinho, J. Gomes, A. Magalhães, C. Gomes, S. Mereiter, C.A. Reis, Glycosylation in cancer: Selected roles in tumour progression, immune modulation and metastasis, *Cell. Immunol.* 333 (2018) 46–57. <https://doi.org/10.1016/j.cellimm.2018.03.007>.
- [33] J.J. Park, M. Lee, Increasing the  $\alpha$  2, 6 sialylation of glycoproteins may contribute to metastatic spread and therapeutic resistance in colorectal cancer, *Gut Liver.* 7 (2013) 629–641. <https://doi.org/10.5009/gnl.2013.7.6.629>.
- [34] R. Garnham, E. Scott, K. Livermore, J. Munkley, ST6GAL1: A key player in cancer (Review), *Oncol. Lett.* (2019) 983–989. <https://doi.org/10.3892/ol.2019.10458>.
- [35] S. Yazawa, R. Takahashi, T. Yokobori, R. Sano, A. Mogi, A.R. Saniabadi, H. Kuwano, T. Asao, Fucosylated glycans in  $\alpha$ 1-acid glycoprotein for monitoring treatment outcomes and prognosis of cancer patients, *PLoS One.* 11 (2016). <https://doi.org/10.1371/journal.pone.0156277>.
- [36] E. Miyoshi, K. Moriwaki, N. Terao, C.C. Tan, M. Terao, T. Nakagawa, H. Matsumoto, S. Shinzaki, Y. Kamada, Fucosylation is a promising target for cancer diagnosis and

- therapy, *Biomolecules*. 2 (2012) 34–45. <https://doi.org/10.3390/biom2010034>.
- [37] O.M.T. Pearce, Cancer glycan epitopes: Biosynthesis, structure and function, *Glycobiology*. 28 (2018) 670–696. <https://doi.org/10.1093/glycob/cwy023>.
- [38] A. Blanas, N.M. Sahasrabudhe, E. Rodríguez, Y. van Kooyk, S.J. van Vliet, Fucosylated antigens in cancer: An alliance toward tumor progression, metastasis, and resistance to chemotherapy, *Front. Oncol.* 8 (2018) 1–14. <https://doi.org/10.3389/fonc.2018.00039>.
- [39] K. Krishna, T. Bekaii-Saab, CA 19-9 as a Serum Biomarker in Cancer, in: V. Preedy, V. Patel (Eds.), *Biomarkers Cancer. Biomarkers Dis. Methods, Discov. Appl.*, Springer, 2015.
- [40] S. Scarà, P. Bottoni, R. Scatena, CA 19-9: Biochemical and Clinical Aspects, in: R. Scatena (Ed.), *Adv. Cancer Biomarkers. Adv. Exp. Med. Biol.*, Springer, 2015.
- [41] U.K. Ballehaninna, R.S. Chamberlain, The clinical utility of serum CA 19-9 in the diagnosis, prognosis and management of pancreatic adenocarcinoma: An evidence based appraisal, *J. Gastrointest. Oncol.* 3 (2012) 105–119. <https://doi.org/10.3978/j.issn.2078-6891.2011.021>.
- [42] U.K. Ballehaninna, R.S. Chamberlain, Serum CA 19-9 as a Biomarker for Pancreatic Cancer-A Comprehensive Review, *Indian J. Surg. Oncol.* 2 (2011) 88–100. <https://doi.org/10.1007/s13193-011-0042-1>.
- [43] J. Munkley, The glycosylation landscape of pancreatic cancer (Review), *Oncol. Lett.* 17 (2019) 2569–2575. <https://doi.org/10.3892/ol.2019.9885>.
- [44] E. Llop, P.E. Guerrero, A. Duran, S. Barrabés, A. Massaguer, M.J. Ferri, M. Albiol-Quer, R. De Llorens, R. Peracaula, Glycoprotein biomarkers for the detection of pancreatic ductal adenocarcinoma, *World J. Gastroenterol.* 24 (2018) 2537–2554. <https://doi.org/10.3748/wjg.v24.i24.2537>.
- [45] T. Hochepped, F.G. Berger, H. Baumann, C. Libert,  $\alpha$ 1-acid glycoprotein: An acute phase protein with inflammatory and immunomodulating properties, *Cytokine Growth Factor Rev.* 14 (2003) 25–34. [https://doi.org/10.1016/S1359-6101\(02\)00054-0](https://doi.org/10.1016/S1359-6101(02)00054-0).

- [46] T. Fournier, N. Medjoubi, D. Porquet, Alpha-1-acid glycoprotein, *Biochim.Biophys.Acta.* 1482 (2000) 157–171. [https://doi.org/10.1016/s0167-4838\(00\)00153-9](https://doi.org/10.1016/s0167-4838(00)00153-9).
- [47] K.N. Kazuaki Taguchi, T.M. and M.O. Victor Tuan Giam Chuang, Chapter 6. Molecular Aspects of Human Alpha-1 Acid Glycoprotein — Structure and Function, in: *Acute Phase Proteins Sev.*, 2013: pp. 139–162. <https://doi.org/10.5772/56101>.
- [48] C.L. Fernandes, R. Ligabue-Braun, H. Verli, Structural glycobiology of human  $\alpha$ 1-acid glycoprotein and its implications for pharmacokinetics and inflammation, *Glycobiology.* 25 (2015) 1125–1133. <https://doi.org/10.1093/glycob/cwv041>.
- [49] R.B. Rodríguez, The Relevance of Alpha-1-Acid Glycoprotein in Human Cancer: A Minireview, *Adv. Cancer Res. Clin. Imaging.* 2 (2019) 10–11. <https://doi.org/10.33552/acrci.2019.02.000526>.
- [50] Q. Zhou, R. Andersson, D. Hu, M. Bauden, A. Sasor, T. Bygott, K. Pawłowski, I. Pla\_Parada, G. Marko-Varga, D. Ansari, Alpha-1-acid glycoprotein 1 (AGP1) as a novel biomarker for pancreatic cancer, *J. Clin. Oncol.* 37 (2019). [https://doi.org/10.1200/JCO.2019.37.15\\_suppl.e15708](https://doi.org/10.1200/JCO.2019.37.15_suppl.e15708).
- [51] J. Liang, J. Zhu, M. Wang, A.G. Singal, M. Odewole, S. Kagan, V. Renteria, S. Liu, N.D. Parikh, D.M. Lubman, Evaluation of AGP Fucosylation as a Marker for Hepatocellular Carcinoma of Three Different Etiologies, *Sci. Rep.* 9 (2019) 1–13. <https://doi.org/10.1038/s41598-019-48043-1>.
- [52] K.D. Smith, J. Behan, G. Matthews-Smith, A.M. Magliocco, Chapter 9. Alpha-1-Acid Glycoprotein (AGP) as a Potential Biomarker for Breast Cancer, in: *Glycosylation, 2012:* pp. 201–222. <https://doi.org/http://dx.doi.org/10.5772/57353>.
- [53] F. Cecilian, V. Pocacqua, The acute phase protein alpha 1-acid glycoprotein: a model for altered glycosylation during diseases., *Curr. Protein Pept. Sci.* 8 (2007) 91–108. <https://doi.org/10.2174/138920307779941497>.
- [54] E. Giménez, M. Balmaña, J. Figueras, E. Fort, C. de Bolós, V. Sanz-Nebot, R. Peracaula, A.

- Rizzi, Quantitative analysis of N-glycans from human alfa-acid-glycoprotein using stable isotope labeling and zwitterionic hydrophilic interaction capillary liquid chromatography electrospray mass spectrometry as tool for pancreatic disease diagnosis, *Anal. Chim. Acta.* 866 (2015) 59–68.  
<https://doi.org/10.1016/j.aca.2015.02.008>.
- [55] M. Balmaña, E. Giménez, A. Puerta, E. Llop, J. Figueras, E. Fort, V. Sanz-Nebot, C. de Bolós, A. Rizzi, S. Barrabés, M. de Frutos, R. Peracaula, Increased  $\alpha$ 1-3 fucosylation of  $\alpha$ -1-acid glycoprotein (AGP) in pancreatic cancer, *J. Proteomics.* 132 (2016) 144–154.  
<https://doi.org/10.1016/j.jprot.2015.11.006>.
- [56] J. Ledingham, N. Snowden, Z. Ide, Diagnosis and early management of inflammatory arthritis, *BMJ.* 358 (2017) 1–8. <https://doi.org/10.1136/bmj.j3248>.
- [57] Q. Guo, Y. Wang, D. Xu, J. Nossent, N.J. Pavlos, J. Xu, Rheumatoid arthritis: Pathological mechanisms and modern pharmacologic therapies, *Nature.* 6 (2018).  
<https://doi.org/10.1038/s41413-018-0016-9>.
- [58] D. Aletaha, J.S. Smolen, Diagnosis and Management of Rheumatoid Arthritis: A Review, *JAMA - J. Am. Med. Assoc.* 320 (2018) 1360–1372.  
<https://doi.org/10.1001/jama.2018.13103>.
- [59] D.T. Felson, J.S. Smolen, G. Wells, B. Zhang, L.H.D. Van Tuyl, J. Funovits, D. Aletaha, C.F. Allaart, J. Bathon, S. Bombardieri, P. Brooks, A. Brown, M. Matucci-Cerinic, H. Choi, B. Combe, M. De Wit, M. Dougados, P. Emery, D. Furst, J. Gomez-Reino, G. Hawker, E. Keystone, D. Khanna, J. Kirwan, T.K. Kvien, R. Landewé, J. Listing, K. Michaud, E. Martin-Mola, P. Montie, T. Pincus, P. Richards, J.N. Siegel, L.S. Simon, T. Sokka, V. Strand, P. Tugwell, A. Tyndall, D. Van Der Heijde, S. Verstappen, B. White, F. Wolfe, A. Zink, M. Boers, American college of rheumatology/European league against rheumatism provisional definition of remission in rheumatoid arthritis for clinical trials, *Ann Rheum Dis.* 70 (2011) 404–413. <https://doi.org/10.1136/ard.2011.149765>.

- [60] A. Sebastian, M.A. Alzain, C.O. Asweto, H. Song, L. Cui, X. Yu, S. Ge, H. Dong, P. Rao, H. Wang, H. Fang, Q. Gao, J. Zhang, D. He, X. Guo, M. Song, Y. Wang, W. Wang, Glycan Biomarkers for Rheumatoid Arthritis and Its Remission Status in Han Chinese Patients, *Omi. A J. Integr. Biol.* 20 (2016) 343–351. <https://doi.org/10.1089/omi.2016.0050>.
- [61] X. Li, L. Ding, X. Li, H. Zhu, E.A. Gashash, Z. Li, P.G. Wang, C. Ma, An integrated proteomic and glycoproteomic study for differences on glycosylation occupancy in rheumatoid arthritis, *Anal. Bioanal. Chem.* 411 (2019) 1331–1338. <https://doi.org/10.1007/s00216-018-1543-3>.
- [62] K.R. Reiding, G.C.M. Vreeker, A. Bondt, M.R. Bladergroen, J.M.W. Hazes, Y.E.M. van der Burgt, M. Wuhler, R.J.E.M. Dolhain, Serum protein N-glycosylation changes with rheumatoid arthritis disease activity during and after pregnancy, *Front. Med.* 4 (2017). <https://doi.org/10.3389/fmed.2017.00241>.
- [63] C. Thiel, C. Körner, Mouse models for congenital disorders of glycosylation, *J. Inherit. Metab. Dis.* 34 (2011) 879–889. <https://doi.org/10.1007/s10545-011-9295-7>.
- [64] R.L. Perlman, Mouse Models of Human Disease: An Evolutionary Perspective, *Evol. Med. Public Heal.* (2016) 170–176. <https://doi.org/10.1093/emph/eow014>.
- [65] D.D. Brand, K.A. Latham, E.F. Rosloniec, Collagen-induced arthritis, *Nat. Protoc.* 2 (2007) 1269–1275. <https://doi.org/10.1038/nprot.2007.173>.
- [66] J.J. Inglis, G. Criado, M. Medghalchi, M. Andrews, A. Sandison, M. Feldmann, R.O. Williams, Collagen-induced arthritis in C57BL/6 mice is associated with a robust and sustained T-cell response to type II collagen, *Arthritis Res. Ther.* 9 (2007) 1–8. <https://doi.org/10.1186/ar2319>.
- [67] W. Schrödl, R. Büchler, S. Wendler, P. Reinhold, P. Muckova, J. Reindl, H. Rhode, Acute phase proteins as promising biomarkers: Perspectives and limitations for human and veterinary medicine, *Proteomics - Clin. Appl.* 10 (2016) 1077–1092. <https://doi.org/10.1002/prca.201600028>.

- [68] A. Barroso, E. Giménez, F. Benavente, J. Barbosa, V. Sanz-Nebot, Analysis of human transferrin glycopeptides by capillary electrophoresis and capillary liquid chromatography-mass spectrometry. Application to diagnosis of alcohol dependence, *Anal. Chim. Acta.* 804 (2013) 167–175. <https://doi.org/10.1016/j.aca.2013.09.044>.
- [69] M. Gudowska, E. Gruszewska, A. Wrona, E. Gindzienska-Sieskiewicz, I. Domyslawska, K. Lipartowska-Klimuk, B. Cylwik, S. Sierakowski, L. Chrostek, The Profile of Serum Transferrin Isoforms in Rheumatoid Arthritis, *J. Clin. Rheumatol.* 25 (2019) 159–162. <https://doi.org/10.1097/RHU.0000000000000808>.
- [70] A. Rosal-Vela, A. Barroso, E. Giménez, S. García-Rodríguez, V. Longobardo, J. Postigo, M. Iglesias, A. Lario, J. Merino, R. Merino, M. Zubiaur, V. Sanz-Nebot, J. Sancho, Identification of multiple transferrin species in the spleen and serum from mice with collagen-induced arthritis which may reflect changes in transferrin glycosylation associated with disease activity: The role of CD38, *J. Proteomics.* 134 (2016) 127–137. <https://doi.org/10.1016/j.jprot.2015.11.023>.
- [71] M.E. Lalonde, Y. Durocher, Therapeutic glycoprotein production in mammalian cells, *J. Biotechnol.* 251 (2017) 128–140. <https://doi.org/10.1016/j.jbiotec.2017.04.028>.
- [72] A.G. Vulto, O.A. Jaquez, The process defines the product: what really matters in biosimilar design and production?, *Rheumatology (Oxford).* 56 (2017) 14–29. <https://doi.org/10.1093/rheumatology/kex278>.
- [73] F. Higel, T. Sandl, C.Y. Kao, N. Pechinger, F. Sörgel, W. Friess, F. Wolschin, A. Seidl, N-glycans of complex glycosylated biopharmaceuticals and their impact on protein clearance, *Eur. J. Pharm. Biopharm.* 139 (2019) 123–131. <https://doi.org/10.1016/j.ejpb.2019.03.018>.
- [74] A. Eon-Duval, H. Broly, R. Gleixner, Quality attributes of recombinant therapeutic proteins: An assessment of impact on safety and efficacy as part of a quality by design development approach, *Biotechnol. Prog.* 28 (2012) 608–622.



- <https://doi.org/10.1002/btpr.1548>.
- [75] V.R. Gómez-Román, J.C. Murray, L.M. Weiner, Chapter 1 - Antibody-Dependent Cellular Cytotoxicity (ADCC), in: *Antib. Fc*, 2014: pp. 1–27.
- [76] M. Fukuda, H. Sasaki, L. Lopez, M. Fukuda, Survival of recombinant erythropoietin in the circulation: the role of carbohydrates, *Blood*. 73 (1989) 84–89.
- [77] R. Lange, V. Pavlovic-Kentera, Erythropoietin, *Prog Hematol*. 4 (1964) 72–96.
- [78] S. Krantz, Erythropoietin, *Blood*. 77 (1991) 419–434.
- [79] W. Jelkmann, Physiology and pharmacology of erythropoietin, *Transfus. Med. Hemotherapy*. 40 (2013) 302–309. <https://doi.org/10.1159/000356193>.
- [80] A.S. Ramos, C.A. Schmidt, S.S. Andrade, M. Fronza, B. Rafferty, S.L. Dalmora, Biological evaluation of recombinant human erythropoietin in pharmaceutical products, *Brazilian J. Med. Biol. Res*. 36 (2003) 1561–1569. <https://doi.org/10.1590/S0100-879X2003001100014>.
- [81] J.S. Lee, T.K. Ha, S.J. Lee, G.M. Lee, Current state and perspectives on erythropoietin production, *Appl. Microbiol. Biotechnol*. 95 (2012) 1405–1416. <https://doi.org/10.1007/s00253-012-4291-x>.
- [82] W. Jelkmann, Use of Recombinant Human Erythropoietin as an Antianemic and Performance Enhancing Drug, *Curr. Pharm. Biotechnol*. 1 (2005) 11–31. <https://doi.org/10.2174/1389201003379068>.
- [83] M. Citartan, S.C.B. Gopinath, Y. Chen, T. Lakshmipriya, T.H. Tang, Monitoring recombinant human erythropoietin abuse among athletes, *Biosens. Bioelectron*. 63 (2015) 86–98. <https://doi.org/10.1016/j.bios.2014.06.068>.
- [84] J.O.M. Plumb, J.M. Otto, M.P.W. Grocott, “Blood doping” from Armstrong to prehabilitation: Manipulation of blood to improve performance in athletes and physiological reserve in patients, *Extrem. Physiol. Med*. 5 (2016) 1–11. <https://doi.org/10.1186/s13728-016-0046-0>.

- [85] 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. <https://doi.org/10.1016/j.aca.2011.10.028>.
- [86] E. Giménez, R. Ramos-Hernan, F. Benavente, J. Barbosa, V. Sanz-Nebot, Capillary electrophoresis time-of-flight mass spectrometry for a confident elucidation of a glycopeptide map of recombinant human erythropoietin, *Rapid Commun. Mass Spectrom.* 25 (2011) 2307–2316. <https://doi.org/10.1002/rcm.5114>.
- [87] E. Giménez, F. Benavente, J. Barbosa, V. Sanz-Nebot, Analysis of intact erythropoietin and novel erythropoiesis-stimulating protein by capillary electrophoresis-electrospray-ion trap mass spectrometry, *Electrophoresis.* 29 (2008) 2161–2170. <https://doi.org/10.1002/elps.200700788>.
- [88] B.Q. Tran, C. Barton, J. Feng, A. Sandjong, S.H. Yoon, S. Awasthi, T. Liang, M.M. Khan, D.P.A. Kilgour, D.R. Goodlett, Y.A. Goo, Comprehensive glycosylation profiling of IgG and IgG-fusion proteins by top-down MS with multiple fragmentation techniques, *J. Proteomics.* 134 (2016) 93–101. <https://doi.org/10.1016/j.jprot.2015.10.021>.
- [89] F.G. Hanisch, Top-down sequencing of O-glycoproteins by in-source decay matrix-assisted laser desorption ionization mass spectrometry for glycosylation site analysis, *Anal. Chem.* 83 (2011) 4829–4837. <https://doi.org/10.1021/ac200493c>.
- [90] H. Geyer, R. Geyer, Strategies for analysis of glycoprotein glycosylation, *Biochim. Biophys. Acta - Proteins Proteomics.* 1764 (2006) 1853–1869. <https://doi.org/10.1016/j.bbapap.2006.10.007>.
- [91] C. Neusüß, U. Demelbauer, M. Pelzing, Glycoform characterization of intact erythropoietin by capillary electrophoresis-electrospray-time of flight-mass spectrometry, *Electrophoresis.* 26 (2005) 1442–1450. <https://doi.org/10.1002/elps.200410269>.

- [92] H. Desaire, Glycopeptide analysis, recent developments and applications, *Mol. Cell. Proteomics*. 12 (2013) 893–901. <https://doi.org/10.1074/mcp.R112.026567>.
- [93] and D.M.Y. Adam Moser, Kevin Range, Comprehensive Native Glycan Profiling with Isomer Separation and Quantitation for the Discovery of Cancer Biomarkers, *Analyst*. 136 (2011) 3663–3671. <https://doi.org/10.1039/c1an15093f>.
- [94] M. V Novotny, W.R. Alley, Isolation and Purification of Glycans from Natural Sources for Positive Identification, *Discov. Subtleties Sugars*. (2014) 133–148.
- [95] M. Hu, Y. Lan, A. Lu, X. Ma, L. Zhang, Glycan-based biomarkers for diagnosis of cancers and other diseases: Past, present, and future, 1st ed., Elsevier Inc., 2019. <https://doi.org/10.1016/bs.pmbts.2018.12.002>.
- [96] 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. <https://doi.org/10.1039/c7an01262d>.
- [97] L.R. Ruhaak, G. Zauner, C. Huhn, C. Bruggink, A.M. Deelder, M. Wuhrer, Glycan labeling strategies and their use in identification and quantification, *Anal. Bioanal. Chem*. 397 (2010) 3457–3481. <https://doi.org/10.1007/s00216-010-3532-z>.
- [98] B. Xia, C.L. Feasley, G.P. Sachdev, D.F. Smith, R.D. Cummings, Glycan reductive isotope labeling (GRIL) for quantitative glycomics, *Anal Biochem*. 387 (2009) 162–170. <https://doi.org/10.1016/j.ab.2009.01.028>.
- [99] 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. <https://doi.org/10.1007/s00216-013-7178-5>.
- [100] X. Xu, T.D. Veenstra, Analysis of biofluids for biomarker research, *Proteomics - Clin. Appl*. 2 (2008) 1403–1412. <https://doi.org/10.1002/prca.200780173>.
- [101] T.D. Veenstra, T.P. Conrads, B.L. Hood, A.M. Avellino, R.G. Ellenbogen, R.S. Morrison,

- Biomarkers: Mining the biofluid proteome, *Mol. Cell. Proteomics*. 4 (2005) 409–418.  
<https://doi.org/10.1074/mcp.M500006-MCP200>.
- [102] B.C. Carlyle, B.A. Trombetta, S.E. Arnold, Proteomic approaches for the discovery of biofluid biomarkers of neurodegenerative dementias, *Proteomes*. 6 (2018).  
<https://doi.org/10.3390/proteomes6030032>.
- [103] C. Wu, J. Duan, T. Liu, R.D. Smith, W.-J. Qian, Contributions of Immunoaffinity Chromatography to Deep Proteome Profiling of Human Biofluids, *J. Chromatogr. B*. 1021 (2016) 57–68. <https://doi.org/10.1016/j.jchromb.2016.01.015>.
- [104] N.E. Thompson, R.R. Burgess, Immunoaffinity Chromatography: Advantages and Limitations, (2015) 483–502.
- [105] A.C. Moser, D.S. Hage, Immunoaffinity chromatography: an introduction to applications and recent developments, *Bioanalysis*. 2 (2010) 769–790.  
<https://doi.org/10.4155/bio.10.31>.
- [106] S. Ongay, A. Boichenko, N. Govorukhina, R. Bischoff, Glycopeptide enrichment and separation for protein glycosylation analysis, *J. Sep. Sci.* 35 (2012) 2341–2372.  
<https://doi.org/10.1002/jssc.201200434>.
- [107] S. Ito, K. Hayama, J. Hirabayashi, Enrichment Strategies for Glycopeptides., in: N.H. Packer, N.G. Karlsson (Eds.), *Glycomics. Methods Mol. Biol. (Methods Protoc., Humana Press, 2009*.
- [108] A. Monzo, G.K. Bonn, A. Guttman, Lectin-immobilization strategies for affinity purification and separation of glycoconjugates, *TrAC - Trends Anal. Chem.* 26 (2007) 423–432. <https://doi.org/10.1016/j.trac.2007.01.018>.
- [109] K.S. Nascimento, A.I. Cunha, K.S. Nascimento, B.S. Cavada, A.M. Azevedo, M.R. Aires-Barros, An overview of lectins purification strategies, *J. Mol. Recognit.* 25 (2012) 527–541. <https://doi.org/10.1002/jmr.2200>.
- [110] Š. Belický, J. Katrlík, J. Tkáč, Glycan and lectin biosensors, *Essays Biochem.* 60 (2016) 37–

47. <https://doi.org/10.1042/EBC20150005>.
- [111] M. Madera, B. Mann, Y. Mechref, M. V. Novotny, Efficacy of Glycoprotein Enrichment by Microscale Lectin Affinity Chromatography, *J Sep Sci.* 31 (2008) 2722–2732. <https://doi.org/10.1002/jssc.200800094>.
- [112] 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. <https://doi.org/10.1002/elps.201400034>.
- [113] T.E. Thingholm, M.R. Larsen, Chapter 9. The use of Titanium Dioxide for selective enrichment of phosphorylated peptides, in: L. von Stechow (Ed.), *Phospho-Proteomics Methods Protoc. Methods Mol. Biol.*, Springer, 2016: pp. 135–46. [https://doi.org/10.1007/978-1-4939-3049-4\\_9](https://doi.org/10.1007/978-1-4939-3049-4_9).
- [114] M. Mazanek, G. Mituloviae, F. Herzog, C. Stingl, J.R. Hutchins, J.M. Peters, K. Mechtler, Titanium dioxide as a chemo-affinity solid phase in offline phosphopeptide chromatography prior to HPLC-MS/MS analysis., *Nat. Protoc.* 2 (2007) 1059–1069. <https://doi.org/10.1038/nprot.2006.280>.
- [115] C. Zhang, Z. Ye, P. Xue, Q. Shu, Y. Zhou, Y. Ji, Y. Fu, J. Wang, F. Yang, Evaluation of Different N-Glycopeptide Enrichment Methods for N-Glycosylation Sites Mapping in Mouse Brain, *J. Proteome Res.* 15 (2016) 2960–2968. <https://doi.org/10.1021/acs.jproteome.6b00098>.
- [116] G. Palmisano, S.E. Lendal, K. Engholm-Keller, R. Leth-Larsen, B.L. Parker, M.R. Larsen, Selective enrichment of sialic acid-containing glycopeptides using titanium dioxide chromatography with analysis by HILIC and mass spectrometry, *Nat. Protoc.* 5 (2010) 1974–1982. <https://doi.org/10.1038/nprot.2010.167>.
- [117] Q. Sheng, X. Li, W. Yin, L. Yu, Y. Ke, X. Liang, Retention mechanism and enrichment of glycopeptides on titanium dioxide, *Anal. Methods.* 5 (2013) 7072–7080. <https://doi.org/10.1039/c3ay41294f>.

- [118] D. Takakura, A. Harazono, N. Hashii, N. Kawasaki, Selective glycopeptide profiling by acetone enrichment and LC/MS, *J. Proteomics*. 101 (2014) 17–30.  
<https://doi.org/10.1016/j.jprot.2014.02.005>.
- [119] S.M. Dahimiwal, D.B. Thorat, N.P. Jain, V.B. Jadhav, P.B. Patil, A review on high performance liquid chromatography, *Int. J. Pharm. Res.* 5 (2013) 1–6.  
<https://doi.org/10.22214/ijraset.2018.2098>.
- [120] R. Malviya, V. Bansal, O.P. Pal, P.K. Sharma, High performance liquid chromatography: A short review, *J. Glob. Pharma Technol.* 2 (2010) 22–26.
- [121] A. Shubhakar, K.R. Reiding, R.A. Gardner, D.I.R. Spencer, D.L. Fernandes, M. Wuhrer, High-Throughput Analysis and Automation for Glycomics Studies, *Chromatographia*. 78 (2015) 321–333. <https://doi.org/10.1007/s10337-014-2803-9>.
- [122] H. Engelhardt, ed., *Practice of High Performance Liquid Chromatography. Applications, Equipment and Quantitative Analysis*, Springer, 1986.
- [123] K. Stavenhagen, R. Plomp, M. Wuhrer, Site-Specific Protein N- and O-Glycosylation Analysis by a C18-Porous Graphitized Carbon-Liquid Chromatography-Electrospray Ionization Mass Spectrometry Approach Using Pronase Treated Glycopeptides, *Anal. Chem.* 87 (2015) 11691–11699. <https://doi.org/10.1021/acs.analchem.5b02366>.
- [124] M. Wuhrer, A.M. Deelder, C.H. Hokke, Protein glycosylation analysis by liquid chromatography-mass spectrometry, *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* 825 (2005) 124–133. <https://doi.org/10.1016/j.jchromb.2005.01.030>.
- [125] B. Buszewski, S. Noga, Hydrophilic interaction liquid chromatography (HILIC)-a powerful separation technique, *Anal. Bioanal. Chem.* 402 (2012) 231–247.  
<https://doi.org/10.1007/s00216-011-5308-5>.
- [126] G. Zauner, A.M. Deelder, M. Wuhrer, Recent advances in hydrophilic interaction liquid chromatography (HILIC) for structural glycomics, *Electrophoresis*. 32 (2011) 3456–3466.  
<https://doi.org/10.1002/elps.201100247>.

- [127] T. Ikegami, Hydrophilic interaction chromatography for the analysis of biopharmaceutical drugs and therapeutic peptides: A review based on the separation characteristics of the hydrophilic interaction chromatography phases, *J. Sep. Sci.* 42 (2019) 130–213. <https://doi.org/10.1002/jssc.201801074>.
- [128] S. Di Palma, P.J. Boersema, A.J.R. Heck, S. Mohammed, Zwitterionic hydrophilic interaction liquid chromatography (ZIC-HILIC and ZIC-cHILIC) provide high resolution separation and increase sensitivity in proteome analysis, *Anal. Chem.* 83 (2011) 3440–3447. <https://doi.org/10.1021/ac103312e>.
- [129] Y. Takegawa, K. Deguchi, H. Ito, T. Keira, H. Nakagawa, S.I. Nishimura, Simple separation of isometric sialylated N-glycopeptides by a zwitterionic type of hydrophilic interaction chromatography, *J. Sep. Sci.* 29 (2006) 2533–2540. <https://doi.org/10.1002/jssc.200600133>.
- [130] L. Veillon, Y. Huang, W. Peng, X. Dong, B.G. Cho, Y. Mechref, Characterization of isomeric glycan structures by LC-MS/MS, *Electrophoresis.* 38 (2017) 2100–2114. <https://doi.org/10.1002/elps.201700042>.
- [131] M.R. Gama, C.H. Collins, C.B.G. Bottoli, Nano-liquid chromatography in pharmaceutical and biomedical research, *J. Chromatogr. Sci.* 51 (2013) 694–703. <https://doi.org/10.1093/chromsci/bmt023>.
- [132] N.W. Smith, C. Legido-Quigley, N.D. Marlin, V. Melin, I. Mutton, Capillary and Micro High Performance Liquid Chromatography, *Ref. Modul. Chem. Mol. Sci. Chem. Eng.* (2013). <https://doi.org/10.1016/b978-0-12-409547-2.00302-4>.
- [133] Milos V. Novotny, Development of Capillary Liquid Chromatography: a Personal Perspective Minireview Milos, *J. Chromatogr. A.* 1523 (2017) 3–16. <https://doi.org/10.1016/j.chroma.2017.06.042>.
- [134] S.R. Wilson, T. Vehus, H.S. Berg, E. Lundanes, Nano-LC in proteomics: Recent advances and approaches, *Bioanalysis.* 7 (2015) 1799–1815. <https://doi.org/10.4155/bio.15.92>.

- [135] L. Song, K.S. Choi, Y.M. Park, A.L. Kazim, K. Marlar, E.S. Kong, E.M. Park, H.K. Yeul, K.H. Koo, Z.C. Ho, Capillary-LC- $\mu$ ESI-MS/MS and nano-LC-nano ESI-MS/MS analysis using a single binary pump capillary LC system: Applications in proteomics, *J. Liq. Chromatogr. Relat. Technol.* 28 (2005) 1271–1289. <https://doi.org/10.1081/JLC-200054754>.
- [136] F. Xie, R.D. Smith, Y. Shen, Advanced proteomic liquid chromatography, *J. Chromatogr. A.* 1261 (2012) 78–90. <https://doi.org/10.1016/j.chroma.2012.06.098>.
- [137] L. García-Descalzo, E. García-López, A. Alcázar, F. Baquero, C. Cid, Chapter 5. Gel electrophoresis of proteins, in: *Gel Electrophor. – Princ. Basics*, 2012: pp. 57–68. [https://doi.org/10.1016/0378-4347\(87\)80050-6](https://doi.org/10.1016/0378-4347(87)80050-6).
- [138] U.K. Laemmli, Cleavage of structural proteins during the assembly of the head of bacteriophage T4, *Nature.* 228 (1970) 680–685. <https://doi.org/10.1038/227680a0>.
- [139] J.M. Manns, SDS-polyacrylamide gel electrophoresis (SDS-PAGE) of proteins, *Curr. Protoc. Microbiol.* (2011) 1–13. <https://doi.org/10.1002/9780471729259.mca03ms22>.
- [140] R.L.C. Voeten, I.K. Ventouri, R. Haselberg, G.W. Somsen, Capillary Electrophoresis: Trends and Recent Advances, *Anal. Chem.* 90 (2018) 1464–1481. <https://doi.org/10.1021/acs.analchem.8b00015>.
- [141] H. Mischak, J.J. Coon, J. Novak, E.M. Weissinger, J. Schanstra, A.F. Dominiczak, Capillary electrophoresis–mass spectrometry as a powerful tool in biomarker discovery and clinical diagnosis: an update of recent developments, *Mass Spectrom Rev.* 28 (2009) 703–724. <https://doi.org/10.1002/mas.20205>.
- [142] R. Gahoual, E. Leize-wagner, P. Houzé, Y. François, Revealing the potential of capillary electrophoresis / mass spectrometry : the tipping point, *Rapid Commun. Mass Spectrom.* (2018). <https://doi.org/10.1002/rcm.8238>.
- [143] N. Volpi, F. Maccari, R.J. Linhardt, Capillary electrophoresis of complex natural polysaccharides, *Electrophoresis.* 29 (2008) 3095–3106. <https://doi.org/10.1002/elps.200800109>.



- [144] G. Lu, C.L. Crieffield, S. Gattu, L.M. Veltri, L.A. Holland, Capillary Electrophoresis Separations of Glycans, *Chem. Rev.* 118 (2018) 7867–7885.  
<https://doi.org/10.1021/acs.chemrev.7b00669>.
- [145] J. Zaia, Capillary electrophoresis-mass spectrometry of carbohydrates, *Methods Mol. Biol.* 984 (2013) 13–25. [https://doi.org/10.1007/978-1-62703-296-4\\_2](https://doi.org/10.1007/978-1-62703-296-4_2).
- [146] M.C. Breadmore, W. Grochocki, U. Kalsoom, M.N. Alves, S.C. Phung, M.T. Rokh, J.M. Cabot, A. Ghiasvand, F. Li, A.I. Shallan, A.S.A. Keyon, A.A. Alhusban, H.H. See, A. Wuethrich, M. Dawod, J.P. Quirino, Recent advances in enhancing the sensitivity of electrophoresis and electrochromatography in capillaries and microchips (2016–2018), *Electrophoresis.* 40 (2019) 17–39. <https://doi.org/10.1002/elps.201800384>.
- [147] S. Medina-Casanellas, F. Benavente, J. Barbosa, V. Sanz-Nebot, Transient isotachopheresis in on-line solid phase extraction capillary electrophoresis time-of-flight-mass spectrometry for peptide analysis in human plasma, *Electrophoresis.* 32 (2011) 1750–1759. <https://doi.org/10.1002/elps.201100017>.
- [148] Z. Malá, P. Gebauer, Recent progress in analytical capillary isotachopheresis, *Electrophoresis.* 40 (2019) 55–64. <https://doi.org/10.1002/elps.201800239>.
- [149] N.A. Guzman, T. Blanc, T.M. Phillips, Immunoaffinity capillary electrophoresis as a powerful strategy for the quantification of low-abundance biomarkers, drugs, and metabolites in biological matrices, *Electrophoresis.* 29 (2008) 3259–3278.  
<https://doi.org/10.1002/elps.200800058>.
- [150] E. Hernández, F. Benavente, V. Sanz-Nebot, J. Barbosa, Evaluation of on-line solid phase extraction-capillary electrophoresis-electrospray-mass spectrometry for the analysis of neuropeptides in human plasma, *Electrophoresis.* 29 (2008) 3366–3376.  
<https://doi.org/10.1002/elps.200700872>.
- [151] F.W.A. Tempels, W.J.M. Underberg, G.W. Somsen, G.J. de Jong, Design and applications of coupled SPE-CE, *Electrophoresis.* 29 (2008) 108–128.

- <https://doi.org/10.1002/elps.200700149>.
- [152] R. Ramautar, G.W. Somsen, G.J. de Jong, Developments in coupled solid-phase extraction-capillary electrophoresis 2013-2015, *Electrophoresis*. 37 (2016) 35–44. <https://doi.org/10.1002/elps.201500401>.
- [153] F. Benavente, S. Medina-Casanellas, J. Barbosa, V. Sanz-Nebot, Investigation of commercial sorbents for the analysis of opioid peptides in human plasma by on-line SPE-CE, *J. Sep. Sci.* 33 (2010) 1294–1304. <https://doi.org/10.1002/jssc.200900669>.
- [154] L. Pont, R. Pero-Gascon, E. Gimenez, V. Sanz-Nebot, F. Benavente, A critical retrospective and prospective review of designs and materials in in-line solid-phase extraction capillary electrophoresis, *Anal. Chim. Acta.* 1079 (2019) 1–19. <https://doi.org/10.1016/j.aca.2019.05.022>.
- [155] K. Mariño, J. Bones, J.J. Kattla, P.M. Rudd, A systematic approach to protein glycosylation analysis: A path through the maze, *Nat. Chem. Biol.* 6 (2010) 713–723. <https://doi.org/10.1038/nchembio.437>.
- [156] H. Liu, N. Zhang, D. Wan, M. Cui, Z. Liu, S. Liu, Mass spectrometry-based analysis of glycoproteins and its clinical applications in cancer biomarker discovery, *Clin. Proteomics*. 11 (2014) 1–9. <https://doi.org/10.1186/1559-0275-11-14>.
- [157] D.S. Dalpathado, H. Desaire, Glycopeptide analysis by mass spectrometry, *Analyst*. 133 (2008) 731–738. <https://doi.org/10.1039/b713816d>.
- [158] A. Banazadeh, L. Veillon, K.M. Wooding, M. Zabet-moghaddam, Y. Mechref, Recent advances in mass spectrometric analysis of glycoproteins, 2017. <https://doi.org/10.1002/elps.201600357>.
- [159] D. Ribeiro, A Short Overview of the Components in Mass Spectrometry Instrumentation for Proteomics Analyses, *Tandem Mass Spectrom. - Mol. Character.* 1 (2013). <https://doi.org/10.5772/54484>.
- [160] T.S. Bugni, Review of Mass Spectrometry: Instrumentation, Interpretation, and

- Applications Mass Spectrometry: Instrumentation, Interpretation, and Applications.  
Edited by Rolf Ekman, Jerzy Silberring, Ann M. Westman-Brinkmalm, and Agnieszka Kraj. Wiley, New York, 2009, *J. Nat. Prod.* 80 (2017) 574–575.  
<https://doi.org/10.1021/acs.jnatprod.7b00030>.
- [161] A. El-Aneed, A. Cohen, J. Banoub, Mass spectrometry, review of the basics: Electropray, MALDI, and commonly used mass analyzers, *Appl. Spectrosc. Rev.* 44 (2009) 210–230. <https://doi.org/10.1080/05704920902717872>.
- [162] M. Mann, J.B. Fenn, Electropray mass spectrometry: principles and methods, in: D.M. Desiderio (Ed.), *Mass Spectrom.*, 1992: pp. 1–35.
- [163] R.D. Smith, J.A. Loo, M. Loo, Rachel R. Ogorzalek; Busman, H.R. Udseth, Principles and practice of electropray ionization-mass spectrometry for large polypeptides and proteins, *Mass Spectrom. Rev.* 10 (1991) 359–452.
- [164] M. Nakano, D. Higo, E. Arai, T. Nakagawa, K. Kakehi, N. Taniguchi, A. Kondo, Capillary electrophoresis-electropray ionization mass spectrometry for rapid and sensitive N-glycan analysis of glycoproteins as 9-fluorenylmethyl derivatives, in: *Glycobiology*, 2009: pp. 135–143. <https://doi.org/10.1093/glycob/cwn115>.
- [165] J.E. Nettleship, R. Aplin, A. Radu Aricescu, E.J. Evans, S.J. Davis, M. Crispin, R.J. Owens, Analysis of variable N-glycosylation site occupancy in glycoproteins by liquid chromatography electropray ionization mass spectrometry, *Anal. Biochem.* 361 (2007) 149–151. <https://doi.org/10.1016/j.ab.2006.11.005>.
- [166] E. Balaguer, C. Neusu, Glycoprotein characterization combining intact protein and glycan analysis by capillary electrophoresis-electropray ionization-mass spectrometry, *Anal. Chem.* 78 (2006) 5384–5393. <https://doi.org/10.1021/ac060376g>.
- [167] R.D. Smith, C.J. Barinaga, H.R. Udseth, Improved electropray ionization interface for capillary zone electrophoresis-mass spectrometry, *Anal. Chem.* 60 (1988) 1948–1952. <https://doi.org/10.1021/ac00169a022>.

- [168] G. Bacher, W.W. Szymanski, S.L. Kaufman, P. Zllner, D. Blaas, G. Allmaier, Charge-reduced nano electrospray ionization combined with differential mobility analysis of peptides, proteins, glycoproteins, noncovalent protein complexes and viruses, *J. Mass Spectrom.* 36 (2001) 1038–1052. <https://doi.org/10.1002/jms.208>.
- [169] M. Karas, U. Bahr, T. Dülcks, Nano-electrospray ionization mass spectrometry: Addressing analytical problems beyond routine, *Fresenius. J. Anal. Chem.* 366 (2000) 669–676. <https://doi.org/10.1007/s002160051561>.
- [170] A.M. Haag, Chapter 7. Mass analyzers and mass spectrometers, in: H. Mirzaei, M. Carrasco (Eds.), *Mod. Proteomics – Sample Prep. Anal. Pract. Appl. Adv. Exp. Med. Biol.*, Springer, 2016: pp. 157–169. [https://doi.org/10.1007/978-3-319-41448-5\\_7](https://doi.org/10.1007/978-3-319-41448-5_7).
- [171] P. Dolashka, Tandem Mass Spectrometry and Glycoproteins, *Tandem Mass Spectrom. - Appl. Princ.* (2012). <https://doi.org/10.5772/30829>.
- [172] N. Leymarie, J. Zaia, Effective use of mass spectrometry for glycan and glycopeptide structural analysis, *Anal Chem.* 84 (2012) 3040–3048. <https://doi.org/10.1021/ac3000573>.
- [173] H.-H. Hsiao, H. Urlaub, Pseudo-neutral-loss scan for selective detection of phosphopeptides and N-glycopeptides using liquid chromatography coupled with a hybrid linear ion-trap/orbitrap mass spectrometer, *Proteomics.* 10 (2010) 3916–3921.
- [174] R.I. Viner, S. Snovida, E. Bodnar, H. Perreault, A Novel Workflow for Glycopeptide Analysis Using Cellulose-Based Separation Cartridges, TMT-Labeling and LTQ Orbitrap ETD, *J Biomol Tech.* 21 (2010).
- [175] M.S. Bereman, T.I. Williams, D.C. Muddiman, Development of a nanoLC LTQ Orbitrap Mass Spectrometric Method for Profiling Glycans Derived from Plasma from Healthy, Benign Tumor Control, and Epithelial Ovarian Cancer Patients, *Anal Chem.* 81 (2009) 1130–1136. <https://doi.org/10.1021/ac802262w>.
- [176] K.L. Ford, W. Zeng, J.L. Heazlewood, A. Bacic, Characterization of protein N-

- glycosylation by tandem mass spectrometry using complementary fragmentation techniques, *Front. Plant Sci.* 6 (2015) 1–6. <https://doi.org/10.3389/fpls.2015.00674>.
- [177] M.J. Levy, M.P. Washburn, L. Florens, Probing the Sensitivity of the Orbitrap Lumos Mass Spectrometer Using a Standard Reference Protein in a Complex Background, *J. Proteome Res.* 17 (2018) 3586–3592. <https://doi.org/10.1021/acs.jproteome.8b00269>.
- [178] S. Li, Y. Zhou, K. Xiao, J. Li, Z. Tian, Selective fragmentation of the N-glycan moiety and protein backbone of ribonuclease B on an Orbitrap Fusion Lumos Tribrid mass spectrometer, *Rapid Commun. Mass Spectrom.* 32 (2018) 2031–2039. <https://doi.org/10.1002/rcm.8273>.
- [179] M. Gay, E. Sanchez-Jiménez, L. Villarreal, M. Vilanova, R. Huguet, G. Arauz-garofalo, M. Díaz-Lobo, D. López-Ferrer, M. Vilaseca, Chapter 12. Top-Down Proteomics Applied to Human Cerebrospinal Fluid, in: E. Santamaria, J. Fernandez-Irigoyen (Eds.), *Cerebrospinal Fluid Proteomics Methods Protoc. Methods Mol. Biol.*, Springer, 2019: pp. 193–219. [https://doi.org/10.1007/978-1-4939-9706-0\\_12](https://doi.org/10.1007/978-1-4939-9706-0_12).
- [180] D.J. Harvey, L. Royle, C.M. Radcliffe, P.M. Rudd, R.A. Dwek, Structural and quantitative analysis of N-linked glycans by matrix-assisted laser desorption ionization and negative ion nanospray mass spectrometry, *Anal. Biochem.* 376 (2008) 44–60. <https://doi.org/10.1016/j.ab.2008.01.025>.
- [181] D.J. Harvey, Fragmentation of negative ions from carbohydrates: Part 3. Fragmentation of hybrid and complex N-linked glycans, *J. Am. Soc. Mass Spectrom.* 16 (2005) 647–659. <https://doi.org/10.1016/j.jasms.2005.01.006>.
- [182] Y. Mu, B.L. Schulz, V. Ferro, Applications of ion mobility-mass spectrometry in carbohydrate chemistry and glycobiology, *Molecules.* 23 (2018) 1–17. <https://doi.org/10.3390/molecules23102557>.
- [183] C.S. Lane, K. McManus, P. Widdowson, S.A. Flowers, G. Powell, I. Anderson, J.L. Campbell, Separation of Sialylated Glycan Isomers by Differential Mobility

- Spectrometry, *Anal. Chem.* 91 (2019) 9916–9924.  
<https://doi.org/10.1021/acs.analchem.9b01595>.
- [184] C. Manz, M. Grabarics, F. Hoberg, M. Pugini, A. Stuckmann, W.B. Struwe, K. Pagel, Separation of isomeric glycans by ion mobility spectrometry – the impact of fluorescent labelling, *Analyst.* 144 (2019) 5292–5298. <https://doi.org/10.1039/c9an00937j>.
- [185] J. Hofmann, A. Stuckmann, M. Crispin, D.J. Harvey, K. Pagel, W.B. Struwe, Identification of Lewis and Blood Group Carbohydrate Epitopes by Ion Mobility-Tandem-Mass Spectrometry Fingerprinting, *Anal. Chem.* 89 (2017) 2318–2325.  
<https://doi.org/10.1021/acs.analchem.6b03853>.
- [186] H. Hinneburg, J. Hofmann, W.B. Struwe, A. Thader, F. Altmann, D. Varón Silva, P.H. Seeberger, K. Pagel, D. Kolarich, Distinguishing N-acetylneuraminic acid linkage isomers on glycopeptides by ion mobility-mass spectrometry, *Chem. Commun.* 52 (2016) 4381–4384. <https://doi.org/10.1039/C6CC01114D>.
- [187] R. Cumeras, E. Figueras, C.E. Davis, J.I. Baumbach, I. Gràcia, Review on Ion Mobility Spectrometry. Part 1: Current instrumentation, *Analyst.* 140 (2015) 1376–1390.  
<https://doi.org/10.1039/c4an01100g>.
- [188] W.B. Struwe, K. Pagel, J.L.P. Benesch, D.J. Harvey, M.P. Campbell, GlycoMob: an ion mobility-mass spectrometry collision cross section database for glycomics, *Glycoconj. J.* 33 (2016) 399–404. <https://doi.org/10.1007/s10719-015-9613-7>.
- [189] K. Pagel, D.J. Harvey, Ion mobility-mass spectrometry of complex carbohydrates: Collision cross sections of sodiated N-linked glycans, *Anal. Chem.* 85 (2013) 5138–5145.  
<https://doi.org/10.1021/ac400403d>.
- [190] A.A. Shvartsburg, R.D. Smith, Fundamentals of Traveling Wave Ion Mobility Spectrometry, *Anal Chem.* 80 (2008) 9689–9699. <https://doi.org/10.1021/ac8016295>.
- [191] I.D.G. Campuzano, K. Giles, Historical, Current and Future Developments of Traveling Wave Ion Mobility Mass Spectrometry: A Personal Perspective, *TrAC Trends Anal.*

- Chem. 120 (2019) 115620. <https://doi.org/10.1016/j.trac.2019.115620>.
- [192] D.J. Harvey, W.B. Struwe, Structural Studies of Fucosylated N-Glycans by Ion Mobility Mass Spectrometry and Collision-Induced Fragmentation of Negative Ions, *J. Am. Soc. Mass Spectrom.* 29 (2018) 1179–1193. <https://doi.org/10.1007/s13361-018-1950-x>.
- [193] A. Barroso, E. Giménez, A. Konijnenberg, J. Sancho, V. Sanz-Nebot, F. Sobott, Evaluation of ion mobility for the separation of glycoconjugate isomers due to different types of sialic acid linkage, at the intact glycoprotein, glycopeptide and glycan level, *J. Proteomics.* 173 (2018) 22–31. <https://doi.org/10.1016/j.jprot.2017.11.020>.
- [194] D.J. Harvey, C.A. Scarff, M. Edgeworth, K. Pagel, K. Thalassinou, W.B. Struwe, M. Crispin, J.H. Scrivens, Travelling-wave ion mobility mass spectrometry and negative ion fragmentation of hybrid and complex N-glycans, *J. Mass Spectrom.* 51 (2016) 1064–1079. <https://doi.org/10.1002/jms.3828>.
- [195] L. Yi, N. Dong, Y. Yun, B. Deng, D. Ren, S. Liu, Y. Liang, Chemometric methods in data processing of mass spectrometry-based metabolomics: A review, *Anal. Chim. Acta.* 914 (2016) 17–34. <https://doi.org/10.1016/j.aca.2016.02.001>.
- [196] S.E. Richards, E. Holmes, Chapter 3. Chemometrics methods for the analysis of genomics, transcriptomics, proteomics, metabolomics, and metagenomics datasets, in: *Metabolomics as a Tool Nutr. Res.*, Elsevier Ltd., 2015: pp. 37–60. <https://doi.org/10.1016/B978-1-78242-084-2.00003-4>.
- [197] S. Wold, K. Esbensen, P. Geladi, Principal component analysis, *Chemom. Intell. Lab. Syst.* 2 (1987) 37–52.
- [198] I.T. Jolliffe, J. Cadima, Principal component analysis: A review and recent developments, *Philos. Trans. R. Soc. A.* 374 (2016) 1–16. <https://doi.org/10.1098/rsta.2015.0202>.
- [199] K.L. Sainani, Introduction to principal components analysis, *PM R.* 6 (2014) 275–278. <https://doi.org/10.1016/j.pmrj.2014.02.001>.
- [200] S. Wold, A. Ruhe, H. Wold, W.J. Dunn, III, The Collinearity Problem in Linear Regression.

- The Partial Least Squares (PLS) Approach to Generalized Inverses, *SIAM J. Sci. Stat. Comput.* 5 (1984) 735–743. <https://doi.org/10.1137/0905052>.
- [201] S. Wold, M. Sjöström, L. Eriksson, PLS-regression: A basic tool of chemometrics, *Chemom. Intell. Lab. Syst.* 58 (2001) 109–130. [https://doi.org/10.1016/S0169-7439\(01\)00155-1](https://doi.org/10.1016/S0169-7439(01)00155-1).
- [202] T. Rajalahti, O.M. Kvalheim, Multivariate data analysis in pharmaceuticals: A tutorial review, *Int. J. Pharm.* 417 (2011) 280–290. <https://doi.org/10.1016/j.ijpharm.2011.02.019>.
- [203] P. Geladi, B.R. Kowalski, Partial least-squares regression: A tutorial, *Analy.* 185 (1986) 1–17.
- [204] R.G. Brereton, G.R. Lloyd, Partial least squares discriminant analysis: Taking the magic away, *J. Chemom.* 28 (2014) 213–225. <https://doi.org/10.1002/cem.2609>.
- [205] P.S. Gromski, H. Muhamadali, D.I. Ellis, Y. Xu, E. Correa, M.L. Turner, R. Goodacre, A tutorial review: Metabolomics and partial least squares-discriminant analysis - a marriage of convenience or a shotgun wedding, *Anal. Chim. Acta.* 879 (2015) 10–23. <https://doi.org/10.1016/j.aca.2015.02.012>.
- [206] 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. <https://doi.org/10.1002/elps.201500255>.
- [207] T.E. Thingholm, M.R. Larsen, The Use of Titanium Dioxide Micro-Columns to Selectively Isolate Phosphopeptides from Proteolytic Digests, *J. Proteome Res.* 8 (2009) 57–66. <https://doi.org/10.1007/978-1-60327-834-8>.
- [208] R. Peró-Gascón, L. Pont, F. Benavente, J. Barbosa, V. Sanz-Nebot, Analysis of serum transthyretin by on-line immunoaffinity solid-phase extraction capillary electrophoresis



- mass spectrometry using magnetic beads, *Electrophoresis*. 37 (2016) 1220–1231.  
<https://doi.org/10.1002/elps.201500495>.
- [209] R. Pero-Gascon, F. Benavente, Z. Minic, M. V. Berezovski, V. Sanz-Nebot, On-line Aptamer Affinity Solid-Phase Extraction Capillary Electrophoresis-Mass Spectrometry for the Analysis of Blood  $\alpha$ -Synuclein, *Anal. Chem.* (2020).  
<https://doi.org/10.1021/acs.analchem.9b04802>.
- [210] F. Benavente, M.C. Vescina, E. Hernández, V. Sanz-Nebot, J. Barbosa, N.A. Guzman, Lowering the concentration limits of detection by on-line solid-phase extraction-capillary electrophoresis-electrospray mass spectrometry, *J. Chromatogr. A*. 1140 (2007) 205–212. <https://doi.org/10.1016/j.chroma.2006.11.092>.
- [211] W. Cao, J. Cao, J. Huang, L. Zhang, J. Yao, H. Xu, P. Yang, Enhanced N-glycosylation site exploitation of sialoglycopeptides by peptide IPG-IEF assisted TiO<sub>2</sub>chromatography, *Glycoconj. J.* 29 (2012) 433–443. <https://doi.org/10.1007/s10719-012-9404-3>.
- [212] M. Shan, D. Yang, H. Dou, L. Zhang, Chapter 4. Fucosylation in cancer biology and its clinical applications, in: *Prog. Mol. Biol. Transl. Sci.*, 1st ed., Elsevier Inc., 2019: pp. 93–119. <https://doi.org/10.1016/bs.pmbts.2019.01.002>.
- [213] K. Deguchi, Y. Takegawa, H. Ito, N. Miura, S. Yoshioka, S. Nagai, H. Nakagawa, S.I. Nishimura, Structural assignment of isomeric 2-aminopyridine-derivatized monosialylated biantennary N-linked oligosaccharides using negative-ion multistage tandem mass spectral matching, *Rapid Commun. Mass Spectrom.* 20 (2006) 412–418.  
<https://doi.org/10.1002/rcm.2320>.
- [214] D.J. Harvey, Fragmentation of negative ions from carbohydrates: Part 3. Fragmentation of hybrid and complex N-linked glycans, *J. Am. Soc. Mass Spectrom.* 16 (2005) 647–659.  
<https://doi.org/10.1016/j.jasms.2005.01.006>.
- [215] D.J. Harvey, Structural determination of N-linked glycans by matrix-assisted laser desorption/ionization and electrospray ionization mass spectrometry, *Proteomics*. 5

- (2005) 1774–1786. <https://doi.org/10.1002/pmic.200401248>.
- [216] D.J. Harvey, Fragmentation of negative ions from carbohydrates: Part 2. Fragmentation of high-mannose N-linked glycans, *J. Am. Soc. Mass Spectrom.* 16 (2005) 631–646. <https://doi.org/10.1016/j.jasms.2005.01.005>.
- [217] C. Michael, A.M. Rizzi, Tandem mass spectrometry of isomeric aniline-labeled N-glycans separated on porous graphitic carbon: Revealing the attachment position of terminal sialic acids and structures of neutral glycans, *Rapid Commun. Mass Spectrom.* 29 (2015) 1268–1278. <https://doi.org/10.1002/rcm.7208>.
- [218] S.F. Wheeler, D.J. Harvey, Negative ion mass spectrometry of sialylated carbohydrates: Discrimination of N-acetylneuraminic acid linkages by MALDI-TOF and ESI-TOF mass spectrometry, *Anal. Chem.* 72 (2000) 5027–5039. <https://doi.org/10.1021/ac000436x>.
- [219] D.J. Harvey, R.L. Martin, K.A. Jackson, C.W. Sutton, Fragmentation of N-linked glycans with a matrix-assisted laser desorption/ionization ion trap time-of-flight mass spectrometer, *Rapid Commun. Mass Spectrom.* 18 (2004) 2997–3007. <https://doi.org/10.1002/rcm.1709>.
- [220] J. Hofmann, K. Pagel, Glycan Analysis by Ion Mobility–Mass Spectrometry, *Angew. Chemie - Int. Ed.* 56 (2017) 8342–8349. <https://doi.org/10.1002/anie.201701309>.
- [221] M. Guttman, K.K. Lee, Site-Specific Mapping of Sialic Acid Linkage Isomers by Ion Mobility Spectrometry, *Anal. Chem.* 88 (2016) 5212–5217. <https://doi.org/10.1021/acs.analchem.6b00265>.
- [222] X. Zheng, X. Zhang, N.S. Schocker, R.S. Renslow, D.J. Orton, J. Khamsi, R.A. Ashmus, I.C. Almeida, K. Tang, C.E. Costello, R.D. Smith, K. Michael, E.S. Baker, Enhancing glycan isomer separations with metal ions and positive and negative polarity ion mobility spectrometry-mass spectrometry analyses, *Anal. Bioanal. Chem.* 409 (2017) 467–476. <https://doi.org/10.1007/s00216-016-9866-4>.
- [223] C. Manz, K. Pagel, Glycan analysis by ion mobility-mass spectrometry and gas-phase

- spectroscopy, *Curr. Opin. Chem. Biol.* 42 (2018) 16–24.  
<https://doi.org/10.1016/j.cbpa.2017.10.021>.
- [224] C. Reily, T.J. Stewart, M.B. Renfrow, J. Novak, Glycosylation in health and disease, *Nat. Rev. Nephrol.* 15 (2019) 346–366. <https://doi.org/10.1038/s41581-019-0129-4>.
- [225] A. Rosal-Vela, A. Barroso, E. Giménez, S. García-Rodríguez, V. Longobardo, J. Postigo, M. Iglesias, A. Lario, J. Merino, R. Merino, M. Zubiaur, V. Sanz-Nebot, J. Sancho, Identification of multiple transferrin species in the spleen and serum from mice with collagen-induced arthritis which may reflect changes in transferrin glycosylation associated with disease activity: The role of CD38, *J. Proteomics.* 134 (2016) 127–137.  
<https://doi.org/10.1016/j.jprot.2015.11.023>.
- [226] A. Rosal-Vela, S. García-Rodríguez, J. Postigo, M. Iglesias, V. Longobardo, A. Lario, J. Merino, R. Merino, M. Zubiaur, J. Sancho, Distinct serum proteome profiles associated with collagen-induced arthritis and complete Freund's adjuvant-induced inflammation in CD38<sup>-/-</sup> mice: The discriminative power of protein species or proteoforms, *Proteomics.* 15 (2015) 3382–3393. <https://doi.org/10.1002/pmic.201400536>.
- [227] R.A. Feelders, G. Vreugdenhi, G. de Jong, A.J.G. Swaak, H.G. van Eijk, Transferrin microheterogeneity in rheumatoid arthritis, *Rheumatol. I.* 12 (1992) 195–199.  
<https://doi.org/10.1007/bf00302152>.
- [228] S.Y. Gwon, K.J. Rhee, H.J. Sung, Gene and protein expression profiles in a mouse model of collagen-induced arthritis, *Int. J. Med. Sci.* 15 (2018) 77–85.  
<https://doi.org/10.7150/ijms.22345>.
- [229] B. Coddeville, E. Regoeczi, G. Strecker, Y. Plancke, G. Spik, Structural analysis of trisialylated biantennary glycans isolated from mouse serum transferrin. Characterization of the sequence Neu5Gc(α2-3)Gal(β1-3)[Neu5Gc(α2-6)]GlcNAc(β1-2)Man, *Biochim. Biophys. Acta.* 1475 (2000) 321–328. [https://doi.org/10.1016/S0304-4165\(00\)00083-0](https://doi.org/10.1016/S0304-4165(00)00083-0).

- [230] S. Albrecht, L. Unwin, M. Muniyappa, P.M. Rudd, Glycosylation as a marker for inflammatory arthritis, *Cancer Biomarkers*. 1 (2014) 17–28.  
<https://doi.org/10.3233/CBM-130373>.



