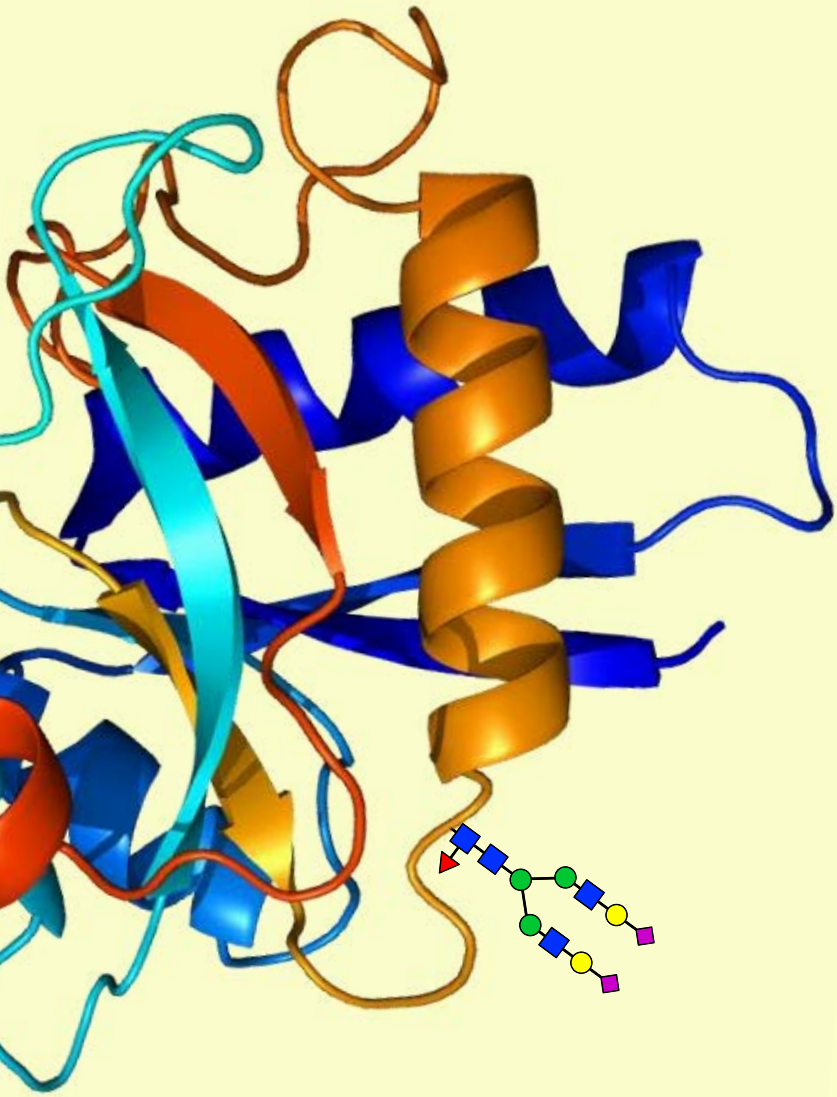
UNIVERSITAT DE BARCELONA

Desarrollo de metodologías analíticas para la separación, detección y caracterización de biomarcadores glicoproteicos en muestras biológicas

Albert Barroso Ramos

Desarrollo de metodologías analíticas para la separación, detección y caracterización de biomarcadores glicoproteicos en muestras biológicas

Albert Barroso Ramos



UNIVERSITAT DE BARCELONA

FACULTAD DE QUÍMICA

DEPARTAMENTO DE INGENIERÍA QUÍMICA Y QUÍMICA ANALÍTICA

Programa de Doctorado:

Química Analítica del Medio Ambiente y la Polución

Desarrollo de metodologías analíticas para la separación, detección y caracterización de biomarcadores glicoproteicos en muestras biológicas

---

Memoria presentada por:

Albert Barroso Ramos

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: "**Desarrollo de metodologías analíticas para la separación, detección y caracterización de biomarcadores glicoproteicos en muestras biológicas**", ha sido realizada bajo su dirección por el Sr. Albert Barroso Ramos y que todos los resultados presentados son fruto de las experiencias realizadas por el citado doctorando.

Y para que así conste, expiden y firman el presente certificado.

Barcelona, Junio 2017

Dra. Victoria Sanz Nebot Dra. Estela Giménez López

Ahora que por fin hemos llegado al final de este camino, no podría empezar este trabajo sin agradecer primero a todos aquellos que de alguna manera han formado parte de este trayecto y me han ayudado a llegar al final.

En primer lugar, querría agradecer todo el trabajo y el esfuerzo realizado por mis directoras de tesis, la Dra. Victoria Sanz y la Dra. Estela Giménez. Vicki, gracias por la orientación recibida y por todos los consejos e ideas que has aportado a lo largo de estos años para que no faltase ningún detalle. Estela, muchas gracias por guiarme durante todos estos años, por todo lo que me has enseñado y por todos los consejos e ideas que han hecho que este trabajo haya llegado a un buen final. Muchas gracias por ayudarme cuando parecía que los resultados que se obtenían no eran los que nosotros esperábamos y por buscar siempre una alternativa.

Quería agradecer también al Dr. José Barbosa por la confianza recibida al permitirme formar parte de este grupo de investigación, por sus consejos y por sus palabras alentadoras. Gracias también al Dr. Fernando Benavente, gracias por tu ayuda en este trabajo, por las nuevas ideas que has aportado y por esa actitud positiva durante todos estos años.

Agradecer también al Dr. Jaime Sancho del Instituto de Parasitología y Biomedicina "López - Neyra" por su colaboración, su profesionalidad, por aportar nuevas ideas a este trabajo y por su predisposición a iniciar nuevos proyectos.

I also wanted to thank Dr. Frank Sobott, from the University of Antwerp, for allowing me to work in his research group and for his orientation, continuous support and kind attitude during the stay and the following months. I am also grateful to Albert Konijnenberg and Sneha Chatterjee for solving all my problems during the stay and for their support and help.

Sobre todo, muchísimas gracias a todos mis compañeros de laboratorio, con los que he compartido todos estos años y los que han hecho que este largo viaje haya sido mucho más ameno.

Muchísimas gracias a mi gran compañera y amiga, Laura Pont, gracias por estar conmigo casi desde el principio, por aguantarme todos estos años, por darme ánimos y apoyarme siempre

que lo he necesitado. Gracias por estar siempre ahí, porque sin ti hubiera sido imposible llegar al final. Muchas gracias por hacer la estancia en el país vecino, por esos fines de semana donde visitábamos ciudades helados hasta los pies pero siempre con ganas e ilusión. Gracias por hacer que los cuatro meses de estancia pasasen volando. Espero que allá donde vayamos esta amistad no la perdamos nunca.

Muchas gracias Montse, por formar parte de esta última mitad del doctorado y porque sé que también me llevo una amiga. Gracias por tu alegría, tu risa, tus grandes momentos y, por qué no, también por tu locura. Gracias por entrar dándolo todo, por ser como eres, porque has traído diversión al laboratorio pero también por ayudar y estar ahí cuando te he necesitado. Espero que nunca pierdas ese toque y que podamos seguir viéndonos durante muchos años más.

Muchas gracias también a Roger e Irena, por todos los momentos compartidos y todas las risas, por ser siempre tan atentos y amables y por preguntar y ofrecer vuestra ayuda cuando las cosas no iban bien. Muchas gracias por vuestra visita a Holanda porque hicisteis de ese fin de semana un momento único.

Gracias a Laura Pey y Lorena Villegas, por todas las risas compartidas y porque a pesar de ser las recién llegadas habéis encajado muy rápido en este grupo contribuyendo a crear un gran ambiente. Espero que el máster y todos los proyectos que inicies después os vayan muy bien.

Gracias a Silvia por todas esas conversaciones, por las risas en el laboratorio y por ayudarme siempre que lo he necesitado. Gracias por los buenos momentos durante el desayuno y la comida y por compartir conmigo más de la mitad de este camino.

Gracias a Fran, Marcos y Elena, por todos esos días que nos quedábamos más tarde en el bar, por esas salidas para desconectar y pasarlo bien. Gracias por hacer los viajes a Torribera mucho más llevaderos, por esas conversaciones y por los buenos momentos que me habéis hecho pasar.

Gracias a Gerard porque nos hiciste pasar un año muy divertido y porque has continuado ahí con tu buen humor y tu actitud positiva todos estos años, ayudándome a desconectar y a reírme incluso en los momentos más duros. Gracias también a Leo por haber formado parte de este camino, por los buenos ratos en el bar que me han ayudado a olvidarme de los problemas y a desconectar y por tener siempre esa actitud positiva.

Thank you very much Kader, thank you for being the kindest person I have ever met, thank you for your good wishes, your words of caution and for taking care of all the people of the lab as if we were your family, I wish you all the best.

También me gustaría agradecer a todas aquellas personas que han pasado por el laboratorio por su compañerismo y por crear un buen ambiente, ya que ellos también han formado parte de este trayecto: Lorena Ortiz, Anna, Marta, Raquel, Mulu, Sandra, Sergio Català, Rafa, Pedro, Sergio Barbosa, Mardjan, Noelia, Pili… Agradecer también a todos los compañeros de Química Analítica, profesores y personal de secretaria porque todos ellos han contribuido de alguna manera a que esta tesis doctoral sea posible. En especial agradecer a Sergio Carneado su amabilidad, su ayuda cuando la he necesitado y los buenos momentos a la hora de la comida.

Gracias a mis compañeros de universidad, en especial a Juanma y Noelia, por escuchar mis quejas, y por esas quedadas para apoyarnos mutuamente y darnos ánimos en este camino.

Finalmente, agradecer a toda mi familia por todo el apoyo recibido durante estos años de tesis, porque siempre estáis ahí si os necesito y por acompañarme en los momentos más duros y alegraros conmigo en los momentos más buenos.

Abbreviations

Symbol / Number

β-ME	2-mercaptoethanol
Ω/z	Collision cross-section-to-charge ratio
1Ant	Monoantennary
2Ant	Biantennary
2D-PAGE	Two-dimensional polyacrylamide gel electrophoresis
3Ant	Triantennary

A

ACN	Acetonitrile
Ab	Antibody
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

Ant	Antenna
A_{norm}	Normalized area
ANOVA	Analysis of variance
APO-CIII	Apolipoprotein C-III
ATD	Arrival time distribution
ATP	Adenosine triphosphate

B

bAGP	Bovine α-1-acid glycoprotein
BGE	Background electrolyte

C

CapLC (μLC)	Capillary liquid chromatography
CapLC-MS	Capillary liquid chromatography mass spectrometry
CapLC-TOF-MS	Capillary liquid chromatography time-of-flight mass spectrometry
CCS (Ω)	Collision cross-section
CDG	Congenital disorders of glycosylation
CDG-I	Congenital disorders of glycosylation type I
CDG-II	Congenital disorders of glycosylation type II
CDT	Carbohydrate deficient transferrin
CD38-KO	Mice deficient in CD38 gene
CE	Capillary electrophoresis
CEC	Capillary electrochromatography
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
CFG	Consortium for Functional Glycomics
CGE	Capillary gel electrophoresis
CHO	Chinese hamster ovary cells
CIA	Collagen induced arthritis
CID	Collision induced dissociation
CIEF	Capillary isoelectric focusing
CITP	Capillary isotachophoresis
Col-II	Chicken collagen type II
CZE	Capillary zone electrophoresis

D

Da	Dalton
DiGE	Two-dimensional fluorescence difference gel electrophoresis
DNA	Deoxyribonucleic acid
Dol-P	Dolichol phosphate
DTIMS	Drift-time ion mobility spectrometry
DTT	Dithiothreitol

E

ECD	Electron-capture dissociation
EIC	Extracted ion chromatogram

EIE	Extracted ion electropherogram
EOF	Electroosmotic flow
EPO	Human erythropoietin
ESI	Electrospray ionization
ESI-L	Electrospray ionization low concentration tuning mix
ETD	Electron-transfer dissociation

F

FAIMS	Field asymmetric waveform ion mobility spectrometry
FT-ICR	Fourier transform ion cyclotron resonance

H

HAc	Acetic acid
hAGP (or AGP)	Human α-1-acid glycoprotein
hEPO	Human erythropoietin
HFor	Formic acid
HILIC	Hydrophilic interaction liquid chromatography
HLB	Hydrophilic-Lipophilic-Balanced
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 (or Tf)	Human transferrin

I

IAA	Iodoacetamide
IAC	Immunoaffinity chromatography
i.d.	Inner or internal diameter
IEF	Isoelectric focusing
Ig	Immunoglobulin
IM	Ion mobility
IMS	Ion mobility spectrometry
IM-MS	Ion mobility mass spectrometry
iPrOH	2-propanol or isopropanol

K

K_a	Acid dissociation constant

L

LacNAc	N-acetyllactosamine

LC	Liquid chromatography
LC-MS	Liquid chromatography mass spectrometry
LOD	Limit of detection
L_T	Total length
LV	Latent variable

M

M or M_r	Molecular mass
m_e (μ_e)	Electrophoretic mobility
M_{exp}	Experimental molecular mass
M_{theo} / M_{teo}	Theoretical molecular mass
MALDI	Matrix-assisted laser desorption/ionization
MALDI-TOF/TOF-MS	Matrix assisted laser desorption ionization tandem time-of-flight mass spectrometry
ME	Microchip electrophoresis
MEKC	Micellar electrokinetic chromatography
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	Neu5Ac	S
N-glycolylneuraminic acid	Neu5Gc	S
Sialic acid or neuraminic acid	SiA	S

MRM	Multiple reaction monitoring
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
nanoUPLC	Nano-ultra performance liquid chromatography
NH_4Ac	Ammonium acetate

O

oa	Orthogonal acceleration
oa-TOF	Orthogonal acceleration time-of-flight mass analyzer
o.d.	Outer diameter

P

PC	Principal component
PCA	Principal component analysis
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
qTOF	Hybrid quadrupole time-of-flight analyzer
q/M	Charge-to-mass ratio

R

RA	Rheumatoid arthritis
rhEPO	Recombinant human erythropoietin
rhEPO-T	Recombinant human erythropoietin digested with trypsin
rhEPO-TN	Recombinant human erythropoietin digested with trypsin and neuraminidase
RSD (or %RSD)	Relative standard deviation (%)
R^2	Regression coefficient

S

s	Standard deviation
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SiOH	Silanol groups
SPE	Solid phase extraction

S/N	Signal-to noise ratio

T

t’	Relative migration time
TFA	Trifluoroacetic acid
TOF	Time-of-flight mass analyzer
TOF-MS	Time-of-flight mass spectrometry
t_R	Retention time
TWIMS	Traveling wave ion mobility spectrometry

U

UPLC	Ultra-performance liquid chromatography
UV	Ultraviolet spectrophotometry

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

Abstract

Glycomics is often considered one of the most convoluted *omic* sciences, reason why advancements in this field have been comparatively scarcer. The main focus of glycomics is the study of the **glycosylation** of biomolecules, being glycosylated proteins, or **glycoproteins**, the most important ones. Glycosylation of proteins is the covalent attachment of carbohydrates, also known as **glycans**, to the peptide backbone. Glycosylation is a **co-** and **posttranslational modification (PTM)** that occurs in the final steps of the protein synthesis which immensely enhances its functional diversity. Glycoproteins are known to be responsible for a wide variety of biological processes, such as signaling events, cell-cell interaction or even protein folding. The fact that in complex organisms, such as mammals, more than half of all secretory and cellular proteins are glycosylated serves as proof of their great importance.

Glycans can have a myriad of structures and be composed of several different monosaccharide units. All these possible variations in a glycoprotein contribute to their **microheterogeneity** and are the main obstacle the scientific community faces, often considered the bottleneck in glycomics advances. A given glycoprotein exists as a heterogeneous mixture of glycoconjugate species known as **glycoforms** that differ in the amount, size and structure of their glycans or the amino acid to which they are bound. A single glycoprotein can have multiple glycoforms, ranging from less than five to even more than a hundred.

Understandably then, a modification of the normal glycosylation profile of a certain glycoprotein can have dire consequences, being several biological or pathological processes the result of an alteration in the relative abundance of some glycoforms. Hence, the work presented in this thesis is focused on the development of analytical methodologies to characterize the glycosylation pattern of proteins suspected to be altered, whether due to pathological processes or unhealthy practices, and identify new or altered glycoforms as potential **biomarkers**.

Additionally, the developed methodologies could also be applied to characterize the glycosylation of **recombinant glycoproteins**. These proteins are quite often used as biopharmaceuticals and their glycosylation pattern, which is frequently different from that of the endogenous glycoprotein, may cause an adverse immune response. Therefore, the characterization of their glycosylation is vital to properly asses their stability, biological activity and pharmacokinetic to ensure their proper therapeutic function.

One of the first glycosylation studies carried out involved the detection and characterization of **human transferrin (hTf)** glycopeptide glycoforms obtained after tryptic digestion by capillary liquid chromatography coupled to mass spectrometry (**CapLC-MS**), using an orthogonal acceleration time-of-flight mass analyzer (**oa-TOF**). The aim was to verify whether its glycosylation was altered in individuals with **chronic alcoholism** and **congenital disorders of glycosylation (CDG)**, a rare genetic disease. After CapLC-MS was successfully optimized and properly applied to characterize glycopeptide glycosylation of a hTf standard, the proposed methodology was, then, applied to serum samples where hTf was isolated from the rest of the components of the serum using a home-made immunoaffinity column.

In the case of chronic alcoholism, after immunoaffinity purification and tryptic digestion, serum samples from a teetotaler individual (as control) and from individuals with low and high alcohol dependence were analyzed by the aforementioned established CapLC-MS method. Several glycoforms were observed to be altered in chronic alcoholism, but mainly, the most important ones were the H5N4S2 which decreased and the non-glycosylated glycoform (peptide glycoform) which increased in relative abundance.

With regard to the CDG serum samples, the same methodology was applied to study the alteration of hTf glycosylation produced by this disease. However, the data set originated in this study was larger and more complex due to the higher number of samples. Reason why, multivariate data analysis was necessary to properly determine which altered glycoforms were

the most important. Principal component analysis (**PCA**) and partial least squares discriminant analysis (**PLS-DA**) were evaluated for the classification of the different samples (control, CDG-I and CDG-II) and for providing a novel insight into hTf glycopeptide glycoforms alteration in CDG. Six out of the fourteen glycoforms detected were considered as the most important ones to differentiate between the three groups of samples, and could be regarded as potential novel biomarkers for this disease.

The following work involved the analysis of **mouse transferrin (mTf)**, a suspected altered glycoprotein in **collagen induced arthritis (CIA)**. CIA is an inflammatory and autoimmune disease which resembles human **rheumatoid arthritis (RA)** in terms of its disease course, histological findings, and in its response to commonly used anti-arthritic drugs. Using the developed CapLC-MS method to separate and detect mTf glycopeptide glycoforms, an increase in fucosylation and glycan branching was observed in sera from mice with CIA versus control. Additionally, the effect of a specific gene was evaluated to assess its contribution to the development of CIA. In particular, genetically modified mice deficient in CD38 (CD38-KO mice) were compared to standard mice (WT) in regards to mTf glycosylation alteration in CIA. The results showed that in both types of mice, the glycosylation of hTf was affected, albeit the alteration was somewhat different between WT and CD38-KO. These results validated the developed methodology as a proper and reliable approach to study the glycosylation of proteins in the glycopeptide level.

Additionally, a methodology based on capillary electrophoresis coupled to mass spectrometry detection (**CE-MS**), using an orthogonal acceleration time-of-flight mass analyzer, was also evaluated as an alternative method to CapLC-MS for the analysis of the glycosylation of proteins. However, the use of a mandatory anionic surfactant to correctly perform the digestion of proteolytic-resistant glycoproteins, such as human or mouse transferrin, proved detrimental to CE-MS analysis. After surfactant hydrolysis, one of the two remaining products

seemed to interact with the inner surface of the silica capillary walls, resulting in wide and distorted electrophoretic peaks. Therefore, as CE is often regarded as an excellent analytical technique in glycosylation studies due to its ability to separate glycopeptide glycoforms, especially when they differ in the number of sialic acids, some approaches regarding sample treatment were evaluated to completely eliminate the anionic surfactant. The so-called µElution plates (96-well plates with Oasis® HLB (WATERS®) stationary phase), were the only method that provided proper sample clean-up with good retention of the glycopeptides, minimal loss of sample and complete elimination of the remaining surfactant product. The robustness of the developed method was afterwards evaluated with the analysis of hTf glycopeptides in serum samples.

Another interesting advantage of using CE as separation technique prior MS detection is the fact that the electrophoretic migration behavior of ions can be easily predicted using the classical semiempirical relationships between **electrophoretic mobility** and **charge-to-mass ratio** ($\mathbf{m_e}$ vs $\mathbf{q/M^{\alpha}}$). Prediction of electrophoretic migration behavior is an appealing modelization tool to speed-up method development in CE-MS, as well as to refine the structural assignments based on the measured molecular mass (M). In this thesis, the **classical semiempirical relationships** were used to predict and model the migration behavior of peptides and glycopeptides originated from the digestion of **recombinant human erythropoietin (rhEPO)**, a therapeutically relevant glycoprotein, which was considered a good model glycoprotein due to its high glycosylation and microheterogeneity. Three different relationships were evaluated, being the one that predicted the migration behavior of peptides different than the one for the glycopeptides. Simulated electropherograms were elaborated with these models, which matched almost perfectly the experimental electropherograms. Results were later validated predicting the migration and simulating the separation of a different set of rhEPO glycopeptides and also hTf peptides and glycopeptides, which also agreed with the experimental electropherograms.

As stated before, the study of protein glycosylation is often regarded as an intricate, albeit important, task. Complexity of glycan structures can escalate rather quickly when we take into account that some monosaccharide units can have different orientation or be bound to different carbohydrates, giving rise to the formation of **isomeric** glycan structures. For instance, the sialic acid residue can have mainly two types of linkage in complex N-glycans, referred to as **α-2→3** or **α-2→6** linkages. In this regard, **ion mobility mass spectrometry (IM-MS)** has gain popularity for the separation of isomeric compounds. For this reason, the last work presented in this thesis aimed to assess the capability of IM-MS for the separation of isomeric glycoconjugates due to the type of sialic acid linkage, using mTf as model glycoprotein. Even though separation of isomeric glycans was evaluated at the intact glycoprotein, glycopeptide and glycan levels, straightforward separation of isomers was achieved with the analysis of glycans, as opposed to the glycopeptides, which was relatively more complicated, or the intact glycoprotein, which was not possible. Positive results were obtained with the analysis of the free glycans not only in standard samples but also in serum samples from mice control and mice with CIA, acknowledging the applicability of the developed methodology to study real complex samples as well as its potential for the separation of isomeric glycans.

Aims

The work presented in this thesis is focused on the development and optimization of analytical methodologies for the characterization of the **glycosylation** of biologically and therapeutically relevant **glycoproteins**, using **capillary liquid chromatography**, as well as **capillary electrophoresis**, both hyphenated to mass spectrometry (**CapLC-MS** and **CE-MS**, respectively), with an orthogonal acceleration time-of-flight mass analyzer (oa-TOF). Additionally, **ion mobility mass spectrometry (IM-MS)** has been evaluated, as an alluring platform for glycan isomer separation, at three different levels: intact glycoprotein, glycopeptide and glycans.

Specifically, in this work we have focused on the study of **human transferrin (hTf)** and **mouse transferrin (mTf)** as glycoproteins whose normal glycosylation pattern is suspected to be altered whether due to pathological processes or unhealthy practices. Additionally, **human recombinant erythropoietin (rhEPO)** has been selected as an excellent model glycoprotein, not only due to its relevance as a biopharmaceutical and in the doping field, but also due to its high percentage of glycosylation and microheterogeneity.

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

- Optimization of the digestion procedure of hTf, as a proteolytic resistant glycoprotein, in order to improve the obtained yield and enhance the sensitivity of the method.
- Development and optimization of a CapLC-MS method for the characterization of the glycosylation of hTf glycopeptides. Evaluation of this methodology for the diagnosis and control of chronic alcohol consumption, an unhealthy practice suspected to alter the normal glycosylation pattern of hTf in a certain way, although still in debate in the scientific community.
- Evaluation of principal component analysis (**PCA**) and partial least squares discriminant analysis (**PLS-DA**) as potent exploratory and classificatory mathematical tools for the classification of **congenital disorders of glycosylation (CDG)** on the basis of hTf glycopeptides analysis by the developed CapLC-MS method.

- Identification of the most discriminant glycoforms that allow differentiating between healthy, CDG-I and CDG-II samples, which could become a novel biomarker panel for this disease.
- Application of the previous optimized CapLC-MS method for the study of mTf in mice with **collagen induced arthritis (CIA)**, an inflammatory relevant disease homologous to human **rheumatoid arthritis (RA)** in certain aspects. The alteration in the glycosylation of this protein due to CIA will be evaluated.
- Development and optimization of a CE-MS methodology as an appealing alternative to CapLC-MS for the analysis of the glycosylation of proteins. Optimization of the sample treatment to avoid incompatibility problems with CE when using a surfactant in the digestion of proteolytic resistant glycoproteins.
- Evaluation of the classical semiempirical relationships to study and predict the electrophoretic migration behavior in CE-MS of peptides and glycopeptides, originated from the digestion of rhEPO, which can be considered a model glycoprotein due to its high percentage of glycosylation and microheterogeneity.
- Assessment of the power of IM-MS for the separation of isomeric glycoconjugates differing in the type of sialic acid linkage, using mTf as an example of a glycoprotein, at the intact glycoprotein, glycopeptide and glycan level.

Chapter 1. Introduction

Omic sciences encompass an increasingly large group of diverse disciplines in the field of biology, chemistry or medicine, which aim to comprehend and study the origin, evolution and biologic behavior of all living organisms [1, 2]. The first *omics* to come to life was genomics[a]. The most notable advancement in this extensive field has been the precise mapping and complete sequencing of the human genome, most commonly known as the Human Genome Project [3], which, upon completion, has shifted the focus to new fields. Together with the great advancement in experimental sciences and new instrumentation, there has been a considerable explosion of new *omic* sciences in the last two decades, being transcriptomics[b] or proteomics[c] two of the most important *omic* sciences in terms of advancement and dedication of the scientific community, in comparison to other more recent *omics*, such as nutrigenomics[d] or interactomics[e] [1, 2].

One of the relatively recent *omic* sciences, which has aroused great interest in the last years, is **glycomics**, described as the discipline that aims to study and characterize the carbohydrates in an organism, whether in free form or bound to another molecule, such as proteins or lipids [4]. Of all the known *omic* sciences, glycomics can be considered one of the most convoluted due to the inherent complexity of carbohydrates, which has made advances in this field scarcer in comparison to other *omics* [5]. The intricacy of glycomics studies often lies in the fact that carbohydrates can form complex structures, severely branched, in contrast to more linear assembled biological molecules such as proteins or oligonucleotides. Additionally, the biosynthetic pathways of carbohydrates are still an obscure area and, even nowadays, not fully understood. When focusing on the study of the carbohydrate structure of proteins, glycomics is usually referred to as **glycoproteomics** [4].

[a]Genomics is the study of the entire genome of an organism, i.e., determine their entire deoxyribonucleic acid (DNA) sequence and detailed genetic mapping.
[b]Transcriptomics is the study of the acid ribonucleic or RNA and gene expression.
[c]Proteomics is the study of whole proteins or peptides.
[d]Nutrigenomics is the study of the interactions between genes and nutrients at a molecular level.
[e]Interactomics is the study of the interactions between biomolecules, particularly proteins, and the consequences of those interactions in a biosystem.

1.1. Glycosylation

Glycosylation is the covalent attachment of carbohydrate moieties mainly to lipids or proteins and is considered one of the most common and complex **cotranslational** and **posttranslational** modifications in proteins [6, 7]. The attachment of carbohydrates, also known as **glycans**, to the peptide backbone of proteins can occur while the protein is being synthesized (cotranslational) or after its synthesis (posttranslational). Glycosylated proteins, commonly known as **glycoproteins**, are synthesized in the endoplasmic reticulum and Golgi apparatus [8]. They can be secretory proteins or they can be located inside the cell, mainly in the nucleus or cytosol, and as constituents of the cell membrane.

The presence of carbohydrate chains in proteins alters considerably their physicochemical properties, such as charge, solubility, viscosity or stability; however, the effect is highly variable depending on the glycan or the peptide sequence. It is also worth mentioning that glycosylation of proteins immensely enhances their functional diversity, as the presence of glycans is understood to be involved in a wide variety of biological processes, which range from signaling events to cell-cell interaction, recognition of other organisms or even protein folding [9–11]. In complex organisms, the presence of glycosylated proteins is rather considerable with more than half of all secretory and cellular proteins being glycosylated in mammals.

Glycoproteins are well known for their high **microheterogeneity** [11], which can be defined as the slight variations in the chemical structure of almost identical molecules. Specifically, in glycoproteins, the high microheterogeneity is due to three main reasons:

1. Number of **glycosylation points** or **glycosites** (the amino acid of the peptide backbone where the glycan can be bound).

2. Occupancy of the different glycosites.

3. Structure and composition of the carbohydrate moiety of each glycosite.

The complexity of glycosylation studies can escalate rather quickly when we take into account the diversity of structures the glycans can adopt, from a simple monosaccharide unit to severely branched polysaccharide structures. Additionally, different bonds between monosaccharide residues can be formed or, even, between the carbohydrate moiety and the amino acid of the protein. However, in general, glycan chains are composed of well-known monosaccharide units covalently bound to each other by **glycosidic bonds**, either in the α or β form [11, 12], as depicted in Figure 1.1.

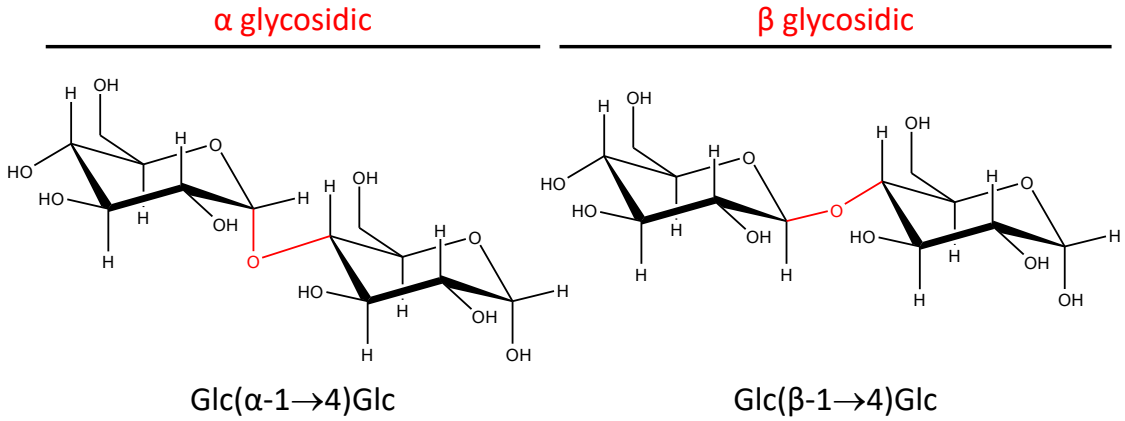
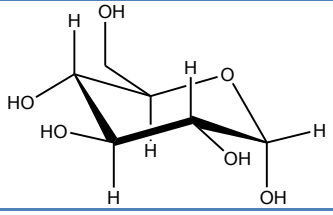
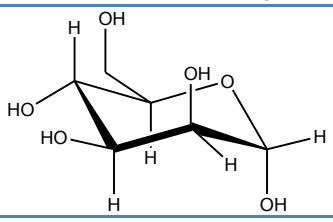
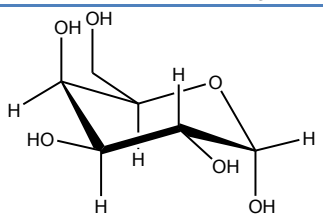
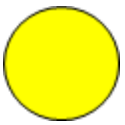
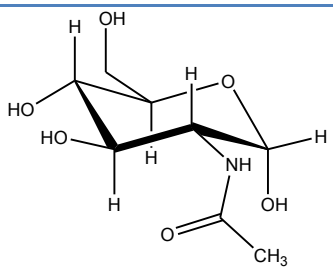
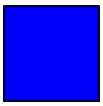
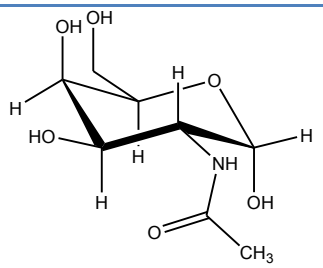
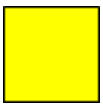
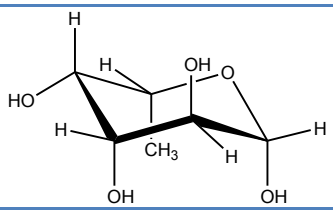


Figure 1.1: Schematic representation of the α and β forms of the glycosidic bond.

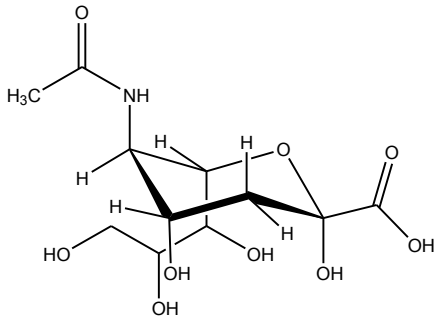
Table 1.1 shows the most common monosaccharide units that can be found in protein glycans, with the standardized symbol nomenclature proposed by the Consortium for Functional Glycomics (CFG) [13].

Table 1.1: Most common monosaccharide units found in protein glycans.

Monosaccharide	Structure	Abbreviation[a]	Symbol
Glucose		**Glc (H)**	
Mannose		**Man (H)**	
Galactose		**Gal (H)**	
N-acetylglucosamine		**GlcNAc (N)**	
N-acetylgalactosamine		**GalNAc (N)**	
Fucose		**Fuc (F)**	

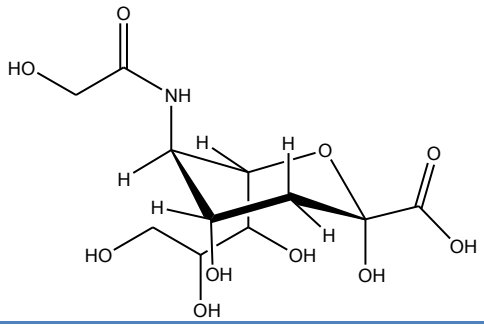
Sialic acids (SiA)

N-acetylneuraminic acid

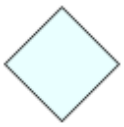


Neu5Ac (S)

N-glycolylneuraminic acid



Neu5Gc (S)



Symbol nomenclature follows the rules of the Consortium for Functional Glycomics (CFG) [13]. [a]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.

Protein glycosylation can be mainly divided in two groups depending on the type of amino acid of the glycosite [11, 14]. In Figure 1.2, the different types of protein glycans are depicted following the symbol nomenclature recommended by the CFG showing an example of each type [13].

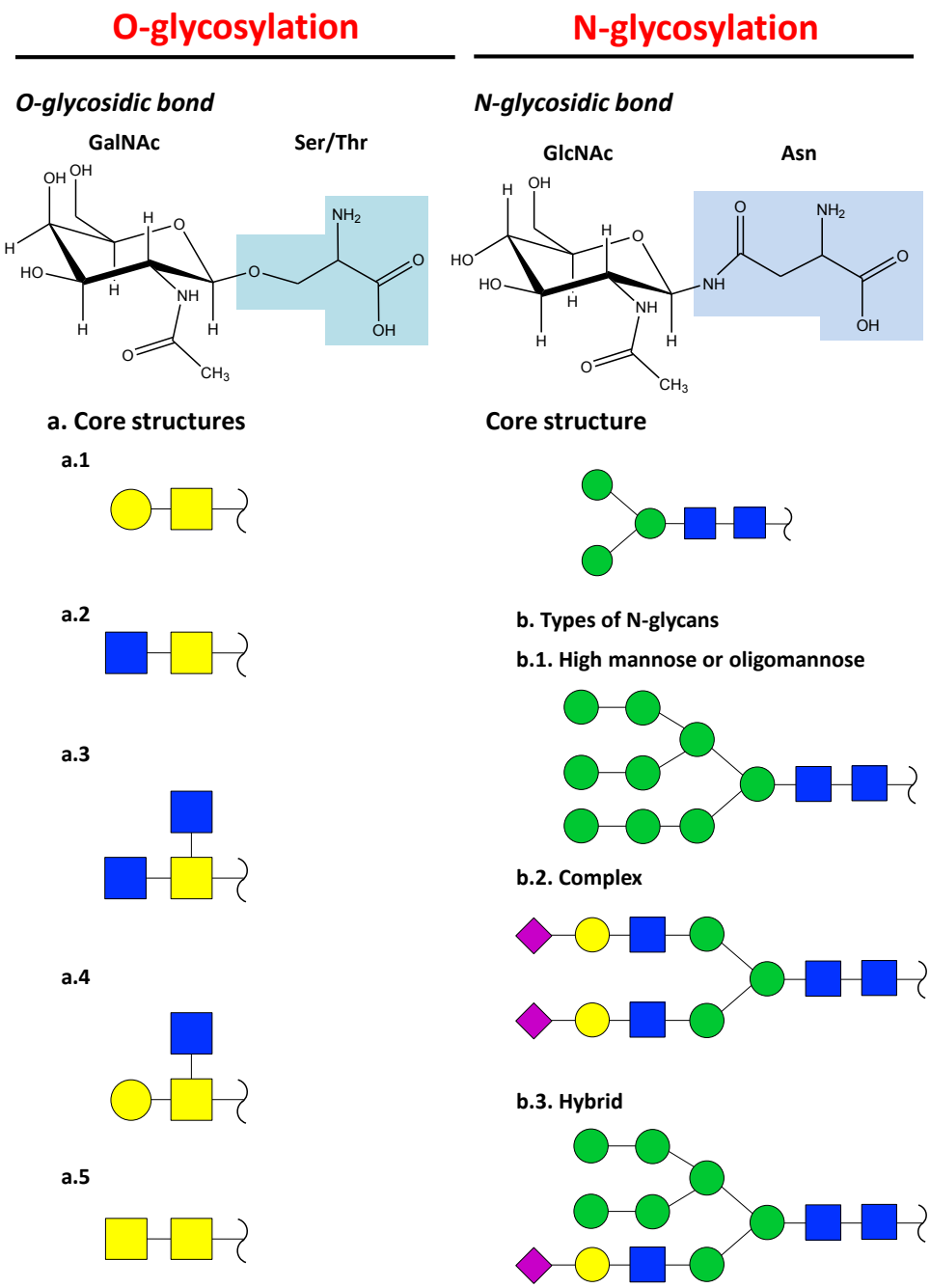


Figure 1.2: Type of glycosidic bonds and type of O- and N-glycan. Symbol nomenclature follows the rules of the Consortium for Functional Glycomics (CFG) [13].

a) **O-glycosylation**: glycans are bound to the oxygen atom of the hydroxyl group of a serine (Ser) or a threonine (Thr). This bond is usually referred to as an **O-glycosidic bond** [15, 16] (see Figure 1.2). O-glycans can be diverse ranging from a single sugar unit to more branched structures. There are several types of O-glycans, being the most common the so-called mucin

O-glycans or O-GalNAc glycans. This type of O-glycans can be classified depending on their central structure, known as **core**. The main core structures in mucin O-glycans are included in Figure 1.2 (a.1-5).

b) **N-glycosylation**: through an N-acetylglucosamine residue, glycans are bound to the nitrogen atom of the amide group of the side chain of an asparagine (Asn) only within a consensus amino acid sequence, Asn-X-Ser/Thr, as long as X is not proline. This union forms the so-called **N-glycosidic bond**, which is also illustrated in Figure 1.2. All N-glycans share a common pentasaccharide core as detailed in Figure 1.2. However, to this core structure, several more monosaccharide units can be bound to form branched structures, giving rise to the three main groups of N-glycans [17]:

i) **High mannose or oligomannose:** only mannose residues are attached to the core structure of the glycan (Figure 1.2 – b.1).

ii) **Complex:** N-acetylglucosamine residues are bound to the mannose of the core, resulting in further branched structures, commonly denominated **antennae**. These antennae are formed by N-acetyllactosamine residues (LacNAc: N-acetylglucosamine plus galactose) that might end with a sialic acid residue (Figure 1.2 – b.2).

iii) **Hybrid:** can be considered a combination of the other two types of N-glycans, as only other mannose residues are bound to the α-1$\rightarrow$6 mannose arm of the core structure giving rise to a high mannose structure, whereas to the α-1$\rightarrow$3 mannose of the core, one or two N-acetylglucosamine can be bound to form a complex type structure, which, after the addition of a galactose residue, can be further extended with additional LacNAc or sialic acid residues (Figure 1.2 – b.3).

Structural diversity of complex N-glycans

Complex N-glycans have been the main type of glycan studied throughout this thesis, reason why their structural diversity will be explained in more detail. This structural variation is summarized in Figure 1.3.

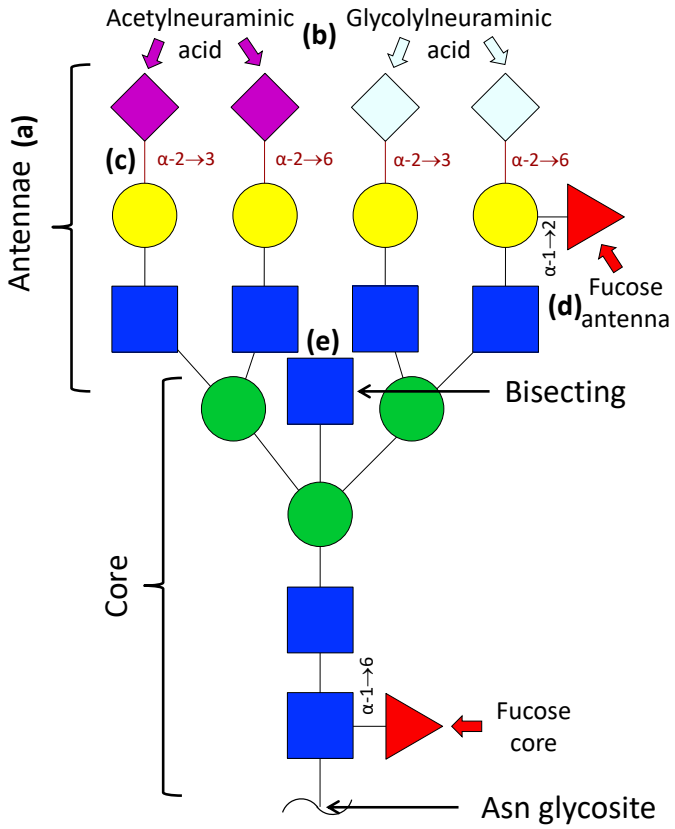


Figure 1.3: Structural diversity of complex N-glycans depicted in a tetraantennary glycan (H7N7F2S4) as an example. The letters in the figure indicate the section of the text where the several causes of this structural variation are explained.

a) Complex N-glycans can form severely branched structures with multiple antennae. Each antenna is often formed by a single LacNAc, however, it can be further extended with the addition of extra LacNAc units. By way of example, a tetraantennary glycan is depicted in Figure 1.3. Commonly, the branching was expressed by the number of antennae of the glycan, which led to their classification as monoantennary (1Ant), biantennary (2Ant), triantennary (3Ant) and so on, usually the number of sialic acids or fucoses was also indicated (e.g.,

2Ant/2SiA1Fuc). However, in the last few years, this nomenclature has fallen into disuse, and has been replaced by a more standardized one that refers to the number of each monosaccharide present in the glycan regardless of the type of glycan (e.g. the 2Ant/2SiA1Fuc glycan would be the H5N4F1S2). For the sake of consistency throughout the thesis, only the novel nomenclature will be used, however, in Figure 1.4, several examples of complex type N-glycans are included expressed in both nomenclatures, most of them appearing during the thesis.

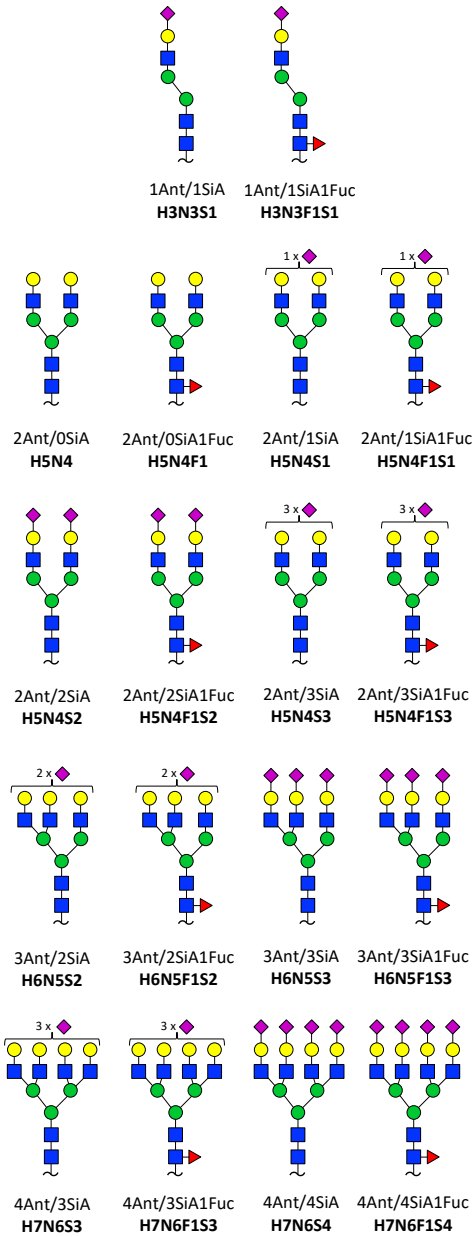


Figure 1.4: Different examples of complex type N-glycans with the former and novel (bold) nomenclature. For the sake of simplicity, the SiA depicted in this picture is only the Neu5Ac.

b) Complex N-glycans can show a sialic acid residue (SiA), also known as neuraminic acid, at the end of the LacNAc unit, as can be observed in Figure 1.3. Of all the monosaccharides that can be present in complex type N-glycans, the sialic acid is the one that can alter considerably the physicochemical properties of the glycan [18, 19]. Sialic acid pKa is around 2.6 due to its carboxylic group (carbon 1 in Figure 1.5). Consequently, at physiological pH, sialic acids are negatively charged, being the monosaccharide that gives the negative charge to the whole glycan. Sialic acids are commonly modified in most glycans, with several natural variations being possible. Quite often carbon 5 can be modified with an N-acetyl group, resulting in the so-called, **N-acetylneuraminic acid** (Neu5Ac), prevalent in humans and birds. Whereas, in other animals, the typical sialic acid is the **N-glycolylneuraminic acid** (Neu5Gc), where a hydroxylated N-acetyl group is added to the neuraminic acid. However, plenty more modifications are quite frequent, such as acetylation, phosphorylation, etc., thus, contributing to the pronounced chemical diversity of sialic acids and their biological relevance [9, 18]. Some of these modifications are illustrated in Figure 1.5.

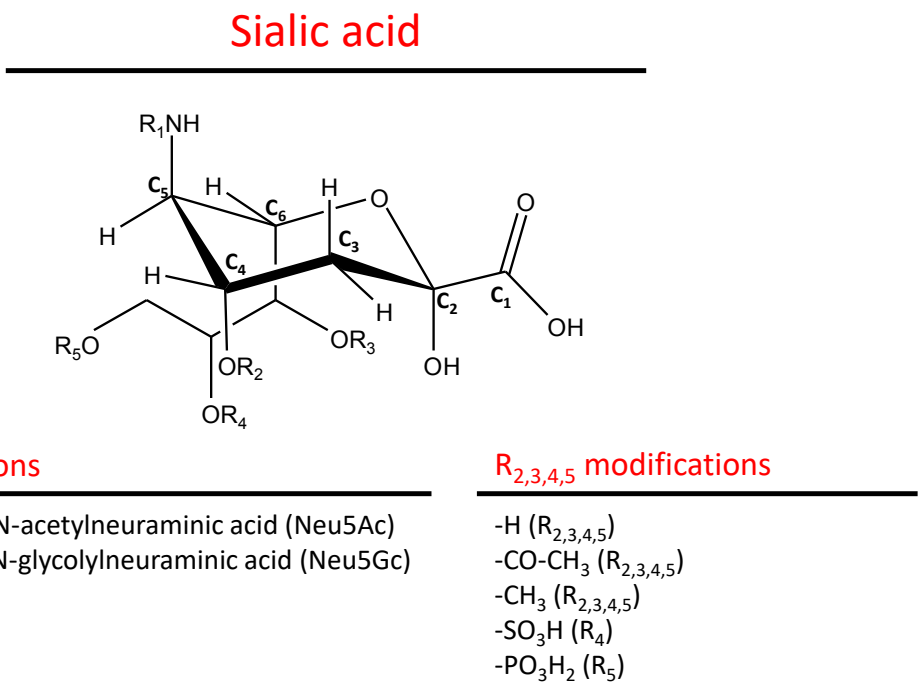


Figure 1.5: Main modifications of the sialic acid (SiA).

c) Additionally, the bond between the several monosaccharide units can be different, which increases even more the heterogeneity of complex N-glycans (see Figure 1.3). For instance, different α linkage might be formed between carbon 2 of the sialic acid and the following monosaccharide residue, usually galactose in complex N-glycans. Specifically, as can be observed in Figure 1.6, the bond with the galactose can take place between carbon 2 of the sialic acid and carbon 3 or 6 of the galactose residue, referred to as **α-2→3** or **α-2→6** linkage, respectively, giving birth to new possible structures, most of them isomers [18].

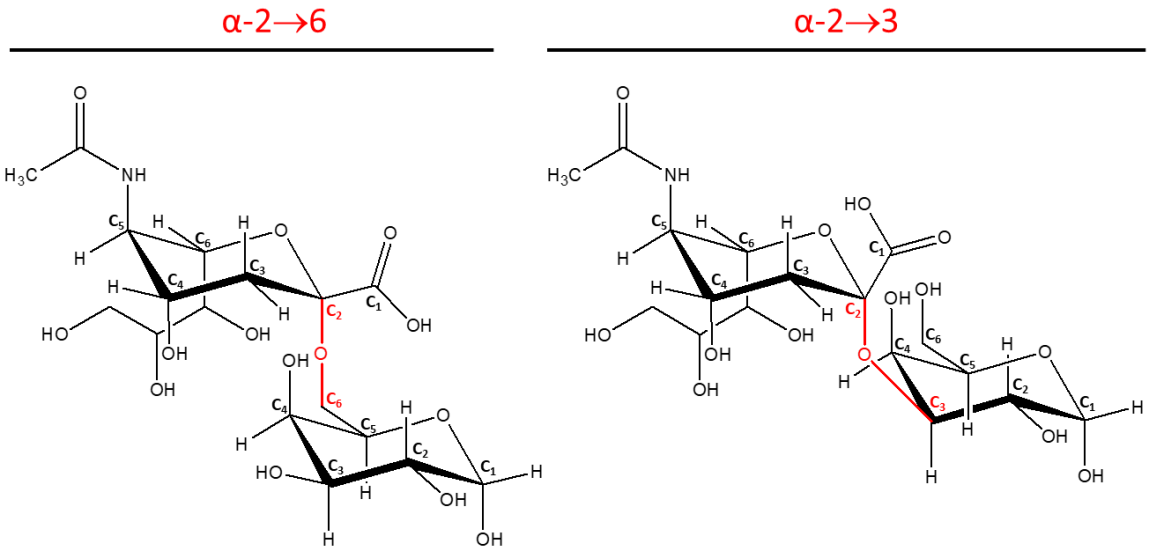


Figure 1.6: Main type of sialic acid linkage in complex type N-glycans.

d) Furthermore, some complex N-glycans may have fucose residues, which can be bond to the glycan core, usually to the first GlcNAc residue (fucose core) or, alternatively, to the GlcNAc or Gal residue of the antenna (fucose antenna) as shown in Figure 1.3.

e) Finally, an additional GlcNAc residue can be bound to the central mannose of the core. In such cases, the glycan is also categorized as **bisecting** (see Figure 1.3). To this extra GlcNAc, a Gal residue can be bound, which, in turn, can be further extended with additional LacNAc units or SiA residues.

All the aforementioned variations that are possible in complex N-glycans contribute to increase the microheterogeneity of glycoproteins and are usually the main obstacle the scientific community faces in glycosylation studies. A given glycoprotein exists as a heterogeneous mixture of glycoconjugates species known as **glycoforms** that differ in the amount, size and structure of their glycans or the amino acid to which they are bound. Therefore, a single glycoprotein can have multiple glycoforms, ranging from less than five to even more than one hundred different glycoforms (e.g. recombinant human erythropoietin [20]).

1.2. Glycoproteins as biomarkers and biopharmaceuticals

Nowadays, the analysis and characterization of the glycosylation pattern of a protein is of great importance as glycosylation is responsible for a wide variety of biological processes and is vital for the proper function of the glycoprotein. However, various biological or pathological processes are understood to alter the normal glycoform composition of a glycoprotein, thus, affecting the normal body functions, which often leads to serious medical complaints or, ultimately, to the death of the individual. Additionally, recombinant and endogenous glycoproteins quite often differ in their glycosylation pattern, as it is dependent on the host cells that synthetize the glycoprotein [21, 22]. Furthermore, the culture media or purification methods used for production of the recombinant glycoprotein can also affect its glycosylation [22, 23]. As recombinant glycoproteins are frequently used as biopharmaceuticals, characterization of their glycosylation is crucial to properly asses their stability, biological activity and pharmacokinetic [24, 25].

Consequently, the work presented in this thesis is focused on the development and optimization of analytical methodologies in the glycoproteomic field, whose two main purposes would be: a) the analysis of certain glycoproteins whose glycosylation is known to be altered, whether due to pathological processes or unhealthy practices, in order to characterize

their glycosylation pattern and identify new or altered glycoforms as potential novel biomarkers; b) the analysis of recombinant glycoproteins that are relevant biopharmaceuticals and whose glycosylation pattern must be properly characterize to ensure their quality and proper function. Specifically, in this thesis, human transferrin (hTf) and mouse transferrin (mTf) have been studied as relevant glycoproteins involved in some pathological processes or unhealthy practices. Whereas, recombinant human erythropoietin (rhEPO) has been selected as a model glycoprotein to develop and optimize analytical methodologies in glycoproteomic studies, not only due to its high microheterogeneity, but also because of its great interest as biopharmaceutical and in the doping field.

1.2.1. Human Transferrin

Human transferrin (hTf) is an iron binding serum glycoprotein of approximately 80kDa of molecular mass, whose main function is the transport of iron through the blood plasma [26]. hTf is a slightly glycosylated protein, with a glycosylation degree of approximately 5.8%. Three glycosites are responsible for hTf glycosylation: one O-glycan with one hexose unit at serine 32 (indicated as O_{32}), and two complex type N-glycans attached to asparagines 413 and 611 (indicated as N_{413} and N_{611}, respectively) of the peptide backbone. Figure 1.7 shows the peptide sequence of hTf with its O- and N-glycosites marked in red and blue, respectively.

hTf

1		V	PDKTVRWCAV	SEHEATKCQS	FRDHMKSVIP	SDGPSVACVK
42	KASYLDCIRA	IAANEADAVT	LDAGLVYDAY	LAPNNLKPVV	AEFYGSKEDP	QTFYYAVAVV
102	KKDSGFQMNQ	LRGKKSCHTG	LGRSAGWNIP	IGLLYCDLPE	PRKPLEKAVA	NFFSGSCAPC
162	ADGTDFPQLC	QLCPGCGCST	LNQYFGYSGA	FKCLKDGAGD	VAFVKHSTIF	ENLANKADRD
222	QYELLCLDNT	RKPVDEYKDC	HLAQVPSHTV	VARSMGGKED	LIWELLNQAQ	EHFGKDKSKE
282	FQLFSSPHGK	DLLFKDSAHG	FLKVPPRMDA	KMYLGYEYVT	AIRNLREGTC	PEAPTDECKP
342	VKWCALSHHE	RLKCDEWSVN	SVGKIECVSA	ETTEDCIAKI	MNGEADAMSL	DGGFVYIAGK
402	CGLVPVLAEN	YNKSDNCEDT	PEAGYFAIAV	VKKSASDLTW	DNLKGKKSCH	TAVGRTAGWN
462	IPMGLLYNKI	NHCRFDEFFS	EGCAPGSKKD	SSLCKLCMGS	GLNLCEPNNK	EGYYGYTGAF
522	RCLVEKGDVA	FVKHQTVPQN	TGGKNPDPWA	KNLNEKDYEL	LCLDGTRKPV	EEYANCHLAR
582	APNHAVVTRK	DKEACVHKIL	RQQQHLFGSN	VTDCSGNFCL	FRSETKDLLF	RDDTVCLAKL
642	HDRNTYEKYL	GEEYVKAVGN	LRKCSTSSLL	EACTFRRP		

Red: O-glycosylated Blue: N-glycosylated

Figure 1.7: Human transferrin (hTf) amino acid sequence (UniProtKB / Swiss-Prot Accession Number: P02787).

If we take into account the microheterogeneity associated with complex type N-glycans, hTf exists as a mixture of glycoforms, each one with a different glycan composition or structure. Figure 1.8 shows the major glycoforms present in intact hTf. In spite of having two glycosylation points, hTf has only four major glycoforms. The main glycoform is the tetrasialoform (also known as **S4**) which presents two disialylated biantennary glycans (H5N4S2), one in each glycosylation point. Therefore, this glycoform shows 4 terminal sialic acids (N-acetylneuraminic acids). In a healthy individual, other less abundant glycoforms are present: the hexasialoform (**S6**) with two trisialylated triantennary glycans (H6N5S3); the pentasialoform (**S5**) with one disialylated biantennary (H5N4S2) and one trisialylated triantennary glycans (H6N5S3); the trisialoform (**S3**) with one disialylated (H5N4S2) and one monosialylated biantennary glycans (H5N4S1) and the disialoform (**S2**) with a single disialylated biantennary glycan (H5N4S2) in one of the glycosites. S2 is often considered a minor glycoform and included, together with the **S1** and **S0** glycoforms (both with less than 2 sialic acid residues and non-existent in a healthy individual), in the so-called carbohydrate deficient transferrin (**CDT**) glycoforms [27–29].

When hTf is digested with a protease like trypsin, three different glycopeptides are obtained. The O-glycopeptide has only one possible glycoform, but the two N-glycopeptides can show several different glycoforms depending on the intact protein glycoform they are derived from. Figure 1.8 also shows the main glycoforms of hTf glycopeptides and the intact glycoform from which they are derived.

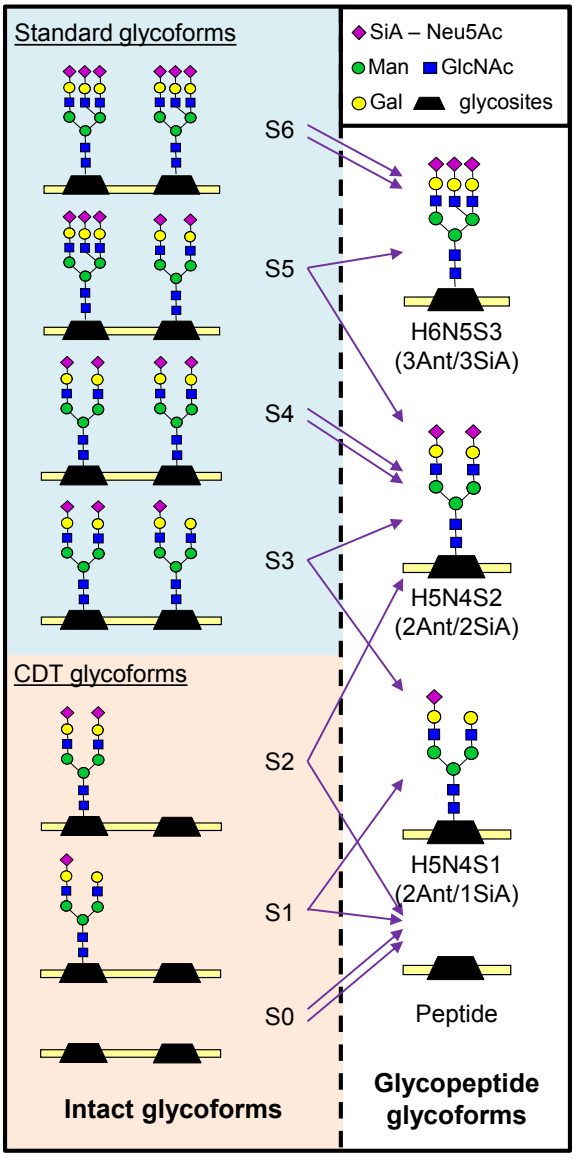


Figure 1.8: Main human transferrin (hTf) glycopeptide glycoforms and the intact protein glycoform from which they are derived. Glycan symbol nomenclature follow the rules recommended by the Consortium for Functional Glycomics (CFG) [13].

hTf glycosylation pattern has been described to be altered in certain unhealthy practices and diseases, such as **chronic alcoholism** and **congenital disorders of glycosylation (CDG)** [30–37]. Additionally, hTf is one of the so-called acute phase proteins, whose concentration is reduced (negative acute phase proteins) in the acute phase of an inflammatory process [38–40] and whose glycosylation is understood to be altered in certain inflammatory diseases [41].

Chronic alcoholism could be defined as the prolonged and uncontrollable intake in large amounts of alcohol-containing products. Chronic alcoholism, considered a modern severe health issue, can lead to a premature death, mainly due to several medical complaints such as liver cirrhosis, internal bleeding, ethylic intoxication, hepatocellular carcinoma, accidents or suicidal behavior [42]. In those individuals with a high intake of ethanol, CDT glycoforms are increased in comparison to a healthy individual, reason why, these glycoforms are considered reliable biomarkers and are, nowadays, the best indicators of chronic alcohol abuse [32, 43, 44]. The mechanism behind the abnormality of hTf glycosylation in chronic alcoholism is still not completely understood [31–33]. Some authors have suggested that ethanol is responsible for the inhibition of galactosyltransferase and sialyltransferase, two important enzymes in the second phase of N-linked glycosylation. Additionally, ethanol is also understood to stimulate the activity of sialidases. Both effects would suggest that ethanol disrupts the synthesis of N-glycans, giving rise to truncated or non-sialylated glycoforms [31, 33, 45, 46]. On the other hand, some authors have implied that ethanol intake completely impedes the initiation of N-linked glycosylation. Ethanol has been suggested to greatly reduce the concentration of dolichol phosphate (Dol-P) and of adenosine triphosphate (ATP). As Dol-P and ATP availability is of great importance for the first phase of N-linked glycosylation, chronic alcoholism is supposed to limit the synthesis of N-linked glycoproteins, thus, an increase of non-glycosylated glycoforms is expected [31, 32, 47, 48]. Although still a matter of debate, the effects that ethanol causes in the first phase of the N-linked glycosylation are often considered predominant [32]. In any case, improvements in the detection and characterization of hTf

glycoforms are mandatory not only for the early diagnosis of chronic alcoholism, but also to monitor the rehabilitation process of alcohol abusers and shed some light on the mechanism behind the alteration of hTf due to ethanol intake.

CDG are a rapidly growing group of genetic diseases that, to date, encompass more than 80 different known disorders. As an ever-expanding family, there are several variants of CDG and new ones are discovered every year, each affecting a different gene [49]. In the majority of CDG, the nervous system is involved, leading to developmental disability, hypotonia, hyporeflexia, ataxia, growth retardation and possible damage in several organs, such as liver, heart or the immune system, among many others [50–52]. Unfortunately, the majority of CDG patients die at a very young age due to the several aforementioned complaints. CDG were classified until recently in two subgroups, CDG type I (**CDG-I**) and CDG type II (**CDG-II**), depending on the defective glycosylation pathway of the N-glycosylation of proteins [36, 53, 54]. Specifically, CDG-I encompassed all defects in the glycan assembly or transfer to the nascent glycoprotein in the cytosol or the endoplasmic reticulum, which led to the lack of complete N-glycans on some glycosites [49]. Whereas, glycosylation defects caused by the abnormal remodeling or processing of the glycan moieties in the Golgi network, generally resulting in truncated or structural deficient carbohydrate chains, were considered CDG-II [55, 56]. However, novel discoveries of new types of CDG affecting not only the N-glycosylation but other glycosylation pathways, such as the O-glycosylation of proteins, have rendered this classification obsolete. Consequently, this classification has been replaced by a much needed standardized nomenclature, which associates the modified gene or genes with the suffix ‘CDG’ [57]. However, when referring to CDG affecting the N-glycosylation pathway of proteins, the former nomenclature (CDG-I or CDG-II) is still used.

Table 1.2 shows some of the main diagnosed CDG affecting the N-glycosylation pathway, their classification in the former CDG type I or CDG type II groups and the affected gene.

Additionally, CDG that only affect the N-glycosylation pathway of proteins are in the blue section of the table, whereas those CDG that affect other pathways apart from the N-glycosylation of proteins are in the orange section.

Table 1.2: Several of the main diagnosed congenital disorders of glycosylation (CDG) affecting the N-glycosylation pathway, their classification in the former CDG type I or CDG type II groups and the affected gene. The information in this table is a compilation from different sources [35, 57–60].

CDG name	Affected gene	CDG type
PMM2-CDG	Phosphomannomutase 2	CDG-I
MPI-CDG	Mannose phosphate isomerase	CDG-I
RFT1-CDG	RFT1 (requiring fifty three 1) homolog	CDG-I
DPAGT1-CDG	dolichyl-phosphate N-acetylglucosaminephosphotransferase 1	CDG-I
ALG6-CDG	Alpha-1,3-glucosyltransferase I	CDG-I
ALG3-CDG	Alpha-1,3-mannosyltransferase VI	CDG-I
ALG12-CDG	Alpha-1,6-mannosyltransferase VIII	CDG-I
ALG8-CDG	Alpha-1,3-glucosyltransferase II	CDG-I
ALG1-CDG	Chitobiosyldiphosphodolichol beta-mannosyltransferase I	CDG-I
ALG9-CDG	Alpha-1,2-mannosyltransferase	CDG-I
MGAT2-CDG	mannosyl (alpha-1,6-)-glycoprotein beta-1,2-N-acetylglucosaminyltransferase	CDG-II
GCS1-CDG	mannosyl-oligosaccharide glucosidase	CDG-II
DPM1-CDG	Dolichyl-phosphate mannosyltransferase subunit 1	CDG-I
MPDU1-CDG	Mannose-P-dolichol utilization defect 1	CDG-I
PGM1-CDG[a]	phosphoglucomutase 1	CDG-I
DOLK-CDG	Dolichol kinase	CDG-I
SRD5A3-CDG	Steroid 5 alpha-reductase 3	CDG-I
COG8-CDG	Component of oligomeric golgi complex 8	CDG-II
SLC35C1-CDG	Solute carrier family 35 member C1	CDG-II
B4GALT7-CDG	Beta-1,4-galactosyltransferase 7	CDG-II
ATP6V0A2-CDG	ATPase H+ transporting V0 subunit a2	CDG-II

[a]This CDG was formerly known as CDG-I_t but, nowadays, there is still no consensus on how it should be categorized.

Like in chronic alcoholism, CDT glycoforms are increased in patients with a CDG that affects the N-glycosylation pathway when compared to a healthy individual. Therefore, the development of an analytical methodology to study the glycosylation of hTf and discover reliable biomarkers for this disease could be of great importance, not only for the early diagnosis and control of CDG, but also to look for a treatment to alleviate or eliminate the symptoms.

Nowadays, there are several reference methods used for the analysis of CDT glycoforms to diagnose chronic alcoholism and CDG. These methods usually involve the use of isoelectric focusing (IEF), capillary electrophoresis (CE) or anion exchange liquid chromatography (LC) for the analysis of hTf intact glycoforms [29, 36, 61–66]. However, these routine methods present several disadvantages, which make the development of new methodologies an imperative task. Specifically, IEF is an arduous and time-consuming technique not suitable for automation. Whereas, in CE and LC, ultraviolet (UV) is used for the detection of hTf intact glycoforms, which does not allow their reliable characterization. Therefore, developing a straightforward methodology able not only to rapidly separate hTf glycoforms but also unequivocally identify them, for instance, with the use of a mass spectrometer, would be essential for the correct diagnosis and control of these diseases.

hTf is also involved in the acute phase of an inflammatory process [38–40]. Inflammation is an adaptive and complex biological response to a harmful stimulus, whether due to endogenous or exogenous inducers, such as, infection, damaged cells, physical trauma, burns, and surgery, among many others [38, 67]. Many systemic physiological and biochemical changes occur when an inflammatory response is activated in the organism, for instance, the concentration of the so-called acute phase proteins can vary or even the glycosylation of certain proteins can be altered. Several authors have shown that an increase in branching (increase of triantennary and tetraantennary glycoforms), sialylation and fucosylation in the N-glycosylation of serum glycoproteins occur when an inflammatory response is triggered [38, 40, 68–71]. In the case of

hTf, an increase of highly sialylated glycoforms with a concomitant decrease in those less sialylated glycoforms has been described in patients with human rheumatoid arthritis, a chronic autoimmune disease [41]. Hence, developing reliable analytical methodologies able to identify and characterize hTf glycosylation could provide a novel insight into the mechanism behind the abnormal glycosylation of this protein in several diseases. These methodologies would also be helpful for the identification of new potential glycan-based biomarkers for these pathologies.

1.2.2. Mouse Transferrin

Mouse transferrin (mTf) is also an iron binding glycoprotein of about 80 kDa, responsible for the transport of iron through the blood plasma, which shares approximately 72% of its peptide sequence with hTf. However, unlike hTf, mTf only has one N-glycosite at asparagine 494, with complex type N-glycans. Hence, its microheterogeneity is somewhat reduced compared to hTf. Moreover, mTf only presents N-glycolylneuraminic acid instead of N-acetylneuraminic acid. Figure 1.9 shows the peptide sequence of mTf, with the N_{494} glycosite marked in blue.

mTf

1	VPDKTVKWCA	VSEHENTKCI	SFRDHMKTVL	PPDGPRLACV	KKTSYPDCIK	AISASEADAM
61	TLDGGWVYDA	GLTPNNLKPV	AAEFYGSVEH	PQTYYYAVAV	VKKGTDFQLN	QLEGKKSCHT
121	GLGRSAGWVI	PIGLLFCKLS	EPRSPLEKAV	SSFFSGSCVP	CADPVAFPKL	CQLCPGCGCS
181	STQPFFGYVG	AFKCLKDGGG	DVAFVKHTTI	FEVLPEKADR	DQYELLCLDN	TRKPVDQYED
241	CYLARIPSHA	VVARKNNGKE	DLIWEILKVA	QEHFGKGKSK	DFQLFSSPLG	KDLLFKDSAF
301	GLLRVPPRMD	YRLYLGHNYV	TAIRNQQEGV	CPEGSIDNSP	VKWCALSHLE	RTKCDEWSII
361	SEGKIECESA	ETTEDCIEKI	VNGEADAMTL	DGGHAYIAGQ	CGLVPVMAEY	YESSNCAIPS
421	QQGIFPKGYY	AVAVVKASDT	SITWNNLKGK	KSCHTGVDRT	AGWNIPMGML	YNRINHCKFD
481	EFFSQGCAPG	YEK**N**STLCDL	CIGPLKCAPN	NKEEYNGYTG	AFRCLVEKGD	VAFVKHQTVL
541	DNTEGKNPAE	WAKNLKQEDF	ELLCPDGTRK	PVKDFASCHL	AQAPNHVVVS	RKEKAARVKA
601	VLTSQETLFG	GSDCTGNFCL	FKSTTKDLLF	RDDTKCFVKL	PEGTTPEKYL	GAEYMQSVGN
661	MRKCSTSRLL	EACTFHKH				

Blue: N-glycosylated

Figure 1.9: Mouse transferrin (mTf) amino acid sequence (UniProtKB / Swiss-Prot Accession Number: Q921I1).

Working with mice models of human diseases can be quite alluring for proteomic and glycomic studies, due to the fact that mice are cost effective as they are small, have the ability to multiply quickly and are easy to look after [72]. Furthermore, mice share several characteristics with humans in terms of anatomy, physiology and genetics. Additionally, samples can be collected under standard operating protocols while the onset and progression of the disease can be closely monitored in well-defined experimental settings [73, 74]. Specifically, **collagen induced arthritis (CIA)** in mice resembles human rheumatoid arthritis not only in terms of its disease course or histological findings, but also in its response to commonly used anti-arthritic drugs [75, 76]. Consequently, CIA in mice can be considered an appealing option to assess the role of certain genes in the induction and/or maintenance of arthritis, to study the efficacy of novel drugs and to evaluate the glycosylation changes derived from this disease.

As previously mentioned, inflammatory diseases have been described to affect the glycosylation pattern expected for certain glycoproteins, in particular, an increase in fucosylation, sialylation and/or glycan branching [71, 77]. This increase in highly sialylated glycoforms has been observed for hTf in patients with rheumatoid arthritis [41], and, as CIA shares multiple characteristics with this disease, a similar alteration in the glycosylation pattern of mTf could be expected.

1.2.3. Recombinant Human Erythropoietin

Recombinant human erythropoietin (rhEPO), also known as **epoetin**, encompasses a group of synthetic glycoproteins produced in mammalian cells using complementary DNA technology. There are several types of rhEPO, most of them are commonly expressed in Chinese hamster ovary cells (CHO) [78], being epoetin alpha and beta the first ones to be synthesized, and also the most used as biopharmaceuticals. In the recent years, new rhEPO have been synthetized with different degrees of glycosylation aimed at improving its pharmacological or physicochemical properties. rhEPO has been widely used in the last decades as a treatment for

anemia, whether from chronic renal failure, cancer, HIV/AIDS or other diseases [79, 80]. However, this recombinant glycoprotein has also been the source of countless disputes in Olympics and top-level sports, due to its misuse by some athletes as a doping agent in several endurance sports, mainly to enhance their performance [80, 81].

rhEPO is a highly glycosylated protein of about 30 kDa and approximately 40% of its weight is due to its glycosylation. rhEPO shows three N-glycosites with complex type N-glycans, at asparagines 24, 38 and 83, and only one O-glycosite in serine 126 [20, 82, 83]. Figure 1.10 shows the peptide sequence of rhEPO, as well as its O- and N-glycosites marked in red and blue, respectively.

rhEPO

1	APPRLICDSR	VLERYLLEAK	EAENITTGCA	EHCSLNENIT	VPDTKVNFYA	WKRMEVGQQA
61	VEVWQGLALL	SEAVLRGQAL	LVNSSQPWEP	LQLHVDKAVS	GLRSLTTLLR	ALGAQKEAIS
121	PPDAASAAPL	RTITADTFRK	LFRVYSNFLR	GKLKLYTGEA	CRTGDR	

Red: O-glycosylated Blue: N-glycosylated

Figure 1.10: Recombinant human erythropoietin (rhEPO) amino acid sequence (UniProtKB / Swiss-Prot Accession Number: P01588).

Besides its high interest as a biopharmaceutical and its importance in the doping field, rhEPO can be considered an excellent model glycoprotein to develop and optimize analytical methodologies in glycoproteomic studies due to its high microheterogeneity. In fact, our research group has already developed several capillary electrophoresis mass spectrometry (CE-MS) methods with the aim of characterizing rhEPO glycosylation, not only by the analysis of the intact glycoprotein but also analyzing the obtained glycopeptides, after the digestion with trypsin or endoproteinase Glu-C [20, 82–84]. In this thesis, some of these methodologies have been used to study and model the migration behavior of glycopeptides and peptides in CE-MS using rhEPO as a model glycoprotein.

1.3. Glycoprotein analysis

1.3.1. Glycoprotein purification techniques

In order to study the glycosylation of proteins present in biological fluids, isolation of the glycoprotein of interest from the rest of the sample components is mandatory. Affinity methodologies for the purification and enrichment of glycoproteins are being developed continuously, mainly due to the great importance of glycosylated proteins and the consequent high demand in glycosylation analysis [85]. Some of these purification techniques include the use of lectins, titanium dioxide (TiO_2) or antibodies. A concise illustration of the different approaches can be found in Figure 1.11.

Lectins are an extensive family of proteins of non-immune origin obtained from natural sources, mainly used for their high binding capacity to both free glycans and glycans attached to glycoconjugates [86, 87]. Their specificity can be rather intricate, as they can recognize not only certain monosaccharides within the glycan chain but also entire glycans or specific fragments, and even different linkages between monosaccharides or glycan branching [88, 89]. Quite often, lectins are immobilized on a stationary support, giving rise to the so-called lectin affinity chromatography. In the last few years, advances have been made in regards to the type of lectin, the methods used in their purification or regarding the stationary supports, resulting in an increase of commercially available lectins and production of state-of-the-art multilectin affinity columns [90, 91].

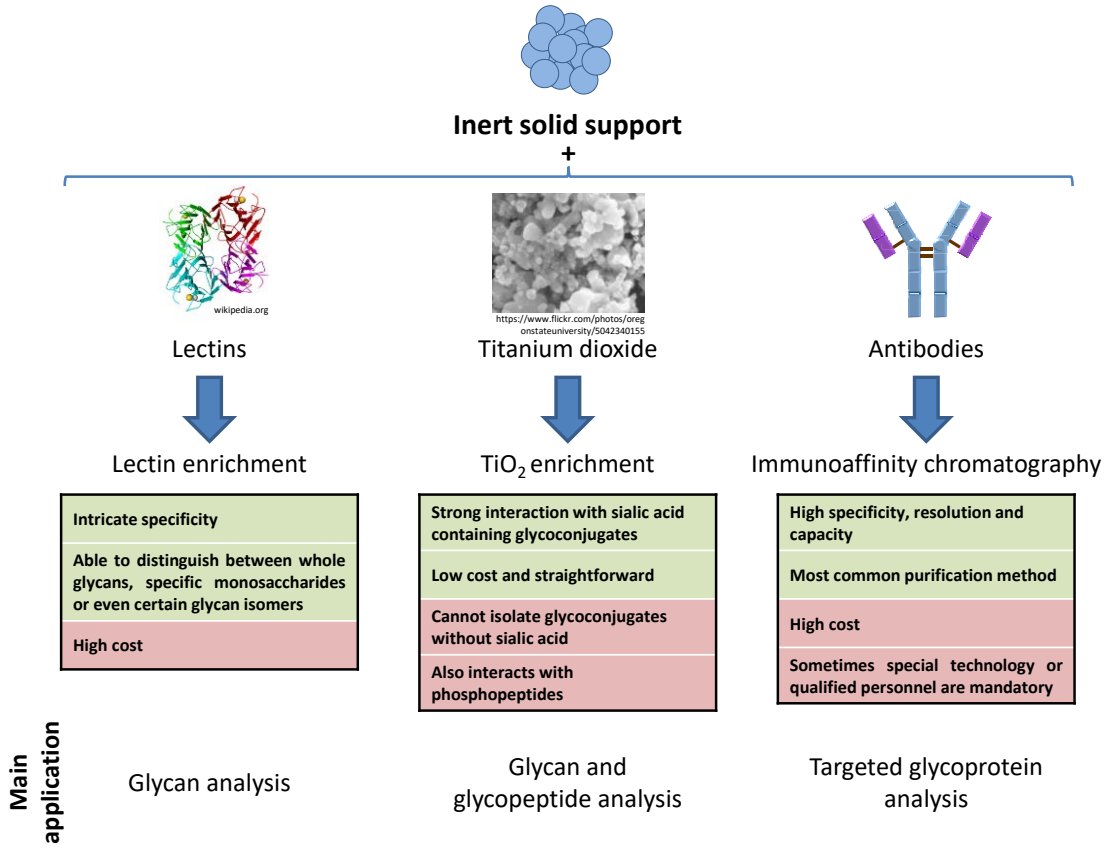


Figure 1.11: Different approaches for the isolation of glycoconjugates and main advantages, drawbacks and applications.

The use of TiO_2 particles is also one of the most promising and noteworthy methods for the purification and enrichment of glycoproteins. This highly specific chemo-affinity solid phase has been demonstrated to have high affinity toward negatively charged compounds. Therefore, it shows an excellent ability to isolate glycoconjugates with sialic acids and has been reported in numerous works [85, 92–94]. Nowadays, similarly to lectins, there are several types of stationary phases for glycoprotein purification with TiO_2, ranging from magnetic particles to chromatographic stationary phases for highly specific columns. One of the main reasons behind their notable affinity is the fact that sialic acids can form a multipoint binding to TiO_2 similar to a multidentate binding, with strong ion exchange and ligand exchange behavior [95, 96]. However, their main advantage is also their principal drawback, as they are only able to isolate sialic acid-rich glycoconjugates.

Finally, in the context of glycoprotein purification, immunoaffinity chromatography (IAC) is one of the most recurrent and powerful methods for targeted glycoprotein analysis, due to its high capacity and specificity [97, 98], being widely used for analytical, clinical and diagnostic purposes. The fundamental behind the effective isolation of a molecule is based on the reversible interaction between the target molecule in a mobile phase and an antibody (Ab), also known as immunoglobulin (Ig), attached to a solid support, creating the so-called stationary phase [97–99]. In theory, any molecule should be effectively purified by IAC as long as the bond with an antibody is possible [100]. The main reason behind that assumption is that the binding antibody-antigen is rather intense as a result of different types of non-covalent interactions [97, 101, 102].

A schematic representation of the steps required to isolate a target molecule by IAC can be found in Figure 1.12. Briefly, the antibody is covalently coupled to the inert solid support to form the stationary phase, which is subsequently transferred and packed in a column. Afterwards, the target antigen, which is in solution in a concentration and condition that favors interaction, is passed through the column. After antigen capture, unwanted antigens that might bind in a non-specific manner are removed with a washing step, to finally release the purified antigen under certain conditions that favor desorption.

With few and simple steps, high purification and simultaneous concentration of the target molecule can be achieved by IAC [102]. For this reason, IAC has been the main affinity technique used for the purification of glycoproteins in this thesis.

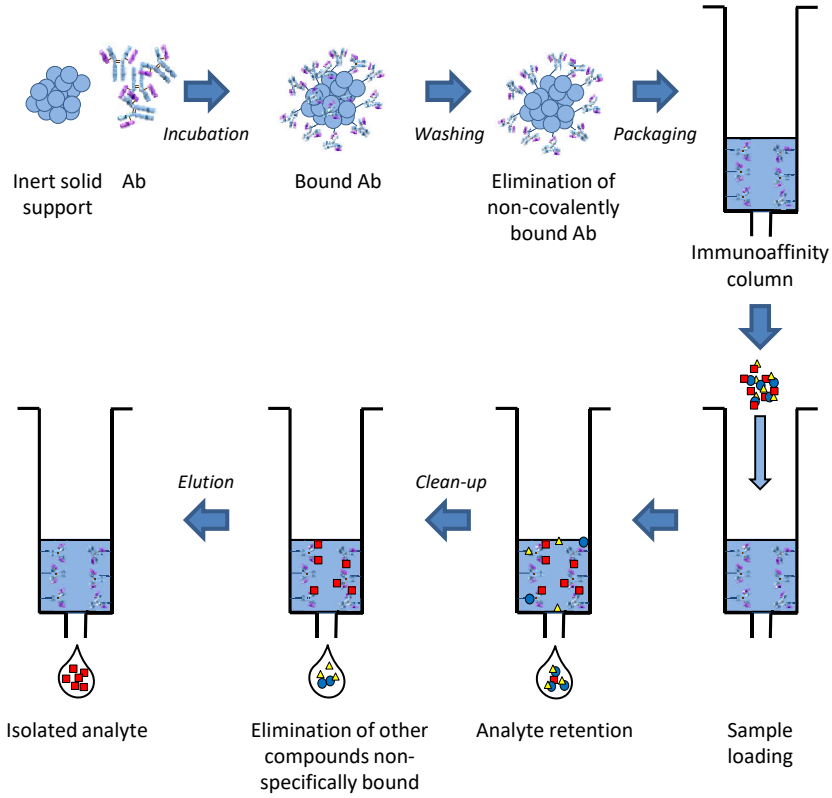


Figure 1.12: Schematic representation of the procedure to isolate a target molecule by immunoaffinity chromatography (IAC).

1.3.2. Mass spectrometry

Before the introduction of **mass spectrometry (MS)**, glycosylation studies were carried out using fluorescence or ultraviolet (UV) detection. However, with those techniques, reliable structural information and unequivocal identification of glycoconjugates was not possible. Nowadays, MS has become the prime option in glycoproteomic studies as it allows high sensitivity profiling and accurate characterization of heterogeneous glycan structures [103–106].

The study of the glycosylation of a certain protein by MS can be carried out at three different levels, each one with their own advantages and drawbacks [103, 107]. A schematic representation of the different approaches can be found in Figure 1.13.

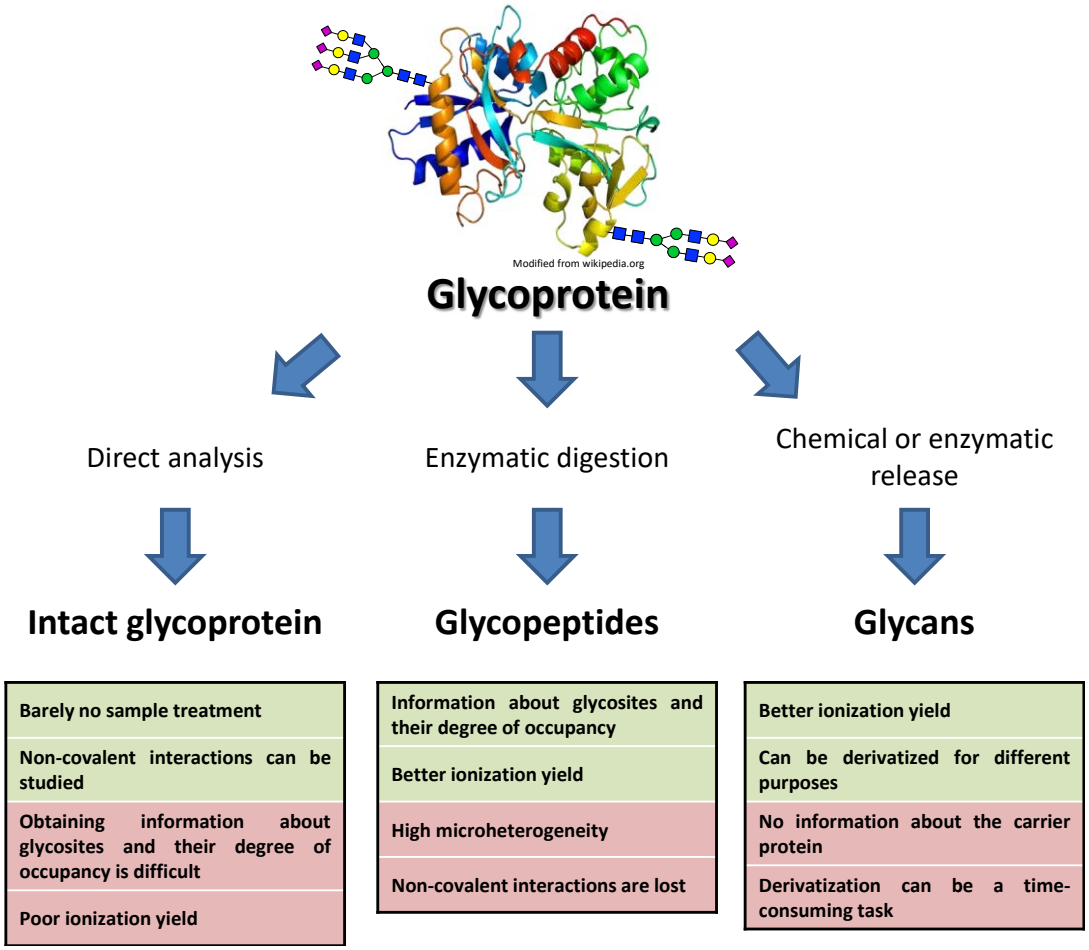


Figure 1.13: Three main approaches for the analysis of protein glycosylation using mass spectrometry detection.

One of these approaches consists in the analysis of the **intact glycoprotein**, which of course, is more direct and considerably faster than the other two alternatives, as no or barely any sample treatment is necessary. However, working with the intact glycoprotein can be quite challenging due to the fact that detection by MS is rather difficult as the ionization of large molecules, such as glycoproteins, is not an easy task [107–109]. Some authors have proposed a top-down approach for the analysis of intact proteins [110–112], although, in the case of glycoproteins, obtaining information about glycosites and their degree of occupancy is almost impossible, due to the limitation of the actual mass spectrometers and the high microheterogeneity of this type of proteins.

The other two approaches involve the use of enzymes or specific reagents to obtain glycosylated structures of lower molecular mass. Hence, following this approach, sample preparation is lengthier and information about non-covalent interactions between the glycoprotein and other molecules cannot be obtained. One of these two alternative approaches consists in the analysis of **glycopeptides** (glycosylated peptides), obtained after the digestion of the intact glycoprotein with a protease[a]. There are several different proteases, each one with their own efficiency and specificity, being trypsin widely employed in a multitude of different studies. The characterization of glycopeptides as low molecular mass glycomarkers allows not only to characterize the glycans of the glycoprotein but also to obtain information about glycosylation sites and their degree of occupancy [113].

The other path to study protein glycosylation is the analysis of the **glycans** released from the intact protein, by chemical release or by enzymatic digestion using O- and N-glycosidases[b]. The release of glycans can be even a more time-consuming task than the protease digestion of the glycoprotein and information of glycosites or the carrier protein is lost. However, isolation of glycans from other components of the sample, such as the peptides from the digest, is relatively easier, as the polarity of glycans is considerably different compared to peptides or the deglycosylated protein. By isolating the glycans, the sensitivity obtained in mass spectrometry can be considerably improved. Additionally, glycans can be derivatized, mainly by permethylation or by reductive amination, to improve their ionization efficiency, their quantification or even to obtain information about the structure of isomeric glycans [114–116].

[a] Proteases are enzymes that catabolize the hydrolysis of the peptide bonds between amino acids in the peptide backbone of a protein.
[b] Glycosidases are enzymes that assist the hydrolyzation of the glycosidic bond between the glycan and the amino acid of the peptide sequence.

All three strategies are complementary as they have their own advantages and drawbacks, reason why, to obtain all the possible information and complete the glycoprotein structural puzzle, quite often all three approaches must be taken into account.

In regards to the instrumentation, a mass spectrometer is a complex and sophisticated analytical instrument that consists of several parts, being the most important ones the ionization source, analyzer and detector [117]. Briefly, the ionization source transforms the analyte in a gaseous ion [118] and the analyzer separates the generated ions on the basis of their mass to charge ratio; however, the mechanism behind the separation highly depends on the type of analyzer, which will also affect sensitivity or resolution of the mass spectrometer [119]. Finally, the detector, as the name implies, will give a signal whenever an ion passes by or hits its surface, thus, acknowledging the presence of that ion [120].

1.3.2.1. Electrospray ionization

Electrospray ionization (ESI) and **matrix-assisted laser desorption/ionization (MALDI)** are the most common ionization sources used in glycoproteomics. Both are considered soft ionization techniques as relatively little energy is imparted to the analyte, thus, less fragmentation occurs during the ionization process. These two revolutionary techniques greatly expanded the use of mass spectrometry to almost all fields in science, due to the fact that not only small neutral molecules could be transformed in gas-phase ions (which was seemingly possible before with other ionization techniques) but also large biomolecules and polar compounds.

Since the revolutionary introduction of ESI in the late sixties and its application to the ionization of large biomolecules in the late eighties, the use of MS hyphenated techniques, like CE-MS, or, to a greater extent, HPLC-MS, has given rise to numerous publications and advances in all *omic* sciences. This has made ESI the ionization technique par excellence in the analysis of proteins, glycoproteins and other biological relevant macromolecules [121–123].

The formation of gas-phase ions from ions or neutral molecules in solution by ESI consists of three different steps, englobed in a schematic representation of the whole process in Figure 1.14: 1) Spray formation is accomplished by the application of a potential difference between the inlet of the mass spectrometer and the end of a conductive capillary, which contains the analyte solution [124, 125]. As a result, a spray of charged droplets (Taylor Cone) is formed at the end of the capillary, often assisted with an auxiliary gas and heating to obtain better and more stable sprays. 2) The solvent in the charged droplets is evaporated, usually using an inert gas, until the charged droplets reach the so-called Rayleigh limit, that is, the state at which the surface tension is equal to the Coulombic repulsion between the charges on their surface. 3) In the Rayleigh limit, new, smaller and more stable droplets are formed from the initial charged droplet (fission process). Afterwards, several successive rounds of desolvation and droplet fission occur until the gas ion is formed.

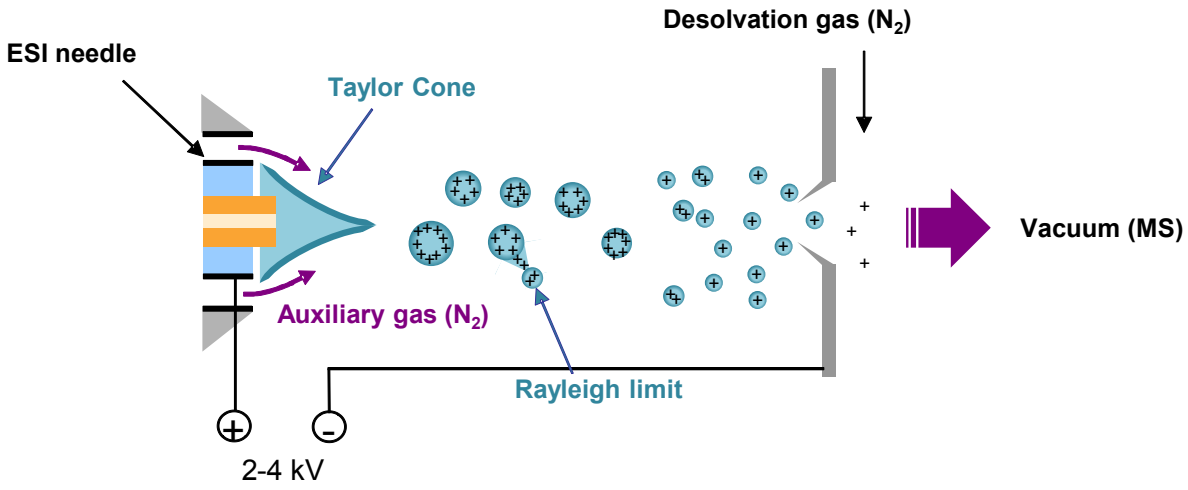


Figure 1.14: Schematic representation of the basics of spray formation in electrospray ionization (ESI).

The principal advantage of ESI is that **multi-charged** ions are formed, reason why it is continuously praised as a ionization technique that allows 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). However, when larger molecules with hundreds of thousands of daltons (Da) are analyzed, obtained spectra can be really convoluted and mathematical approaches might be necessary to fully understand the obtained information.

Although ESI ionization is considered a soft ionization technique, special care has to be taken regarding the potential applied to form the spray. A middle ground has to be found between applying too much voltage, which may cause in-source fragmentation or loss of the native form, or too low, which will considerably aggravate sensitivity. However, several other parameters may affect spray efficiency and, consequently, alter the sensitivity of the method, being the composition and pH of the solution and the number of ionizable groups of the molecule the most noted ones.

In the last few years, the scientific community has seen the rise of a new variant of ESI, which was already introduced by Wilm and Mann [126, 127] in the late nineties, the noteworthy **nanoElectrospray (nanoESI)**. This interface is ideally suited for the ionization of analytes in low flow rate systems, such as nano liquid chromatography or capillary electrophoresis. In this variant, the initial diameter of the formed droplets is reduced considerably, which provides several advantages compared to conventional ESI:

- Less droplet fissions are needed to ionize the analytes and solvent evaporation is much faster.
- Sensitivity is considerably increased.
- Less adduct formation.

Apart from all the aforementioned advantages, one particular and interesting application of nanoESI is the study of molecules in **native-like** conditions, especially proteins and glycoproteins. As ionization conditions can be even milder than normal ESI, when neutral buffers, with relative low ionic strength, are used, proteins and glycoproteins can be sprayed without altering their non-covalent interactions. This allows the study of polymeric structures

and protein complexes. Nowadays, this new approach to study protein structure and functionality is referred to as **Native Mass Spectrometry** and is fully recognized as another powerful tool in structural biology [128, 129].

1.3.2.2. Mass analyzers

Nowadays, there are several types of mass analyzers, ranging from a simple quadrupole to more complex hybrid systems that combine two or more analyzers into one. Each analyzer provides certain advantages over other ones and, quite often, different information can be obtained depending on the type of analyzer. Most high-end analyzers combine the advantages of more than one to form complex hybrid instruments with increased sensitivity and resolution. The proper analyzer will be the one with the required sensitivity, resolving power, accuracy, acquisition speed and dynamic range, among other characteristics.

Some of the most common used analyzers in the study of glycoproteins include time-of-flight (TOF), quadrupole time-of-flight (qTOF), Orbitrap™ and Fourier transform ion cyclotron resonance (FT-ICR). 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.

Separation in a TOF instrument is based on the different velocity of the introduced ions when are accelerated through a high voltage at the beginning of a flight tube. The time taken to traverse this tube to reach the detector is dependent on the m/z rates of ions. Ions with greater charge but lower mass cross the tube before those ions with higher mass and lower charge. As the initial accelerating voltage is pulsed, the output of the detector as a function of time can be conveniently converted into a mass spectrum [130, 131].

The most important advancement in TOF technology was the development of the orthogonal acceleration (oa-TOF) [131, 132]. The most striking difference introduced with oa-TOF was the

use of a separated direction for the separation of ions, orthogonal to the continuous ion-beam of the ion source. This distribution 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) [133, 134].

In recent years, hybrid type analyzers have aroused great interest in the analytical community, not only for their great performance and amplitude of benefits but, particularly, for the fact that more structural information can be obtained as MS/MS experiments are easily carried out. In this regard, quadruple time-of-flight (qTOF) is often noted as a versatile, powerful and robust instrument and considered one of the typical platforms in the glycoproteomic field [103, 135, 136]. This hybrid spectrometer can be regarded as a conventional TOF but with an additional mass-resolving quadrupole before the mass analyzer. The principal benefits of this distribution are the high sensitivity, mass resolution and mass accuracy obtained for the product ions with the relative affordable price, ease of use and simplicity. However, new advances in instrumentation have led to the introduction of novel and more powerful hybrid mass spectrometers, such as the Thermo Scientific™ Orbitrap Fusion™ Lumos™ Tribrid™, being their impact in the glycoproteomic field already reported by some authors [137–139].

1.3.3. Ion mobility mass spectrometry

As has been introduced before, glycans can form complex structures, severely branched, with many monosaccharide constituents, usually resulting in several isomers [140]. Even though, the described MS technology is perfectly suited and capable of obtaining information to characterize carbohydrate structures, when analyzing isomeric glycans, conventional MS approaches fail to separate them [141–143], as they have identical mass and atomic composition. Hence, glycans isomers are undistinguishable unless derivatization or less common stationary phases in liquid chromatography (LC) are used [144–146]. In the last few

years, several MS/MS methods have been reported that allow the identification and structure characterization of glycan isomers [147–151]. Nevertheless, these MS/MS methods are still in the early stages of development, therefore, their robustness is somewhat limited and obtaining specific fragment ions to reliably identify each isomer can be a difficult task.

Ion mobility mass spectrometry (IM-MS) provides an easy and straightforward solution to this problem. IM-MS has aroused great interest in the last years, not only in the glycomics and glycoproteomics fields, but also in other *omic* sciences, as a proficient analytical technique for the separation of isomeric biomolecules [140–143, 152–155]. 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, hence, resolving ions that would be otherwise indistinguishable solely by MS, such as isomers [156–159]. Figure 1.15 illustrates the fundamentals of classical IM-MS separation.

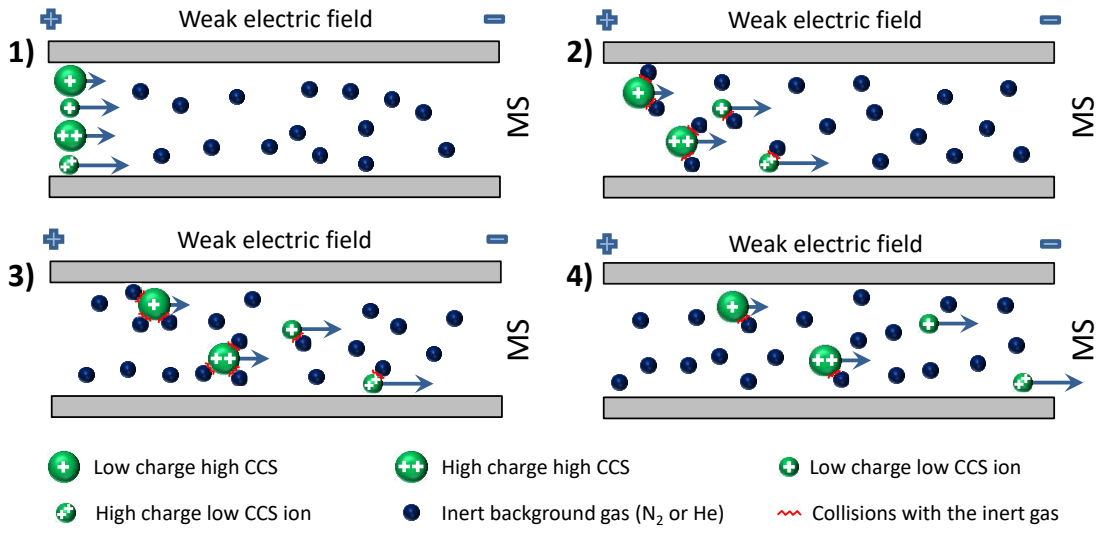


Figure 1.15: Separation in classical IM-MS. Ions with smaller CCS and higher charge are separated before those with higher CCS and smaller charge.

Specifically, IM-MS measures the time (**drift time**) that a particular ion takes to cross a cell filled with a background inert neutral gas (being N_2 and He the most commonly used) at a controlled pressure under the influence of a weak electric field. The drift time of a specific ion is mainly due to the ion-gas collisions; therefore, ions are separated due to their ion-neutral **collision cross-section** (CCS, Ω), related to the overall shape and topology of the ion [156–158]. Small compact ions cross first as a result of their smaller CCS. Moreover, the higher the charge of the ion, the greater the strength of the separation field, and therefore the more quickly the ion will cross the chamber. Hence, those ions with high charge and small CCS will cross the mobility cell before than those with higher CCS and lower charge. Consequently, IM-MS is often considered as being proportional to collision cross-section-to-charge ratio (Ω/z) [158]. IM-MS is regarded as a powerful analytical technique that provides three-dimensional analytical information, shape-to-charge, mass-to-charge and abundance, thus, allowing reliable analyte identification.

Nowadays, there are several IM methods next to the classical drift-time ion mobility spectrometry (DTIMS), such as field asymmetric waveform ion mobility spectrometry (FAIMS), but among them, **traveling wave ion mobility spectrometry (TWIMS)** is the one that has seen a major growth in the last years [160, 161] and the one used in this thesis. Figure 1.16 shows the fundamentals of TWIMS separation.

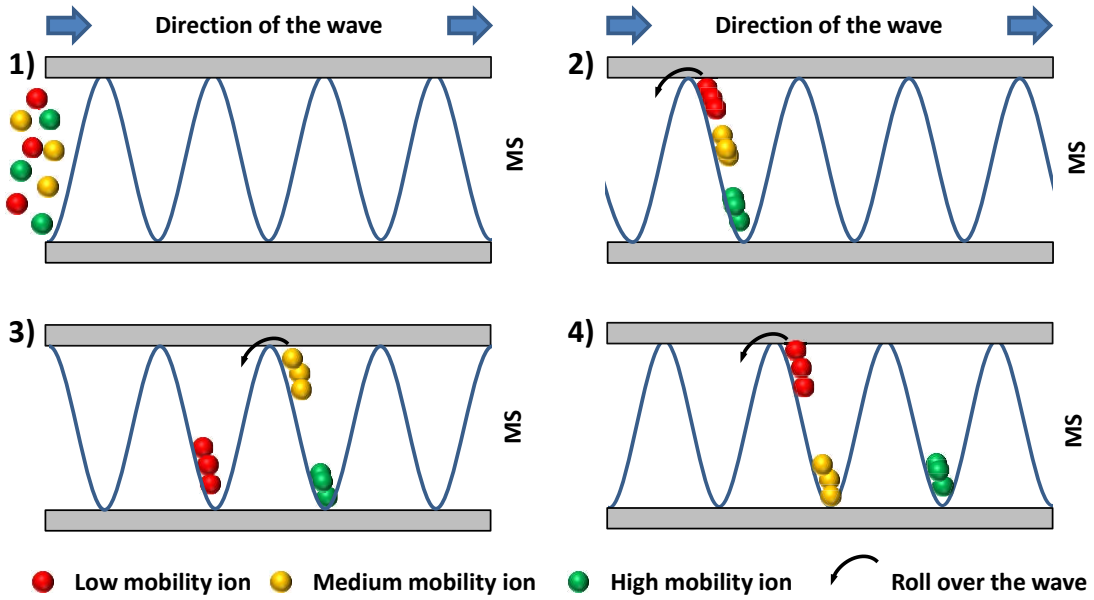


Figure 1.16: Fundamentals of TWIMS separation. High mobility ions travel with the wave, thus, arriving first to the MS. Whereas, low mobility ions roll over the wave, spending more time in the mobility cell and, therefore, arriving later to the MS.

Briefly, in TWIMS, ions are propelled thanks to a sequence of symmetric potential waves continually propagating through a cell. An ion will travel along the wave depending on different parameters, such as ion charge, CCS, the interaction of the ion with the background gas and the characteristics of the wave (i.e., amplitude, velocity). Therefore, those ions with higher mobility will be pushed along with the wave, thus, they will travel the drift cell faster. Whereas, low mobility ions roll over the top of the wave and, as a result, they stay longer in the mobility cell. Particularly, in the case of Figure 1.16, low mobility ions (red) keep rolling over the wave, thus they are the last to arrive to the MS. Medium mobility ions (yellow) roll over fewer times and, then, arrive sooner. Whereas, high mobility ions (green) travel with the wave, arriving first to the MS.

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 [162, 163].

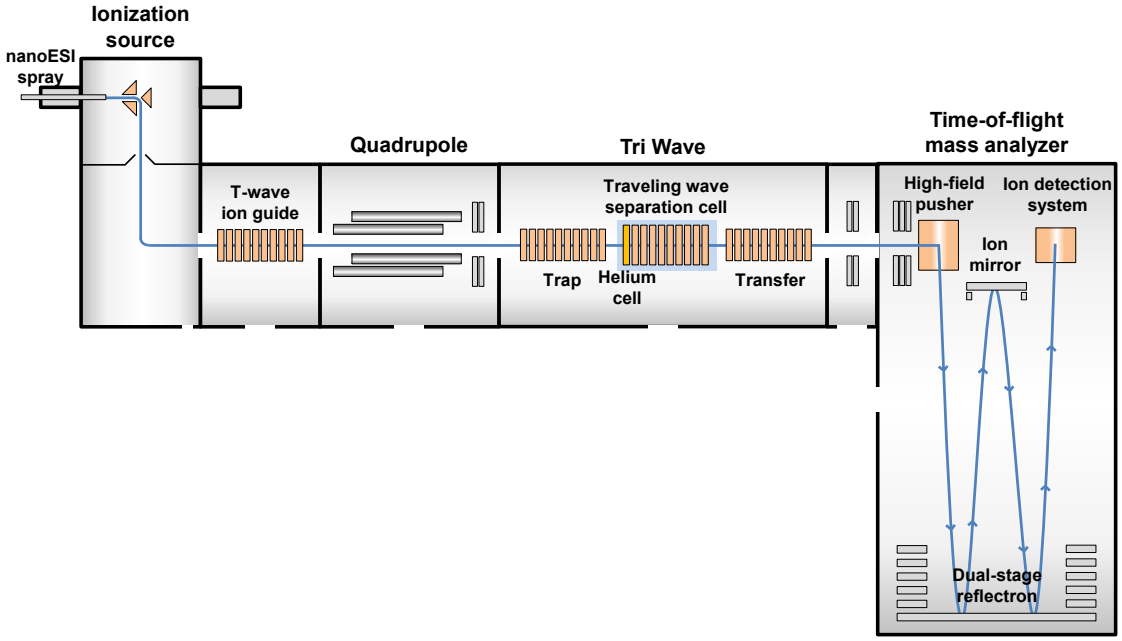


Figure 1.17: Schematic representation of an ion mobility mass spectrometer (IM-MS) with a traveling wave ion mobility cell (Synapt G2 from WATERS® Corp.).

1.3.4. Separation techniques

Even though complex mixtures of different glycans can be analyzed in a single MS experiment, very often it is necessary to couple MS with a separation technique, such as HPLC or CE, to separate the different analytes prior MS analysis. In this way, the potential of MS detection can be considerable increased due to the fact that ion suppression is minimized, enabling the detection of those analytes whose concentration or ionization efficiency is lower.

1.3.4.1. Liquid chromatography

High performance liquid chromatography (HPLC), commonly referred to as simply liquid chromatography (LC), is considered the analytical technique par excellence which, in the last decades, has experienced a worldwide expansion with thousands of applications [164–166].

HPLC allows the separation of the compounds of a complex mixture carried by a liquid, referred to as **mobile phase**, due to the different distribution of the compounds between the

mobile phase and the solid phase within a column, known as **stationary phase**. In HPLC the mobile phase is pumped with high pressure to improve the efficiency of the separation. The time at which a specific compound elutes from the column, is referred to as retention time, and is governed by several factors, such as the physicochemical properties of the compound and the stationary and mobile phases.

In regards to glycoproteomics, HPLC is widely used as an analytical technique for the characterization of protein glycosylation, especially at the glycopeptide and glycan level, mainly due to its good quantification, reproducibility and robustness [167–169]. Nowadays, due to the advances made in regard to the particle size and chemistry of stationary phases, several subtypes or separation modes of HPLC have emerged [170], all of them used in the glycoproteomic field for different purposes, in combination with different types of stationary phases. With so many possibilities, several factors, such as sample availability, required sensitivity or desired separation, among many others, must be taken into account to correctly select the appropriate combination that suits your needs.

The main HPLC mode used during the thesis has been **Capillary Liquid Chromatography (CapLC)**. In CapLC, column dimensions are greatly reduced, particularly, the internal diameter (i.d.). The principal benefit from reducing the i.d. is the ability to work with minute sample volumes, with the consequent improvement in peak shape and reduction of the limits of detection (LODs). The increase in sensitivity is due to the reduced dilution of the chromatographic band during analysis when columns with smaller i.d. are used. Additionally, as low flow rates are used, less mobile phase is consumed, which, in turn, reduces the costs of purchase and disposal of these solvents and diminishes their environmental impact. Moreover, in CapLC the flow rate is high enough to be able to use conventional ESI, with the only addition of a narrow capillary inside the ESI interface to avoid band spread and loss of separation efficiency.

Nano Liquid Chromatography (nanoLC) can be considered one of the ultimate advancements in LC. In this case, the flow rate is extremely reduced as well as the column i.d. [171]. Consequently, the improvement in sensitivity compared to other LC variants is significant, usually at least a 100-fold and a 10-fold increase is expected when compared to HPLC or CapLC, respectively. However, a small inconvenience when using nanoLC is that the flow rates are so low that conventional ESI sources for MS cannot be used. Instead, the more expensive nanoESI interface is often required when coupling nanoLC with MS. Recently, a variation of nanoLC has become popular which uses the small flow rates and low internal diameter of the column with smaller particle size, increasing the system pressure considerably. This system, which is a combination of nanoLC and ultra-performance liquid chromatography (UPLC) and, therefore, referred to as nano-ultra performance liquid chromatography (nanoUPLC), allows separation in less time, improving peak shape and increasing sensitivity when compared to normal nanoLC. Additionally, some authors have been able to separate glycopeptide glycoforms when using common reversed phase stationary phases, such as C18, in nanoUPLC, but also in nanoLC [172–174]. However, the mechanism behind the separation in these cases is still not quite understood and a matter of debate.

Most common stationary phases formerly used in several glycoproteomic studies were C18 and C8 [116, 173, 175], which had several advantages, such as, the commercial availability of columns with these stationary phases in different dimensions, their relative low cost compared to other type of stationary phases and the fact that typical mobile phases are used, which favor the ionization by mass spectrometry. However, because of their inability to properly retain glycans and separate the glycoforms of glycopeptides in most cases, they have been recently discarded in favor of novel stationary phases, such as, hydrophilic interaction liquid chromatography (HILIC) stationary phases. HILIC provides better separation between glycans, sometimes even separating isomeric ones, and also between glycopeptides which differ in the content of sialic acid [175–178]. However, mobile phases used in HILIC are quite saline, thus,

ionization efficiency in mass spectrometry is severely hampered. In the case of glycans, derivatization is an alluring, albeit time-consuming, option to avoid ionization problems and increase sensitivity.

As glycoproteins are negatively charged at physiological pH and in most common mobile phases, due to the presence of sialic acids, anion exchange liquid chromatography has been generally used for the analysis of intact glycoproteins, allowing the separation of intact glycoforms that differ in the content of sialic acid (as they have different total charge). In fact, anion exchange liquid chromatography is used as a routine method for the analysis of hTf intact glycoforms in CDG and chronic alcoholism [43, 66, 179]. However, the principal drawback of this variant is the high content in salt of the mobile phase, which makes the coupling with a mass spectrometer a difficult task [168].

1.3.4.2. Electrophoresis

Electrophoresis encompasses a group of analytical techniques that are able to separate a mixture of biomolecules according to their molecular size and charge. Separation by electrophoresis is based on the fact that charged molecules will migrate at different speed through a matrix upon application of an electric field depending on their charge to mass ratio. Therefore, highly charged small molecules will migrate faster in comparison to bigger less charged ones.

1.3.4.2.1. Gel electrophoresis

In gel electrophoresis, as the name indicates, the aforementioned matrix is often an agarose or polyacrylamide gel, as they are anti-conductive media that minimize the loss of efficiency due to thermal diffusion. This technique is a relatively straightforward, simple and highly sensitive tool to study the properties of certain molecules, for instance, proteins.

In glycoproteomics, as in proteomics, two main variants of gel electrophoresis are employed for the analysis of glycoconjugates: the one dimensional sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and the two-dimensional polyacrylamide gel electrophoresis (2D-PAGE).

SDS-PAGE was first described by Laemmli [180] and can be considered one of the most commonly used techniques among all other variants. In SDS-PAGE, a polyacrylamide gel is used as the matrix due to its ability to separate molecules ranging from low to relatively high molecular weight, such as proteins [181]. It is worth mentioning that the range of molecular weights that can be effectively separated is governed by the porous size of the polyacrylamide gel [181], narrow porous size for small molecules and wide for larger ones. Nowadays, even gels with a gradient in porous size are used to separate proteins with a myriad of molecular weights. Additionally, sodium dodecyl sulfate is used as a surfactant to unfold proteins into a linear shape and impart a distribution of negative charges per unit mass. In this way, protein separation is only due to their approximate size. The typical instrumentation used for SDS-PAGE as well as an example of a common obtained gel is shown in Figure 1.18-a

Perhaps even more common than SDS-PAGE in proteomics and glycoproteomics, 2D-PAGE is a powerful technique that couples isoelectric focusing (IEF) in the first dimension with SDS-PAGE in the second dimension, allowing the separation and fractioning of highly complex mixtures of proteins and glycoproteins on the basis of their isoelectric point (pI) and their molecular weight, respectively. As with monodimensional SDS-PAGE, the gel can be modified to effectively separate molecules of diverse molecular weight and the pH gradient of the IEF strip can also be changed to accommodate specific needs. Then, depending on these two conditions, 2D-PAGE is able to resolve as many as 5000 proteins simultaneously and even detect 1 ng of protein per spot [182, 183]. 2D-PAGE delivers a map of intact proteins of the loaded sample as can be observed in Figure 1.18-b. Changes in protein isoforms, expression

level and posttranslational modification (PTM) can be easily detected, therefore, 2D-PAGE is often regarded as an exceptional tool to study glycoproteins [184]. Another great advantage of 2D-PAGE is the ability to isolate a specific glycoform for further characterization by MS. In this case, digestion of small gel portions, where the glycoprotein of interest has been isolated, is also possible and a common practice nowadays for structural analysis, carbohydrate characterization or amino acid sequencing by MS [185, 186].

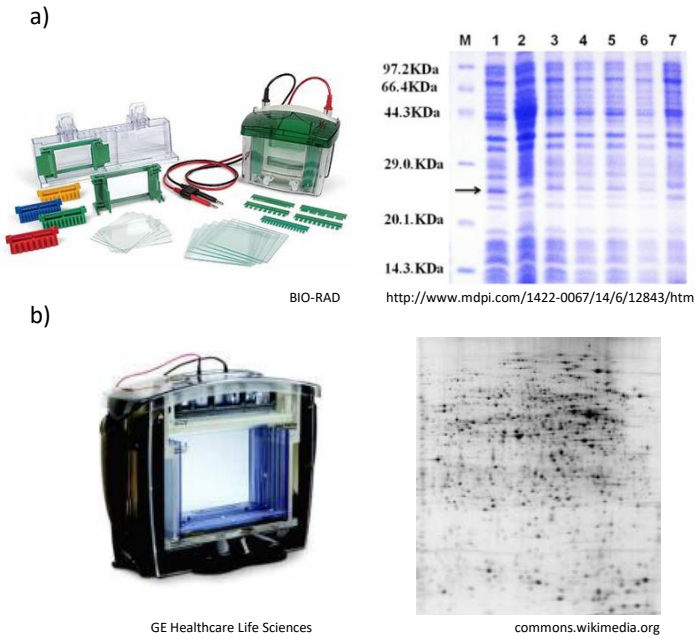


Figure 1.18: Typical instrumentation and obtained gel with a) polyacrylamide gel electrophoresis (SDS-PAGE) and b) two-dimensional polyacrylamide gel electrophoresis (2D-PAGE).

1.3.4.2.2. Capillary electrophoresis

Capillary electrophoresis (CE) is encompassed in the so-called group of microscale separation techniques, which also includes capillary liquid chromatography and chip-based techniques. In CE, as opposed to gel electrophoresis, separation takes place in a narrow capillary, where the loss of efficiency due to thermal diffusion and convection is almost non-existent, generating only minimal amounts of heat. Among all the advantages of CE, the minimal sample and reagent consumption, its high selectivity, resolution and efficiency, the short analysis times,

the ease of use and the potential for automation, are the most prominent ones. However, what sets CE apart from gel electrophoresis based techniques is the fact that it can be coupled to mass spectrometry. All the aforementioned advantages make CE a more robust and competitive technique than those based on gel electrophoresis, and, also, a notably complementary technique to HPLC. With CE, multiple separation modes can be used, reason why it can be considered a versatile and suitable technique for numerous applications, from the analysis of biomolecules to inorganic ions [187–189].

The most common modes of operation of CE are capillary zone electrophoresis (CZE), capillary gel electrophoresis (CGE), micellar electrokinetic chromatography (MEKC), capillary electrochromatography (CEC), capillary isoelectric focusing (CIEF) and capillary isotachophoresis (CITP). However, CZE is the most common mode due to its simplicity and versatility, which is demonstrated by the wide range of analytes that can be separated, including glycoproteins [188]. This is the reason why the use of only CE to refer to CZE has been extensively accepted by the scientific community. Apart from the already mentioned advantages, the only mode used during this thesis has been CZE (which will be referred to from now on as CE) due to the fact that the coupling of the other modes to MS is generally more complicated. To further delve into the basics of this technique, the principles of the separation in CE will be explained in the next section.

- Separation principles in CE

CE is a microscale analytical technique where charged species are separated on the basis of their migration velocity when an electric field is applied in a silica capillary of small internal diameter (approximately 25-75 μm), filled with a conductor background electrolyte. Therefore the migration velocity of a charged specie is dependent on the electric field applied and its electrophoretic mobility (m_e or μ_e), which is a constant of that specie and depends on the electric force that it experiences, balanced by its frictional drag through the medium. When a

steady state is attained between these two opposing forces, the m_e of a given ion can be defined as:

$$m_e = \frac{q}{6\pi\eta r} \qquad \textbf{Eq. 1}$$

where q is the ion total charge, η is the viscosity of the medium (background electrolyte) and r is the solvated ion radius. In turn, the radius of a charged specie depends on its molecular weight, three-dimensional structure and solvation degree; however, quite often only the molecular weight is considered in the above equation for practical purposes. Consequently, small, highly charged species have higher mobilities than larger, slightly charged ones.

Another basic fundament of CE separation is the so-called **electroosmotic flow (EOF)**. When silica capillaries are used, the inner wall of the capillary is composed of several silanol groups (SiOH) that exist in anionic form at pH above 2 (see Figure 1.19). Therefore, under aqueous conditions the inner wall possess an excess of negative charge that favors the formation of a diffuse double-layer near the surface where counterions, cations in this case, build up to maintain the charge balance. As soon as the voltage is applied across the capillary, the cations in the diffuse double-layer are attracted towards the cathode which is usually at the end of the capillary (normal polarity). As cations are solvated, their movement drags the bulk solution inside the capillary towards the cathode, generating the aforementioned EOF (Figure 1.19). Hence, the EOF causes nearly all species to be dragged in the same direction, that is, towards the cathode where the detector is located (normal polarity), as the flow goes from the positive electrode (anode) to the negative electrode (cathode), as can be observed in Figure 1.19. Consequently, the experimental mobility of a given ion, also referred to as effective mobility, is dependent on its electrophoretic mobility and the EOF, which, in turn, is mainly affected by the pH and composition of the background electrolyte. As, in general, the EOF is at least one order of magnitude greater than the electrophoretic mobility of the solutes, cations migrate faster as

they are not only propelled by their electrophoretic mobility but also by the EOF. In contrast, neutral compounds are only dragged by the EOF with its same velocity and anions are the latest to migrate since they are attracted towards the anode because of their electrophoretic mobility, but are, nonetheless, carried towards the cathode by the EOF due to its greater strength (Figure 1.19). Additionally, another advantage of the generation of the EOF is that the flow velocity profile is almost uniform throughout the entire capillary, avoiding the dispersion or broadening of the solute zones.

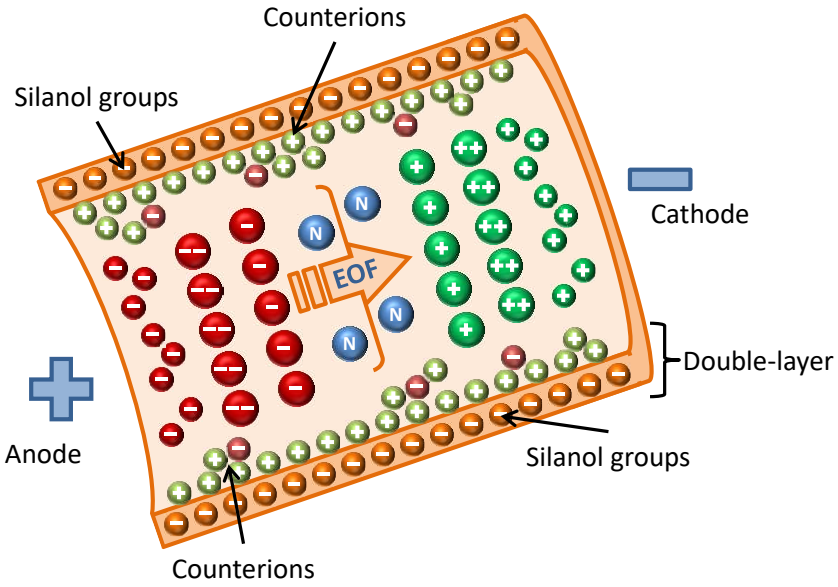


Figure 1.19: Differential ion migration depending on charge and size due to the electroosmotic flow (EOF) in capillary electrophoresis (CE).

One of the reason that makes CE a noteworthy technique in glycoproteomics is the fact that it allows the correct separation of intact glycoforms and glycans or glycopeptides of a glycoprotein digest, within a single run, being for example, the analytical technique of choice in the Pharmacopeia Reference Method for the analysis of recombinant human erythropoietin [190].

- CE-ESI-MS

The coupling between CE and MS using ESI as ionization source (CE-ESI-MS) has not always been an easy task, but since its introduction almost four decades ago, CE-ESI-MS has aroused great interest as a powerful and complementary analytical technique to HPLC-ESI-MS. CE-ESI-MS offers several benefits, as the high resolution and separation efficiency of CE is combined with the use of ESI-MS, which generates multicharged species and allows obtaining structural information and reliable identification through mass determination of the analyzed species. Due to the aforementioned advantages, CE-ESI-MS is well suited for the analysis of glycoconjugates, which is demonstrated by the continuous influx of reviews of numerous CE-ESI-MS applications in several research areas of the glycoproteomic field, such as medicine, pharmacology, biochemistry, among many others [191–194].

Even though ESI was revolutionary to establish a reliable and robust connection between LC and MS, in regards to CE, this connection was problematic. This was mostly due to compatibility problems between the conventional background electrolytes used in CE and the mass spectrometers, which dictate that all the components of the solvent must be volatile. This drawback together with the low flow rates in CE and the fact that a voltage is applied to accomplish the separation, thus, the electric contact needed to be terminated, were the main causes why the first commonly developed interfaces were inappropriate for CE-MS coupling. Not until Smith et al. [195] developed the so-called **sheath-flow interface** was CE successfully hyphenated to MS. In this interface, a coaxial sheath liquid is used to increase the flow rate and also end the electric contact in the terminus end of the CE capillary [187, 196] (see Figure 1.20-a). Usually, an inert gas is employed to assist the ionization and improve spray stability. Nowadays, this interface is commercially available, being Agilent Technologies® Inc. the predominant company in this regard.

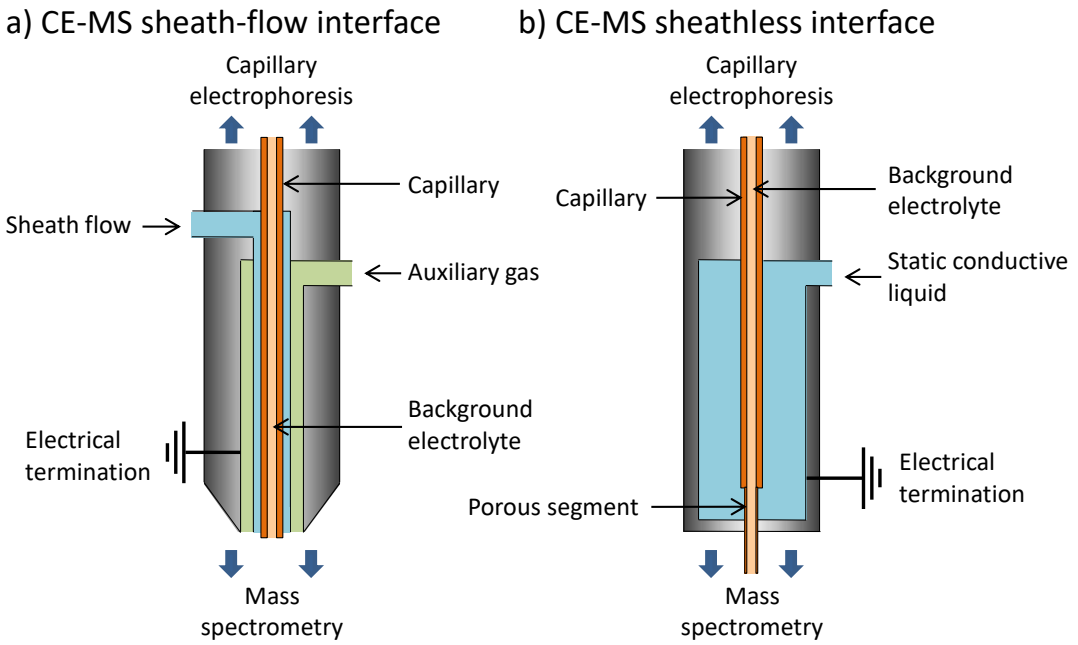


Figure 1.20: Schematic representation of the a) sheath-flow and b) sheathless interface for coupling capillary electrophoresis to mass spectrometry.

This interface provides several advantages, such as increased reproducibility, robustness, ease of use and versatility, among others, and, together with the use of volatile BGEs, has resolved the main problems faced when coupling CE with MS. However, there are still some notable drawbacks when using this interface. First of all, as the flow rate is considerably increased, analytes are consequently diluted, which, in fact, reduces the overall sensitivity of the method. Secondly, the use of a sheath flow at the end of the capillary with a different composition than the background electrolyte might contribute to band spread and affect the separation of the analytes or their migration times [197–199].

The introduction of nanoESI favored the development of new CE-MS interfaces, being the **sheathless interface** the most noteworthy, due to the fact that, several of the sheath-flow interface limitations were solved with this variant as no coaxial liquid was needed (Figure 1.20-b). Even though this interface is relatively new and needs to be improved, in particular its

robustness, several prototypes are being developed, being the sheathless interface from Beckman Coulter® Inc. the most noted one [197–200].

- Modelization of the electrophoretic migration behavior

One additional and appealing advantage of choosing CE as separation technique is that the electrophoretic migration behavior of the compounds of interest can be easily predicted using the classical semiempirical relationships between electrophoretic mobility (m_e) and charge-to-mass ratio (q/M) (equation 2), when using CE in combination with MS [196, 201, 202]:

$$m_e vs\ {}^{q}/_{M^{\alpha}} \qquad \textbf{Eq. 2}$$

the value of α is dependent on the electrophoretic model. This prediction can be quite an alluring modelization tool to speed-up method development or to refine the structural assignments based on the measured molecular mass [203–205]. In glycoproteomics, modelization of the electrophoretic migration behavior could be quite useful to easily generate 'dry-lab' peptide-glycopeptide electrophoretic maps of glycoproteins. In turn, this could facilitate the optimization of the separation and identification of the glycopeptides of interest.

However, this glycopeptide modelization can be a challenging task, mainly, due to the different glycoforms that glycopeptides can show. Different glycoforms not only affect the glycopeptide size (and, hence, molecular mass), but also its charge, as the presence of sialic acids strongly contributes to the global glycopeptide charge and is critical for the electrophoretic separation resolution.

Particularly, in this thesis, three classical semiempirical relationships have been tested to model the migration behavior in CE-MS of peptides and glycopeptides of rhEPO, the Stoke's law (α=1/3, equation 2), the classical polymer model (α=1/2, equation 2) and the Offord's surface law (α=2/3, equation 2). These three models differ in how the compound shape is assumed to be and how the compound might be affected by the forces that it undergoes

during electrophoretic motion [206–208]. Specifically, in the Stoke's law compounds are modelled as spherical particles, whereas in the classical polymer model they are modelled as polymers with lower charge densities, and in the Offord's surface law they are modelled as larger molecules with more rigid structures, which experience frictional forces that are proportional to their surface area.

1.4. Data analysis

In several *omic* sciences, the amount of data generated is quite often too high and convoluted to be able to obtain proper conclusions with a simple exploration without mathematical tools. In glycomics and glycoproteomics studies, datasets can be even more massive and complex due to the inherent complexity and microheterogeneity associated with carbohydrates. In those cases, further data exploration and interpretation might be needed to obtain accurate information and representation of the dataset. Statistical and chemometric data analysis methods play a crucial role in data processing, exploration and classification of the large datasets generated in such studies [209].

Conventional statistic methods, which include the common significance tests, only provide limited information about the dataset, as mean comparison does not take into account multicompound relationships and interactions. To further investigate complex datasets, quite often multivariate data analysis methods are required. In this thesis, **principal component analysis (PCA)** and **partial-least square discriminant analysis (PLS-DA)** have been used and, thus, they will be briefly explained in the next section.

1.4.1. Principal Components Analysis

PCA is regarded as one of the most old and widely used mathematical tools for data decomposition and reduction (i.e. dimensionality reduction), preserving the majority of information, which helps to understand and interpret large and complex datasets [210–212].

In PCA, the matrix that forms the dataset is decomposed into several principal components (**PCs**), which could be defined as new variables that are linear functions of those in the original dataset. These PCs maximize explained variance in the data on each successive component under the constraint of being orthogonal to each other (i.e., they must be uncorrelated to one another) [213–215]. As a result, a bilinear model is generated which is a product of scores (**T**) and loadings (**P**) matrices:

$$X = TP^T + E \qquad \textbf{Eq. 3}$$

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 regarded as a potent visualization technique as each object (sample) gets a scores value on each PC and, in the same manner, 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 in the data, such as clusters, trends and outliers. Whereas, 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. 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 [213–215].

1.4.2. Partial Least Squares Discriminant Analysis

Introduced approximately three decades ago [216, 217], partial least squares (PLS) resembles PCA in the fact that it reduces the dimensionality of a dataset matrix by means of decomposition into a set of components. In this case, components are referred to as latent

variables (**LV**), but they are created following another criterion than the maximum variance, as opposed to PCA [213].

In PLS, a new matrix comes into play composed of normalized weight vectors ($\mathbf{W}^T$), calculated as the covariance between the response **Y** (groups, class membership) and the data matrix **X** (sample readings). 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. When PLS is used as a supervised classification method, the response variable is just a binary vector of zeros and ones (Figure 1.21), which describes the class membership for each sample in the investigated groups. In this case, the method is referred to as **partial least squares discriminant analysis** (PLS-DA) [218–220].

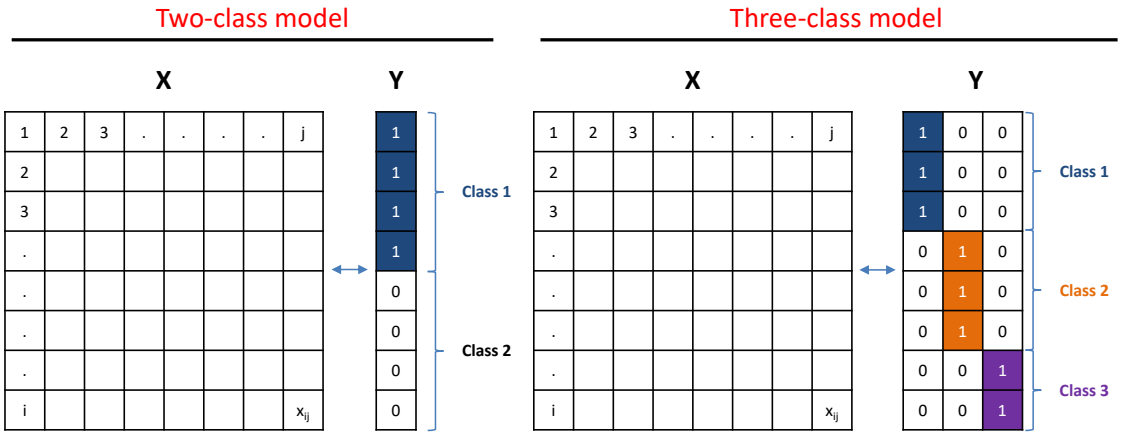


Figure 1.21: Representation of partial least squares-discriminant analysis (PLS-DA) for models including two classes and three classes.

This method provides important statistics such as loadings weight, regression coefficient and **variable importance on projection** (VIP, a measure of how important a specific variable is for the discrimination of a specific group from the others). All of them help in the identification of the most important variables. Additionally, the comparison of the loadings and scores plots provides information about the relationship between important variables and the groups which they are related and permits to detect outliers and samples that are not correctly

classified. All in all, this mathematical tool provides a visual interpretation of really complex datasets and helps to discern which variables are the most important in the classification of samples in groups [218–220].

1.5. Trends and future applications

Glycomics and glycoproteomics are ever expanding fields that are surely to continue to grow with new advances in the following years. In this regard, MS has become the prime option for any glycosylation study whose purpose is to obtain reliable structural information about protein glycoforms and glycan composition.

One of the fundamental aspects of MS-based glycosylation analysis is the need to selectively enrich and isolate the glycoprotein of interest. There are several purification and enrichment methods available nowadays [104, 221, 222], some of the most important ones have been already discussed in section 1.3.1. In the last few years, solid phase extraction (SPE) with HILIC and porous graphitic carbon (PGC) stationary phases have gained popularity for the purification of glycoconjugates due to their excellent performance in the separation between glycans or glycopeptides [104, 223]. Other SPE strategies involving the use of boronic acid [224] or hydrazide [225] have also been used in glycoproteomics due to their high affinity to glycoconjugates. In addition, several authors have proposed, quite recently, different strategies that involve the use of nanomaterials for the purification of glycoconjugates [226–229]. These new strategies offer several advantages, such as better enrichment capacity, quicker adsorption process and higher recovery efficiency, mainly due to the larger specific surface area of the nanomaterials. In the following years, most technological advances will focus on improving the enrichment and purification rates of glycoconjugates, thus, increasing the sensitivity of glycomic-based methods. Furthermore, the automation of the whole enrichment process will surely be another important research area.

Glycoprotein digestion is another vital aspect of the sample treatment in studies involving glycopeptide and glycan analysis. Usually, the digestion is carried out in-solution or in-gel, however, the use of reactors where the enzyme is immobilized in a solid support has recently gained popularity due to reduced digestion times and their suitability for automation [230, 231]. Additionally, it has become a common practice to use a combination of different enzymes, mainly to increase digestion efficiency and decrease the complexity of the glycosylation study, obtaining smaller glycopeptides with a single glycosite or, even, breaking down the protein into its individual amino acid components [232]. The automation and enhancement of the digestion procedure will probably be important areas of investigation in glycopeptide-based studies. In the case of glycans, several strategies are being developed continuously in regards to their derivatization and labelling aimed at improving their quantification and detection efficiency. Permethylation and reductive amination have been the most used derivatization strategies so far, but other strategies have been reported recently, each one focusing on improving one specific aspect of the glycan characterization [104, 114, 233]. The use of novel derivatization or labelling strategies to improve fragmentation efficiency, to enhance ionization of glycoconjugates, to avoid sialic acid fragmentation or to improve glycan quantification are just some examples of the main purpose of the myriad of different procedures available in the literature nowadays [104, 114, 167, 233]. Developing new strategies for the derivatization or labelling of glycans in order to improve sensitivity and accuracy of glycan profiling will probably continue to be an important area of research in the following years.

In regards to instrumentation, MS is present in the majority of glycoproteomic-related publications. However, quite often a separation technique before MS detection must be used in order to improve glycan characterization and enhance ionization efficiency. As stated in the introduction, HPLC and CE techniques are the most frequent separation techniques used in glycoproteomics. However, recent advances in HPLC focus on the use of the most novel

stationary phases, such as HILIC or PGC, to improve glycoconjugate separation and detection. The use of these stationary phases orthogonally to other conventional ones, like reverse phases (C_{18}), or the tendency to use chip-based technology in the separation of glycoconjugates are two of the most recent advancements in this regard [234–237]. Additionally, monolithic columns are novel tools that have also been used in glycoproteomics due to their low back pressure, efficient mass transfer and large surface area [238, 239]. As with HPLC, there have also been new advancements in the miniaturization of CE, with the development of the so-called microchip electrophoresis (ME), which has shown to be a promising separation technique in glycoproteomics [240–243]. Additionally, CE coupling to MS will still be an important area of investigation in the near future, as the most recent developed sheathless interfaces still need to be perfected. Furthermore, capillary coating is still relevant in CE-based glycosylation studies [244–246], and the search of more stable, compact and homogenous coatings, which will allow improved separation efficiencies, will surely be important in the future.

In addition to the already mentioned advances in separation techniques, MS/MS approaches for the structural characterization of glycans have been important for more than a decade and will continue to be an important strategy, even more with the recent advances in MS instrumentation. Particularly, for the analysis of glycoconjugates, collision-induced dissociation (CID) and electron transfer or capture dissociation (ETD or ECD) have been widely employed for the fragmentation of glycans and glycopeptides, in order to characterize their structure [105, 136, 247]. However, each fragmentation approach provides different and complementary information, thus, to obtain reliable structural information, sometimes, both approaches must be used [105, 136, 248, 249]. Moreover, multiple reaction monitoring (MRM) is a common practice nowadays to maximize sensitivity in targeted glycoproteomics [222, 250]. Recently, MS/MS approaches have also shown to be important in the differentiation of

isomeric glycans [147–151], although obtaining specific fragments for each isomer is still a difficult task.

Separation of isomeric glycans has also been an important research area in the last few years, mainly due to the fact that differences in the abundance of some glycan isomers could be the key for the early diagnosis, control or differentiation of certain diseases, such as cancer [251–254]. Therefore, new techniques and methods that will permit the detection of different glycoconjugates isomers in a straightforward manner will be fundamental in the years to come. In this regard, HILIC and PGC stationary phases, as well as CE, have already been evaluated for the separation of isomeric glycoconjugates, being PGC the most promising one [241, 255–257]. Additionally, IM-MS has proved, quite recently, to be a proficient analytical technique in the separation of isomeric glycoconjugates, with few, very recent publications about their separation and characterization [140–143, 152–155]. Despite its promising application on isomer separation, to obtain complete and reliable information about the glycan isomers of a certain glycoprotein, IM-MS will need to be complemented with other studies, such as the use of specific sialidases or MS/MS analysis.

For quite some time, MS-based methodologies aimed to study the glycosylation of proteins at glycopeptide and glycan levels (bottom-up strategies). However, in recent years, with the implementation of new hybrid spectrometers, such as the Orbitrap Fusion™ Lumos™ Tribrid™ (Thermo Scientific™), the interest in top-down (glyco)proteomic analysis has been revitalized [110–112, 258–260]. This alternative approach is appealing as not only the glycosylation of proteins can be studied, but also the interaction between the glycoprotein and other molecules or organisms can be evaluated. Furthermore, the laborious digestion and/or derivatization protocols, inherent in glycopeptide and glycan analysis, can be avoided.

In regards to the importance of glycosylated proteins in the organism, it has been already introduced that glycoproteins are responsible for a myriad of biological processes, reason why,

a deficiency or alteration of the glycosylation of certain proteins can have dire consequences and be the cause of several diseases. For instance, it is widely accepted that the glycosylation of acute phase proteins, such as transferrin, might be subjected to marked changes during inflammation. Several authors have shown that generally an increase of branching, sialylation or fucosylation occurs during an inflammatory process [38, 40, 68–71]. However, these changes in glycosylation are not exclusive of an inflammatory process, as it has been suggested that other diseases might affect the glycosylation in a similar way [38, 69]. Therefore, the search of more specific glycosylated biomarkers that will help to differentiate an inflammatory process from other diseases might be the focus of future research.

Closely tied with inflammation, cardiovascular diseases and strokes have been described to be associated with changes in glycosylation, specifically with increased levels of total serum sialic acid [40, 261]. Additionally, glycosylation has been reported to have an important role in the development and aggravation of neurodegenerative diseases, such as Alzheimer [262–265]. Although still under-researched, aberrant N-glycosylation, in certain key proteins, has been suggested to be involved in Alzheimer disease [262, 263]. Moreover, O-GlcNAc glycosylation (a type of O-glycosylation), which is suspected to have a neuroprotective role, has been observed to decrease in certain brain proteins in neurodegenerative diseases [263–265].

However, among the most worrying diseases that affect the normal glycosylation pattern of proteins is cancer [252, 253, 266–269], which is the second leading cause of death in developed countries. For instance, increased fucosylation has been observed in several glycoproteins, such as haptoglobulin and fetuin A, in liver cancer [270, 271] and the glycosylation of alpha-1-acid glycoprotein has been described to be altered in several types of cancer, such as pancreatic [251] or ovarian cancer [272], among others. Additionally, several studies have also reported an increase in fucosylation and sialylation of prostate specific antigen in prostate cancer [273, 274]. Moreover, the implication of glycosylation in cancer has

been broadened quite recently as it has been suggested that the initiation, advancement in the organism and metastasis of cancer is also related with changes in glycosylation [270]. Consequently, this increasing implication of glycosylation during all steps of tumor progression in various cancers must be further investigated, which will surely make cancer the main focus of glycomic and glycoproteomic studies in the following years.

In the context of glycosylation-related diseases, CDG can be considered one of the clearest examples of the expanding nature that is quite often seen in glycoproteomics. Even though over 80 disorders have been already described [49, 51, 275, 276], new types are continuously being discovered due to the fact that new techniques and instrumentation are being employed in the study of CDG. Some of these new types are complex multisystemic diseases that might affect different pathways in the synthesis of both N- and O-glycosylated proteins. Consequently, new types will certainly be discovered in the following years. In this regard, a novel group of glycosylation defects have emerged quite recently, the so-called secondary disorders of glycosylation, which are, in general, rare autosomal recessive disorders that affect the metabolism of certain carbohydrates, such as galactosemia or hereditary fructose intolerance. Further research of the effects causing the alteration in these diseases as well as the search of new treatment methods and more sensitive biomarkers will surely be important topics in glycoproteomics in the following years, which could further our understanding of the role and importance of glycosylation in the organism. Additionally, it is generally accepted that glycoconjugate-related diseases affect the metabolism of glycoconjugates mainly due to alterations in their synthesis, transport, glycosylation or secretion. However, little is known about the mechanism involving the degradation or elimination of these glycoconjugates and how it might be affected by these diseases [31]. Further research on this topic may improve our understanding of how these diseases might alter the normal function of the organism.

Chapter 2. Glycosylation study of glycoproteins by capillary liquid chromatography mass spectrometry

The study of the **glycosylation** of molecules, which is, of course, the focus of **glycomics**, has become an important topic nowadays, with multiple works being published every year. In complex organisms, glycosylated proteins, i.e. **glycoproteins**, are the most abundant and important glycoconjugates. Glycosylation of proteins is a vital **posttranslational modification** (PTM) in complex organisms, involved in numerous biological processes from signaling to cell-cell interaction or protein folding. When focusing on the study of the glycosylation of proteins, glycomics is usually referred to as **glycoproteomics**.

It should come as no surprise, then, giving the implication of glycosylation in the proper functioning of the organism, that the main reason behind the several medical complaints observed in various biological or pathological processes is the alteration of the standard glycosylation pattern of glycoproteins. Additionally, as glycoproteins are important in a multitude of biological processes, some relevant biopharmaceuticals are, in fact, **recombinant glycoproteins**. The glycosylation of recombinant glycoproteins is quite often different than that of their endogenous homologous. Consequently, the characterization of protein glycosylation is mandatory not only to comprehend, diagnose and propose effective treatments to the myriad of diseases associated with glycosylation defects, but also to guarantee the quality and proper function of the biopharmaceutical.

Microscale separation techniques, such as capillary liquid chromatography (**CapLC**), coupled to mass spectrometry (**MS**), can be regarded as a powerful analytical approach for the proper characterization of the glycosylation of proteins. The combination of both analytical techniques offers high quantification, reproducibility and robustness in the separation of glycoconjugates and unequivocal identification and characterization of their **glycoforms** on the basis of exact mass measurements.

Protein glycosylation can be studied by MS at three different levels, with the analysis of the intact protein, the obtained **glycopeptides** after proteolysis or the released **glycans**. However, the analysis of glycopeptides allows not only to characterize the glycans of the glycoprotein but also to obtain information about glycosylation sites and their degree of occupancy.

Three different works have been published and included in this chapter of the thesis that focus on the analysis of the glycosylation of transferrin glycopeptides by CapLC-MS and the application to the diagnosis and study of alcohol dependence, congenital disorders of glycosylation and arthritis:

- **Publication 2.1:** Analysis of human transferrin glycopeptides by capillary electrophoresis and capillary liquid chromatography-mass spectrometry. Application to diagnosis of alcohol dependence. Albert Barroso, Estela Giménez, Fernando Benavente, José Barbosa, Victoria Sanz-Nebot. *Analytica Chimica Acta* 2013, **804**, 167-175.
- **Publication 2.2:** Classification of congenital disorders of glycosylation based on analysis of transferrin glycopeptides by capillary liquid chromatography-mass spectrometry. Albert Barroso, Estela Giménez, Fernando Benavente, José Barbosa, Victoria Sanz-Nebot. *Talanta* 2016, **160**, 614-623.
- **Publication 2.3:** 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. Antonio Rosal-Vela, Albert Barroso, Estela Giménez, Sonia García-Rodríguez, Victoria Longobardo, Jorge Postigo, Marcos Iglesias, Antonio Lario, Jesús Merino, Ramón Merino, Mercedes Zubiaur, Victoria Sanz-Nebot, Jaime Sancho. *Journal of Proteomics* 2016, **134**, 127-137.

Analytica Chimica Acta 804 (2013) 167–175

Contents lists available at ScienceDirect

Analytica Chimica Acta

journal homepage: www.elsevier.com/locate/aca

Analysis of human transferrin glycopeptides by capillary electrophoresis and capillary liquid chromatography-mass spectrometry. Application to diagnosis of alcohol dependence

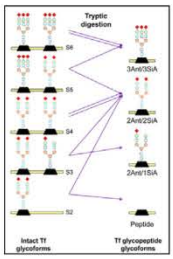
Albert Barroso, Estela Giménez*, Fernando Benavente, José Barbosa, Victoria Sanz-Nebot

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

HIGHLIGHTS

- CE-TOF-MS and μLC-TOF-MS methods separate and detect Tf glycopeptide glycoforms.
- The surfactant used to enhance the enzymatic digestion interferes in CE analyses.
- Tf alteration in alcoholism is mainly due to entire loss of the carbohydrate chains.
- Ethanol intake has no effect in the activity of sialyltransferase and sialidase.
- This method could be used to elucidate Tf glycosylation in different subtypes of CDGs.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 8 July 2013
Received in revised form
17 September 2013
Accepted 20 September 2013
Available online 9 October 2013

Keywords:
Capillary liquid chromatography
Transferrin
Glycopeptides
Glycoforms
Mass spectrometry
Alcohol dependence

ABSTRACT

In this study, capillary electrophoresis and capillary liquid chromatography coupled to mass spectrometry (CE-TOF-MS and μLC-TOF-MS) were used to detect and characterise human transferrin (Tf) glycopeptide glycoforms obtained by tryptic digestion. After selecting μLC-TOF-MS because of improved performance in analysis of N_{413} and N_{611} glycopeptide glycoforms, the proposed methodology was applied to serum samples. Two immunoaffinity columns were employed to isolate Tf from serum samples. Both columns were activated with the same anti-Tf antibody but using two different bonding chemistries. After immunoaffinity purification and digestion, serum samples from a teetotal individual (as control) and from individuals with low and high alcohol dependence were analysed by μLC-TOF-MS. Relative abundance of each glycoform was useful to estimate the degree of alcohol dependence of each individual. Finally, the established methodology was used to analyse serum samples from specific individuals with an unknown degree of alcohol dependence.

1. Introduction

Human transferrin (Tf) is an iron-binding serum glycoprotein of ~80 kDa, the main function of which is the transport of iron through the blood plasma [1]. Approximately 5.8% of its total molecular mass corresponds to one O-glycan with one hexose unit at serine 32, and two complex type N-glycans attached to asparagines 413 and 611 of the polypeptide backbone, respectively. Consequently, owing to the microheterogeneity associated with complex type N-glycans, Tf exists as a mixture of glycoforms which differ in composition, structure and charge. The main glycoform of Tf is tetrasialoform (S4) and it presents two disialylated biantennary glycans (globally, it shows 4 antennae with one terminal sialic acid each (SiA = N-acetylneuraminic acid)). The less abundant

* Corresponding author. Tel.: +34 934039123; fax: +34 934021233.
E-mail address: estelagimenez@ub.edu (E. Giménez).

0003-2670/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.aca.2013.09.044

glycoforms in a healthy individual comprise disialoform (S2) with a disialylated biantennary glycan (2Ant/2SiA), trisialoform (S3) with a disialylated and a monosialylated biantennary glycan (2Ant/2SiA and 2Ant/1SiA), pentasialoform (S5) with one biantennary and one triantennary glycan (2Ant/2SiA and 3Ant/3SiA) and hexasialoform (S6) with two fully sialylated triantennary glycans (3Ant/3SiA). S1 and S0 glycoforms, with less than 2 sialic acid residues, are barely existent in a healthy individual [2–4]. These unusual glycoforms are known as carbohydrate deficient transferrin (CDT) and they have been proposed as biomarkers for the diagnosis of different diseases such as chronic alcohol dependence or congenital disorders of glycosylation (CDG) [2–4].

Chronic alcohol dependence represents a serious human health risk that usually ends in premature death due to different causes, such as hepatic cirrhosis, internal haemorrhage, alcohol poisoning, hepatocarcinoma, accidents or suicide. Whereas the mechanism that alters Tf glycoform composition in CDG is quite well understood [5], the process that causes abnormality in Tf glycosylation in cases of heavy alcohol consumption is still a matter of debate. On the one hand, some authors have suggested that ethanol intake may alter the biosynthesis and/or transfer of dolichol-oligosaccharide intermediates, impeding initiation of N-linked glycosylation and thus causing the loss of the entire carbohydrate chain, as happens in CDG type I [6,7]. On the other hand, other authors have proposed that ethanol also decreases the activity of sialyltransferase and stimulates sialidase activity, which may result in partial N-linked oligosaccharide or sialic acid loss [8,9].

Capillary electrophoresis (CE) and anion-exchange liquid chromatography (LC) with UV detection are the current techniques employed for the analysis of CDT glycoforms [4]. However, UV detection does not permit unequivocal identification of Tf glycoforms [10]. Moreover, these methodologies are insufficient to fully clarify the mechanism behind transferrin carbohydrate abnormality, not only in cases of heavy alcohol consumption but also in some CDG subtypes. Nowadays, mass spectrometry (MS) is the most reliable way to obtain structural information about protein glycosylation [11,12]. The analysis of Tf glycoforms by MS can be approached either by analysing the intact glycoprotein [10] or by analysing the glycans or glycopeptides obtained by enzymatic digestion [13,14]. However, the first approach is rather challenging due to difficulties in detecting intact glycoproteins with sufficient sensitivity by MS [15,16]. The detection of specific glycosylation markers of lower molecular mass (glycopeptides or glycans) could be a better alternative for detecting CDT glycoforms by MS. In this context, most authors have studied Tf glycoforms through analysis of released N-glycans. However, in contrast to glycans, glycopeptides not only provide information about the structure and composition of the oligosaccharides, but also about glycosylation sites and their degree of occupancy. Some authors have studied CDT glycoforms by MALDI-MS; however, under typical vacuum source conditions, MALDI may result in dissociation of labile glycosidic bonds in glycan and glycoconjugate analytes [3,13,17–20]. Because it can be tuned to be considerably softer than MALDI, ESI seems to be more appropriate for profiling intact native glycans or glycopeptides without dissociation of labile sugars such as sialic acid and fucose.

In the present study, capillary electrophoresis and capillary liquid chromatography coupled to electrospray time-of-flight mass spectrometry (CE-TOF-MS and μLC-TOF-MS) were evaluated for the analysis of human Tf glycopeptides. Due to the complexity of serum samples, prior to enzymatic digestion and analysis by μLC-TOF-MS, an off-line immunoaffinity chromatography purification of the glycoprotein was performed using an anti-Tf polyclonal antibody. After desalting and tryptic digestion, the resulting digests were analysed by μLC-TOF-MS, obtaining a glycopeptide glycoform map of Tf with which teetotallers could be distinguished from individuals with different degrees of alcohol dependency.

2. Materials and methods

2.1. Chemicals

All chemicals used in the preparation of buffers and solutions were of analytical reagent grade. Isopropanol (iPrOH), acetic acid (HAc, glacial), formic acid (HFor, 98–100%), trifluoroacetic acid (TFA, ≥99%), glycine (≥99.7%), ammonia (25%), ammonium acetate (NH_4Ac) and sodium hydroxide were supplied by Merck (Darmstadt, Germany). CNBr-activated Sepharose 4B was from GE Healthcare (Waukesha, WI, USA). DL-Dithiothreitol (DTT, ≥99%), iodoacetamide (IAA), ammonium hydrogen carbonate, water (LC-MS grade) and acetonitrile (LC–MS grade) were supplied by Sigma–Aldrich (Madrid, Spain). Ethylenediaminetetraacetic acid (EDTA, ≥99%) was supplied by Panreac (Barcelona, Spain). Trypsin (sequencing grade modified) was provided by Promega (Madison, WI, USA). RapiGest® from Waters (Bedford, MA, USA) was used to enhance enzymatic digestion. Goat polyclonal antibody against Tf (immunogen affinity purified) was purchased from Abcam (Cambridge, UK). ESI low concentration (ESI-L) tuning mix was supplied by Agilent Technologies (Waldbronn, Germany) for tuning and calibration of the oa-TOF mass spectrometer.

2.2. Tf samples

Tf standard was purchased from Sigma–Aldrich (Madrid, Spain). Tf standards were prepared at 1500 mg L^{-1} in 50 mM NH_4HCO_3 buffer (pH 7.9).

Serum samples with a non-pathological profile (SERODOS® plus) and serum control samples from individuals with high and low alcohol dependence were purchased from Human GmbH (Wiesbaden, Germany) and Bio-Rad (Hercules, CA, USA), respectively. Two serum samples from individuals with an unknown degree of alcohol dependence were provided by Balagué Center S.A. (Barcelona, Spain).

Tf standards as well as serum samples previously purified by immunoaffinity chromatography were reduced, alkylated and immediately subjected to enzymatic digestion. Briefly, 3.75 μL of 0.5 M DTT in 50 mM NH_4HCO_3 buffer was added to an aliquot of 100 μL of Tf 1500 mg L^{-1} with 0.1% (w/v) RapiGest®. The mixture was incubated in a TS-100 Thermo-Shaker digester (Bio-San, Riga, Latvia) at 56 °C for 30 min and then alkylated with 50 mM IAA for 30 min at room temperature in the dark (10.5 μL of 0.73 M IAA). Excess reagent was removed by ultracentrifugation with Microcon YM-10 (MW cut-off 10 kDa, Millipore, Bedford, MA, USA), washing 3 times with NH_4HCO_3 buffer [21,22]. The final residue was recovered from the upper reservoir by being centrifuged upside down in a new vial, and reconstituted in 100 μL NH_4HCO_3 buffer with 0.1% RapiGest®. Trypsin in an enzyme to sample ratio of 1:40 by mass was added and the mixture was carefully vortexed and incubated at 37 °C in the digester for 18 h. At the end of digestion, RapiGest® was hydrolysed to avoid MS interferences. TFA was added to a concentration of 0.5% (v/v) and the mixture was incubated in the digester at 37 °C for 45 min. Then, the solution was centrifuged for 10 min at 12,000 rpm to separate the RapiGest® residues. The supernatant was carefully collected and stored at −20 °C until analysis. pH measurements were performed with a Crison 2002 potentiometer and a Crison electrode 52-03 (Crison Instruments, Barcelona, Spain). Centrifugation procedures were carried out in a Mikro 20 centrifuge (Hettich, Tuttlingen, Germany) at room temperature.

2.3. Immunoaffinity chromatography (IAC)

2.3.1. Preparation of IAC columns

2.3.1.1. Silica-hydrazide column. The silica-hydrazide column was prepared as reported elsewhere [23]. Briefly, 0.05 g of silica-diol

particles were oxidised with HIO_4 protected from the light. The resulting aldehyde silica particles were mixed with a solution containing an excess of oxalic dihydrazide and incubated for 2 h. Afterwards, the resulting hydrazide particles were reduced with a solution containing an excess of $NaBH_4$ and the mixture was incubated for 90 min. The final activated hydrazide particles were stored at 4 °C. The commercial buffer containing anti-Tf antibody was substituted by ultracentrifugation and the antibody was oxidised as explained elsewhere [23]. Then, the activated hydrazide silica particles were mixed with the oxidised anti-Tf antibody and were allowed to react with constant stirring for 24 h at 4 °C. Afterwards, the immunoaffinity sorbent was washed twice with 0.1 M phosphate buffer (pH 7.0), twice with the same buffer containing 2 M sodium chloride (pH 7.0) and twice with water. Finally, the immunoaffinity sorbent was transferred and packed in a 1.5 mL empty plastic column. The column was stored at 4 °C filled with 10 mM Tris–HCl and 0.01% (w/v) NaN_3 buffer (pH 7.6–7.7).

2.3.1.2. Cyanogen bromide-sepharose column. The CNBr-sepharose column was prepared following the manufacturer's instructions. Briefly, 0.5 mL of cyanogen bromide-sepharose resin was washed and swelled with several aliquots of 1 mM HCl for at least 30 min, removing the supernatant each time. The commercial buffer containing the anti-Tf antibody was exchanged for the coupling buffer (0.1 M $NaHCO_3$ and 0.5 M de NaCl, pH 8.3) by ultracentrifugation as described in Section 2.2. The antibody solution was mixed with the gel for 2 h at room temperature. Afterwards, the gel was washed with coupling buffer several times, transferred to 0.1 M Tris–HCl buffer (pH 8.0) and incubated for 2 h at room temperature to block unreacted groups. The gel was transferred to a 1.5 mL empty plastic column and was extensively washed three times with two washing steps to remove the blocking solution. The first washing step consisted of washing five times with 0.1 M NaAc and 0.5 M NaCl buffer (pH 4.0) and the second washing step consisted of washing five times with 0.1 M Tris–HCl and 0.5 M NaCl buffer (pH 8.0). Finally, the column was stored at 4 °C filled with 10 mM Tris–HCl and 0.01% (w/v) NaN_3 buffer (pH 7.6–7.7) to preserve the gel and the immobilised antibody.

2.3.2. Purification of Tf by IAC

The same purification procedure was used in both immunoaffinity columns to isolate Tf from the rest of serum proteins. 50 μL of serum sample was diluted 1/4 in 10 mM Tris–HCl (pH 7.6) in order to improve antigen-antibody interaction and consequently, to ensure maximum Tf recoveries. Before passing the sample through the column, a conditioning step was carried out which consisted of washing twice with 10 mM Tris–HCl. Afterwards, the diluted serum was passed through the column five times. After washing with 10 mM Tris–HCl and 0.5 M NaCl (pH 7.6) to eliminate the non-specifically retained fraction (washing fraction), retained Tf was eluted with 100 mM glycine-HCl (pH 2.5). Eluted Tf was immediately neutralised with 0.5 M Tris. Afterwards, glycine-HCl buffer was exchanged for 50 mM NH_4HCO_3 buffer by ultracentrifugation. Finally, Tf was digested as explained previously in Section 2.2.

2.4. μLC-TOF-MS

The μLC-TOF-MS experiments were performed in the 1200 series capillary liquid chromatography system coupled to a 6220 oa-TOF LC/MS mass spectrometer with an orthogonal G1385-44300 interface (Agilent Technologies). LC and MS control, separation, data acquisition and processing were performed using MassHunter workstation software (Agilent Technologies). The oa-TOF mass spectrometer was tuned and calibrated following the manufacturer's instructions. Once a day, or even twice a day when required, a "Quick Tune" of the instrument was carried out in positive mode followed by a mass-axis calibration to ensure accurate mass assignments. In order to enhance detection sensitivity of glycopeptides, no internal recalibration was used [22]. MS measurement parameters were similar to those described for the analysis of rhEPO glycopeptides by CE-TOF-MS [22], except for the fragmentor voltage value: capillary voltage 4000 V, drying gas (N_2) temperature 200 °C, drying gas flow rate 4 L min^{-1}, nebuliser gas (N_2) 7 psig, fragmentor voltage 190 V, skimmer voltage 60 V, OCT 1 RF Vpp voltage 300 V. Data were collected in profile (continuum) at 1 spectrum s^{-1} (approx. 10,000 transients/spectrum) between m/z 100 and 3200, working in the highest resolution mode (4 GHz).

For separation, a Zorbax 300SB-C18 column (3.5 μm particle diameter, 300 Å pore diameter, 150 mm × 0.3 mm L_T × id, Agilent Technologies) was used. Experiments were performed at room temperature with gradient elution at a flow rate of 4 μL min^{-1}. Eluting solvents were A: water with 0.1% (v/v) HFor, and B: acetonitrile with 0.1% (v/v) HFor. Solvents were degassed for 10 min by sonication before use. The optimum elution programme was: solvent B from 10% to 60% (v/v) within 45 min as linear gradient, followed by cleaning and re-equilibration steps of B: 60% to 100% (v/v) (5 min), 100% (v/v) (10 min), 100% to 10% (v/v) (5 min) and 10% (v/v) (10 min). Before analysis, samples were filtered using a 0.22 μm polyvinylidene difluoride centrifugal filter (Ultrafree-MC, Millipore, Bedford, MA, USA) at 12,000 rpm for 4 min. Sample injection was performed with an autosampler refrigerated at 4 °C and the injection volume was 0.15 μL. All samples were kept at 4 °C and stored at −20 °C when not in use for a long period.

2.5. CE-TOF-MS

The CE-TOF-MS experiments were performed in a HP^{3D}CE system coupled to a 6220 oa-TOF LC/MS mass spectrometer with an orthogonal G1603A sheath-flow interface (Agilent Technologies, Waldbronn, Germany). The sheath liquid was delivered at a flow rate of 3.3 μL min^{-1} by a KD Scientific 100 series infusion pump (Holliston, MA, USA). CE control was performed using ChemStation software running in combination with MassHunter workstation software (both from Agilent Technologies) for control, data acquisition and processing of the oa-TOF mass spectrometer. The oa-TOF was calibrated as explained in Section 2.4. Instrument parameters were the same as those used for μLC-TOF-MS, except for the nebuliser gas (N_2), which in this case was 7 psig.

A bare fused-silica capillary of 70 cm total length (L_T) × 75 μm internal diameter (id) × 375 μm outer diameter (od) (Polymicro) supplied by Composite Metals Service (Worcester, England) was used for CE-TOF-MS. New capillaries were activated with 30 min of 1 M NaOH, water and background electrolyte (BGE, 50 mM HAc and 50 mM HFor, pH 2.3). Capillaries were conditioned every day by rinsing for 5 min with NaOH, 7 min with water and for 10 min with BGE. Activation and conditioning procedures were performed off-line in order to avoid NaOH entering the mass spectrometer. Electrophoretic separations were carried out at 25 °C under normal polarity (18 kV). Between runs, the capillary was conditioned for 1 min with water, 3 min with 1 M HAc, 1 min with water and 5 min with BGE. A sheath liquid of iPrOH:H_2O (50:50, v/v) with 0.05% (v/v). The sheath liquid was degassed for 10 min by sonication before use. Injection was performed hydrodynamically at 50 mbar for 15 s.

3. Results and discussion

3.1. Analysis of Tf standard

3.1.1. Analysis of Tf glycopeptides by CE-TOF-MS

In this study, aimed at characterising Tf glycopeptide glycoforms, we applied a CE-TOF-MS method that had previously been

established for the characterisation of recombinant human erythropoietin (rhEPO) and α_1-acid-glycoprotein (AGP) glycopeptides [22]. The presence of glycoforms in intact Tf implies that the glycopeptides obtained after enzymatic digestion also present glycoforms. After tryptic digestion of Tf, two N-glycopeptides with one glycosylation site each (N_{413} and N_{611}, respectively), and an O-glycopeptide with a single hexose (O_{32}-Hex) were generated. Fig. 1 shows the N-glycopeptide glycoforms expected, in a non-pathological Tf, after tryptic digestion of the intact glycoforms. Fig. 2 shows the extracted ion electropherograms (EIEs) for the O-linked glycopeptide with one hexose unit at Ser 32 (O_{32}-Hex), and the disialylated biantennary glycoforms of both N-linked glycopeptides (N_{413}-2Ant/2SiA and N_{611}-2Ant/2SiA) using the previously established CE-TOF-MS method [22]. These N-glycopeptide glycoforms are the most abundant as they come from intact Tf S4 glycoform. The amino acid coverage of the protein sequence was around 98%. However, the overall sensitivity for the peptides and glycopeptides was rather low in comparison to that obtained for AGP and rhEPO digests following the same methodology [22]. Thus, no 3Ant/3SiA glycoforms were observed for any of the N-glycopeptides, even though this glycoform comes from the second and third most abundant Tf glycoforms in a standard Tf (S5 and S6, respectively. See Fig. 1). Therefore, in order to increase the sensitivity of the method and to detect 3Ant/3SiA glycoforms, an optimisation of the fragmentor voltage value of the mass spectrometer was performed. As previously reported, glycopeptide ionisation improves with increased fragmentor voltage [22]. However, the loss of labile groups such as SiA or N-acetyllactosamine units (HexHexNAc or LacNAc) may be observed if the fragmentor voltage is too high. In this study, fragmentor voltage values of 190, 215, 240 and 270 V were tested. Glycopeptide fragmentation was observed with 240 and 270 V. The main fragments were produced after the loss of 1 SiA and 1 or 2 SiALacNAc groups for both 2Ant/2SiA N-glycopeptide glycoforms. It should be emphasised that these were due to precursor ion fragmentation inside the ionisation source and they were not native glycoforms present in the digest, as the precursor and product ions appeared at the same migration time regardless of their electrophoretic mobilities [22]. At a fragmentor voltage value of 215 V, a significant increase in sensitivity was observed compared to 190 V, and non-fragmentation of the glycopeptides was evident.

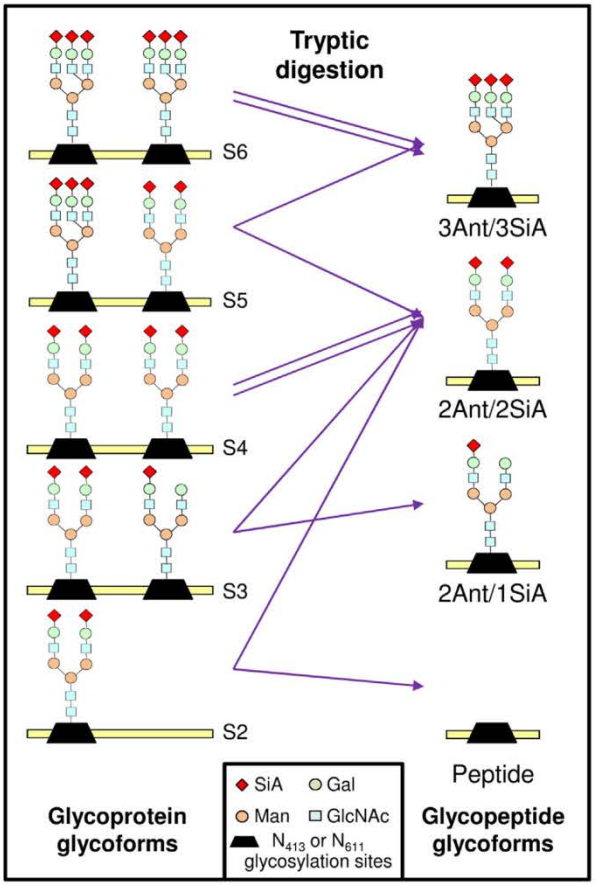


Fig. 1. N-glycopeptide glycoforms produced after tryptic digestion of intact Tf glycoforms in a healthy individual.

Nevertheless, peptide and glycopeptide sensitivity was still too low and no 3Ant/3SiA glycoforms were detected.

An optimisation of the tryptic digestion procedure was carried out in order to increase Tf digestion yield. First, reaction times were

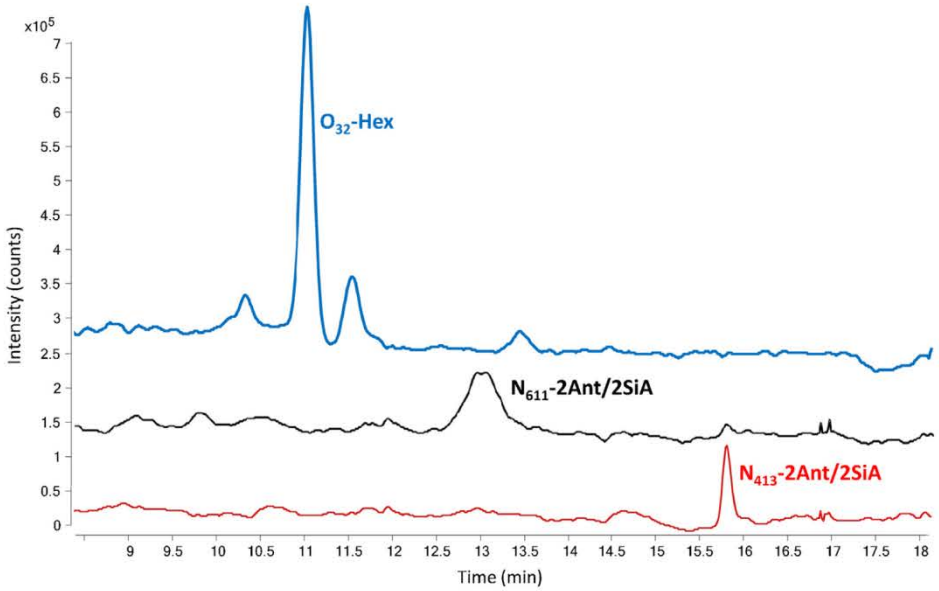


Fig. 2. Extracted ion electropherograms (EIEs) obtained for the O-linked glycopeptide (O_{32}-Hex) and the disialylated biantennary glycoforms of both N-linked glycopeptides (N_{413}-2Ant/2SiA and N_{611}-2Ant/2SiA) in Tf standard by CE-TOF-MS.

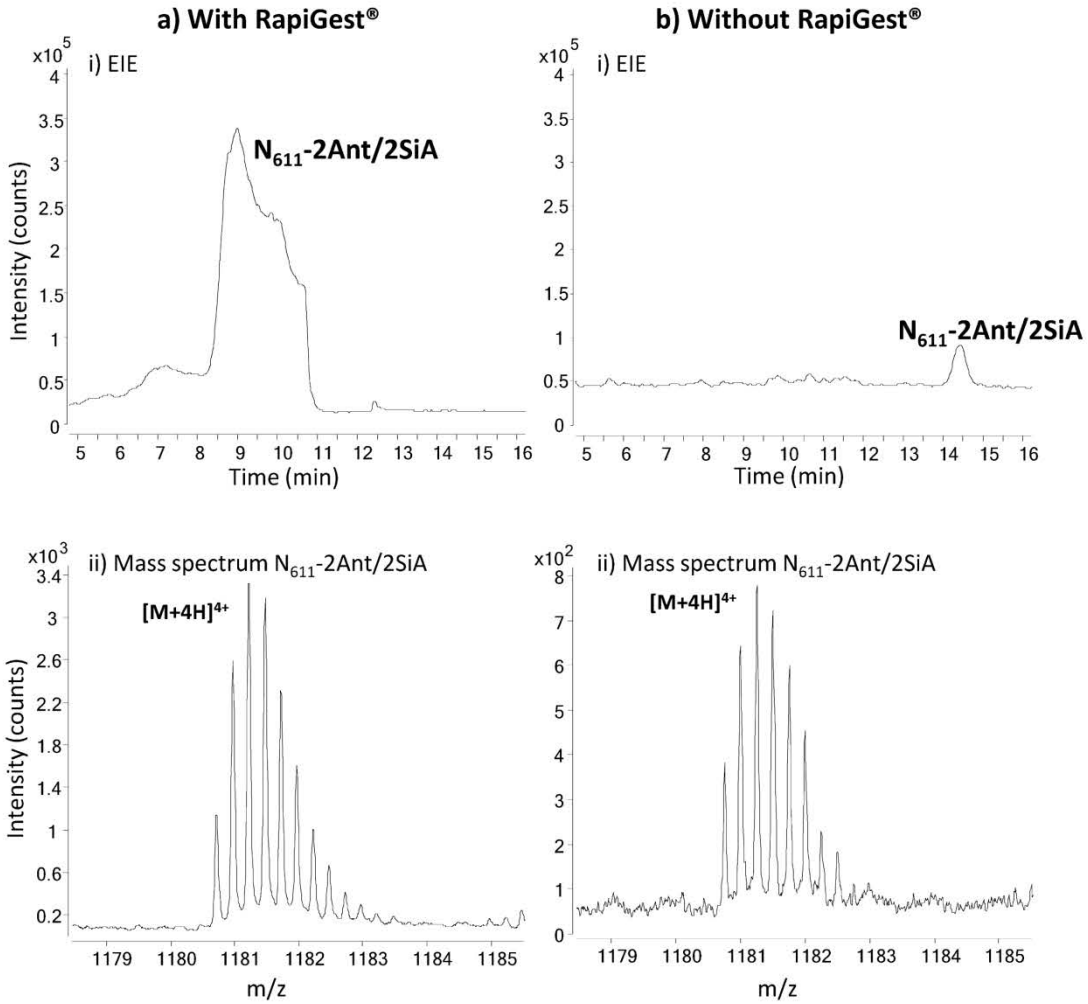


Fig. 3. (i) Extracted ion electropherograms (EIEs) and (ii) mass spectra for N_{611}-2Ant/2SiA glycoform (a) with RapiGest® and (b) without RapiGest® by CE-TOF-MS.

increased and some changes in temperatures and reagents were introduced following the method described by Zaneer M. Segu et al. [24]. The temperature and incubation time with DTT were modified (from 56 °C to 60 °C and from 30 min to 45 min, respectively), as well as the volume of DTT added (from 1.25 μL to 2.75 μL). The incubation time with IAA was also modified (from 30 min to 45 min) and the final concentration of trypsin used for the digestion procedure was increased from an enzyme to sample ratio of 1:40 to 1:30. However, no improvements in digestion yields were observed. Some authors have suggested that metalloproteins such as Tf may show increased proteolytic resistance because the bond between the glycoprotein and iron may prevent the reducing reagent and the enzyme from reaching the polypeptide chain, leading to ineffective digestion. In order to improve protein denaturation, Tf samples containing (a) 25 μM EDTA [25], (b) 0.1% TFA [26], or (c) 4 M guanidine [27], were heated for 5–10 min at 70–90 °C before reduction and alkylation. Nevertheless, none of these three treatments improved the enzymatic digestion yield and the 3Ant/3SiA glycoform was not detected under any of the conditions. As an alternative, a surfactant known as RapiGest® was evaluated to enhance the digestion of Tf. This surfactant has been reported to improve the solubilisation of complex proteins, making them more susceptible to enzymatic cleavage without inhibiting enzyme activity and without modifying peptides. Furthermore, it is compatible with MS detection [28]. The results obtained by CE-TOF-MS using this surfactant showed an improvement in sensitivity, with peptide signals increasing threefold. In addition, all minor glycopeptide glycoforms expected in Tf standard were detected: 3Ant/3SiA, 2Ant/1SiA and the peptide. The CE-TOF-MS method permitted separation of the glycoforms from the N-glycopeptide according to their mass-to-charge ratios (data not shown). However, peptide and glycopeptide electrophoretic peaks were distorted and reproducibility was low, probably due to the interaction of the surfactant with the inner wall of the fused silica capillary. To illustrate this, Fig. 3 shows the EIEs obtained for the N_{611}-2Ant/2SiA glycoform with and without the use of RapiGest®. Hence, this surfactant favoured the detection of Tf glycoforms but its use was detrimental to CE-TOF-MS analysis.

3.1.2. Analysis of Tf glycopeptides by μLC-TOF-MS

μLC was evaluated as an alternative to CE for achieving an adequate characterisation of Tf glycopeptide glycoforms obtained by tryptic digestion assisted by use of RapiGest®. An elution gradient, using acetonitrile:water with 0.1% (v/v) formic acid as mobile phase, was optimised to allow appropriate separation of both N_{413} and N_{611} glycopeptides. Fig. 4 shows the extracted ion chromatograms (EICs) and the mass spectra of 2Ant/2SiA and 3Ant/3SiA glycoforms of both Tf N-glycopeptides. As can be observed, excellent separation was obtained between both N_{413} and N_{611} glycopeptides without peak distortion. As expected, separation between glycoforms from the same glycopeptide was not observed when using C18 reversed-phase chromatography (see Fig. 4, with the 2Ant/2SiA and 3Ant/3SiA glycoforms appearing at

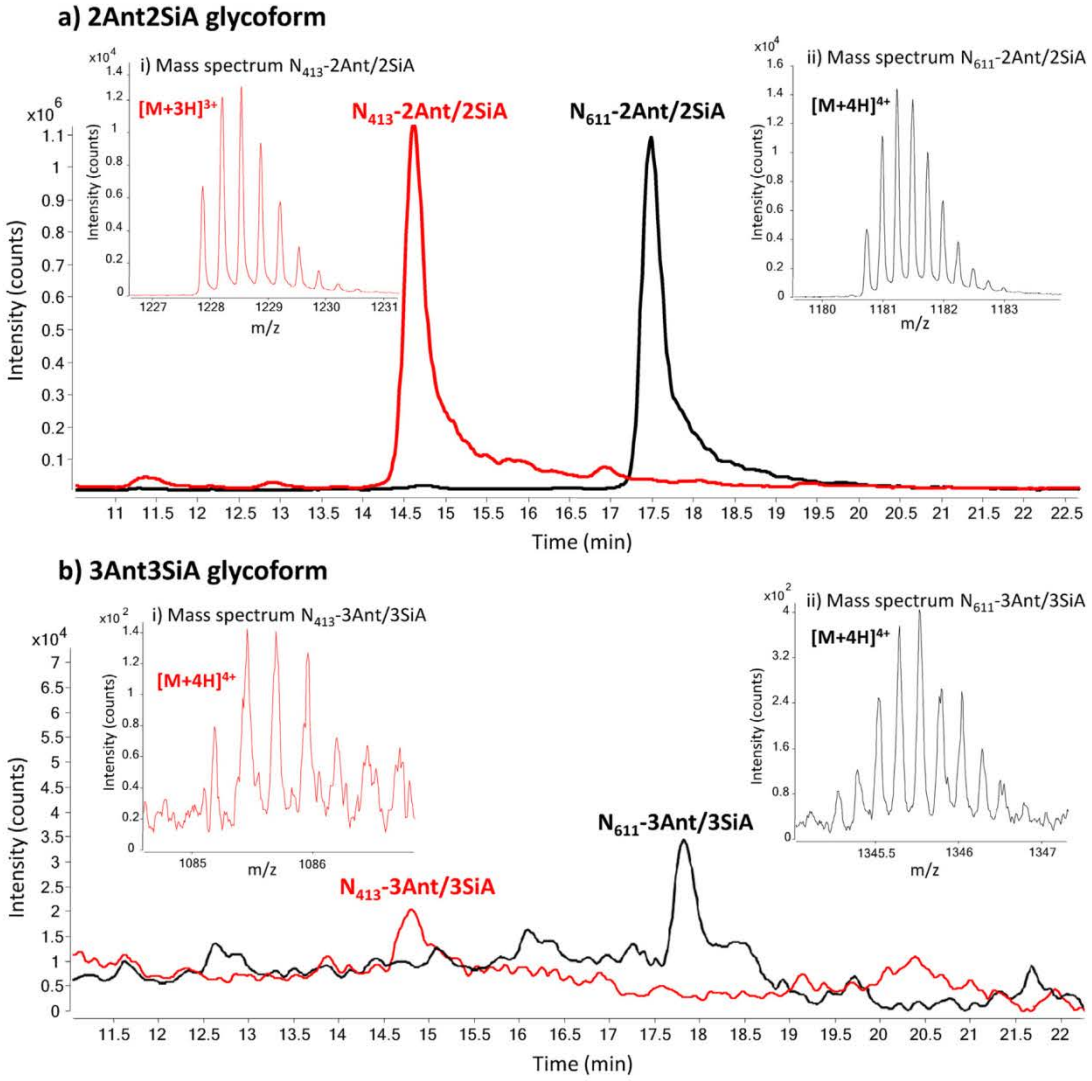


Fig. 4. Extracted ion chromatograms (EICs) and mass spectra obtained for (a) 2Ant/2SiA and (b) 3Ant/3SiA glycoforms of both N_{413} and N_{611} glycopeptides by μLC-TOF-MS.

Table 1
Detected glycoforms for both N-glycopeptides in Tf standard by μLC-TOF-MS.

		M_{theo}[a] (Da)	M_{exp}[b] (Da) ($n=6$)	Mass error[c] (ppm)	Observed ion charges	t_R[d] (min) ($n=6$)	RSD (%) (t_R)	A_{norm}[e] ($n=6$)	RSD (%) (A_{norm})
Glycopeptide	Glycoform								
N_{413}	Peptide	–	–	–	–	–	–	–	–
	2Ant/1SiA	3389.4213	3389.4130	2.4	+2, +3, +4	14.3	1.1	17.0	3.3
	2Ant/2SiA	3680.5167	3680.5081	2.3	+2, +3, +4	14.4	1.1	70.6	3.2
	2Ant/2SiAFuc	3826.5746	3826.5796	1.3	+3, +4	14.3	1.9	1.2	13.9
	3Ant/3SiA	4336.7443	4336.7527	1.9	+4	14.3	1.3	1.0	7.2
N_{611}	Peptide	2514.1169	2514.1022	5.8	+2, +3	17.8	0.9	0.6	13.2
	2Ant/1SiA	4427.7939	4427.7875	1.4	+3, +4	16.9	1.0	14.5	5.8
	2Ant/2SiA	4718.8893	4718.8838	1.2	+3, +4, +5	17.1	0.9	74.5	1.6
	2Ant/2SiAFuc	4864.9472	4864.9336	2.8	+3, +4	16.9	0.9	4.8	2.2
	3Ant/3SiA	5375.1169	5375.1201	0.6	+4, +5	17.1	0.8	1.9	5.0

[a] M_{theo}: Theoretical mass.
[b] M_{exp}: Experimental mass.
[c] Mass error (ppm) = $\left| \frac{M_{theo} - M_{exp}}{M_{theo}} \times 10^6 \right|$.
[d] t_R: Retention time.
[e] Normalized peak areas were calculated as: (Glycoform peak area/peptide 197–206 peak area) × 100. $t_{R\ peptide\ 197-206}$ = 10.4 (±1.3%).

the same retention time). Table 1 shows the glycoforms of both N-glycopeptides detected in a standard of Tf by μLC-TOF-MS, together with their theoretical and experimental masses (M_{theo} and M_{exp}), mass error of the M_{exp} (mass error), observed ion charges and retention time (t_R). Glycoform peak areas measured from their EICs were normalised (A_{norm}) using peptide 197–206 from the tryptic Tf digest as an internal standard. As can be observed in Table 1, the most abundant glycoform was 2Ant/2SiA for both N-glycopeptides. This is in agreement with intact Tf analysis, as this glycan structure is present in the main Tf glycoform in healthy individuals (the S4 glycoform), and also in S3 and S5 intact Tf glycoforms (see Fig. 1). Furthermore, 2Ant/1SiA and 3Ant/3SiA glycoforms were also detected in both glycopeptides, although to a lesser extent as these structures are only present in the less abundant glycoforms S3 and S5–S6 of Tf standard, respectively. Despite being theoretically less abundant, 2Ant1SiA showed higher A_{norm} compared to 3Ant3SiA. This may be because glycoforms with a higher content of sialic acids are more difficult to ionise and consequently, their signals can be lower than expected. Moreover, the peptide (or non-glycosylated glycoform) was detected in N_{611} glycopeptide, although in a very low amount. Detection of the peptide can be explained as being due to the presence of a low amount of S2 glycoform in a non-pathological Tf (Fig. 1) [3,4].

3.2. Analysis of serum samples

3.2.1. Tf purification by immunoaffinity chromatography

A clean-up treatment for purification of Tf was obligatory before the analysis of serum samples. Nowadays, there are different kits that allow the elimination of the most abundant serum proteins, albumin and immunoglobulins (Ig), to avoid their interference in the separation and detection of the protein of interest [29,30]. In this study, the use of some of these albumin and Ig depletion cartridges was discarded since, based on our previous experience, they do not remove these proteins sufficiently and they interfere with MS [10]. Immunoaffinity chromatography (IAC) was used as an improved alternative to the commercial kits, using an antibody (Ab) against human Tf to isolate this glycoprotein from the rest of serum proteins.

Two different IAC columns were prepared, in both cases using the same polyclonal anti-Tf Ab, but a different coupling chemistry. The first one, called a silica-hydrazide column, was prepared using a stationary phase of silica derivatised with hydrazide groups. The anti-Tf Ab was immobilised through the carbohydrate chains present in its heavy chain. This immobilising procedure has been recommended by some authors because it improves orientation of the Ab in the immunoaffinity sorbent [23,31]. The second one, called a CNBr-sepharose column, consisted of a sepharose gel derivatised with CNBr groups and the Ab was immobilised through the amino groups present in lysine and arginine amino acid residues. Hence, in this case, the active sites of the Ab were not supposed to be as well oriented as in the silica-hydrazide column. To evaluate Tf recoveries, a solution of 1500 mg L^{-1} Tf standard was passed through both columns and the eluted Tf was digested and analysed by μLC-TOF-MS. As a reference, a 1500 mg L^{-1} Tf standard digest was also analysed by μLC-TOF-MS. The EIC peak areas of the 2Ant/2SiA glycoforms of both N_{413} and N_{611} glycopeptides were used to estimate the recoveries of the IAC columns (recovery: area of the 2Ant/2SiA after passing 1500 mg L^{-1} Tf solution through the column divided by the area of the same glycoform directly analysed by μLC-TOF-MS). The analysis of blank samples permitted confirmation that none of the IAC columns showed carry-over between samples. Nevertheless, higher recoveries were obtained with the sepharose column (34% and 13% recovery for CNBr-sepharose and silica-hydrazide columns, respectively). Hence the CNBr-sepharose column was chosen for sample clean-up.

To assess the reproducibility of sepharose column preparation, three CNBr-sepharose columns were prepared using the same lot of CNBr-sepharose resin and anti-Tf Ab. The obtained recoveries were: 32% (±9%), 35% (±3%) and 16% (±7%) ($n = 2$). Therefore, the second column was selected to continue our studies. Moreover, to increase column recovery, a 1500 mg L^{-1} solution of Tf standard was diluted 1/8 instead of 1/4 (see Section 2.3.2), before passing through the IAC column, and it was reconstituted to the same volume to maintain the concentration before the digestion procedure. Using the 1/8 sample dilution, recovery increased to a value of 44%. It has been reported in IAC that in general, diluting the sample to a certain extent may improve the interaction between the Ab and the target analyte, increasing recovery. Hence a dilution sample of 1/8 was selected to purify Tf before the analysis of glycopeptide glycoforms in serum samples.

3.2.2. Analysis of Tf glycopeptides by μLC-TOF-MS in serum samples from healthy and alcohol dependent individuals

As was previously explained in Section 1, two theories have been proposed as possible mechanisms of Tf alteration in individuals with alcohol dependence: some authors have suggested that ethanol intake may impede initiation of N-linked glycosylation, thus causing the loss of the entire carbohydrate chain [6,7]. Others have proposed that ethanol intake may result in partial N-linked oligosaccharide or sialic acid loss [8,9]. To determine which mechanism is present in alcohol dependence, we studied the different glycopeptide glycoforms expected for both mechanisms of Tf alteration. Serum samples from healthy individuals (as control) and from individuals with different degrees of alcohol dependence were analysed following the previously established methodology (IAC purification, tryptic digestion and μLC-TOF-MS analysis). Moreover, blanks were also analysed in order to discard the possibility of carry-over between samples by IAC and between injections by μLC-TOF-MS. Table 2 shows the glycoforms of both N_{611} and N_{413} glycopeptides detected in the different serum samples by μLC-TOF-MS, as well as their normalised area (A_{norm}) and %RSD values. Despite serum sample complexity, the healthy control gave similar results to those previously obtained with Tf standard (compare Tables 1 and 2). A_{norm} values were very similar for all glycopeptide glycoforms, as both samples (Tf standard and healthy control) showed a similar composition of intact Tf glycoforms (S2–S6 glycoforms). The 2Ant/0SiA glycoform was also extracted, as it is one of the glycoforms that may be affected in alcohol dependent individuals. Surprisingly, this glycoform was detected in the samples from alcohol dependent individuals but also in the healthy control sample at a similar concentration. It is important to emphasise that this glycoform was present in the Tf digest and was not an artefact caused by ion-source fragmentation, as it was also detected by CE-TOF-MS in a different migration time from the rest of glycoforms (2Ant/2SiA or 2Ant/1SiA) (data not shown).

The two samples from alcohol dependent controls that were analysed corresponded to a: *low control* from an individual with a low degree of alcohol dependence and hence, with a minimal alteration of Tf glycoforms (only presence of higher amounts of S2 glycoform), and *high control*, from an individual with a high degree of alcohol dependence and a high alteration of Tf glycoforms (presence of S0 and S2 CDT glycoforms). As can be seen in Table 2, a decrease in the 2Ant/2SiA glycoform was observed for both N_{611} and N_{413} glycopeptides as the individual's degree of alcohol dependence increased, which is related to a decrease in the S4 glycoform in intact Tf (Fig. 1). Furthermore, the peptides related to the S0 and S2 glycoforms in intact Tf were also clearly affected, and amounts increased at the same time that the individual's degree of alcohol dependence increased. In contrast, 2Ant/1SiA and 2Ant/0SiA glycoforms remained almost unaltered in N_{611} glycopeptide, and they decreased in N_{413} glycopeptide in

Table 2
Normalised peak area of the detected glycoforms for both Tf N-glycopeptides in healthy control and alcoholic controls (control low and high) and in individuals with different degrees of alcoholism.

Serum sample	Peptide		2Ant/0SiA		2Ant/1SiA		2Ant/2SiA		2Ant/2SiAFuc		3Ant/3SiA	
	A_{norm}	RSD (%)	A_{norm}	RSD (%)	A_{norm}	RSD (%)	A_{norm}	RSD (%)	A_{norm}	RSD (%)	A_{norm}	RSD (%)
Normalised peak area	N_{413} glycopeptide glycoforms											
Healthy control	0.0	0.0	2.5	6.0	15.7	2.4	66.3	2.3	1.3	14.8	1.1	2.8
Alcoholic												
Control low	0.0	0.0	1.4	12.1	12.3	8.7	60.9	2.2	1.1	10.7	1.0	5.2
Control high	3.0	11.5	1.6	8.2	11.4	6.7	55.0	1.8	1.4	5.5	1.0	8.0
Individual 1	0.0	0.0	1.7	2.9	11.7	3.4	62.4	3.9	1.1	6.3	1.0	9.9
Individual 2	2.5	7.1	1.3	6.1	11.1	0.3	55.3	1.0	1.2	8.9	0.8	15.1
Normalised peak area	N_{611} glycopeptide glycoforms											
Healthy control	0.8	15.6	0.5	8.9	13.8	1.4	74.1	2.0	4.2	7.5	1.9	7.9
Alcoholic												
Control low	2.9	13.9	0.5	10.2	11.1	8.7	71.9	2.8	3.6	8.0	1.0	8.9
Control high	8.5	4.3	0.4	11.0	12.0	10.0	67.0	0.8	4.4	4.5	1.5	2.7
Individual 1	2.0	1.7	0.5	8.8	11.0	7.5	72.5	1.4	3.9	6.7	1.8	9.4
Individual 2	7.0	3.5	0.4	3.6	12.0	6.4	68.6	0.9	4.6	3.4	1.2	12.7

samples from alcohol dependent individuals. However, in any case these glycoforms increased as it would be expected if there was an increase of sialidase activity or a decrease of sialyltransferase activity, as some authors proposed [8,9]. Consequently, these results support the idea that ethanol intake mainly affects the entire synthesis and transport of N-linked oligosaccharides as happens in CDG type I [6,7].

Finally, once the capacity of the developed methodology to distinguish between samples from individuals with different degrees of alcohol dependence had been demonstrated, two serum samples from individuals with an unknown level of alcohol dependence were analysed. The results obtained are also depicted in Table 2. As can be observed with the obtained A_{norm} values for the peptides and 2Ant/2SiA glycoforms, individual 1 may show a low degree of alcohol dependence whilst individual 2 may show a heavy alcohol dependence profile, since A_{norm} values were similar to those obtained in the high control. Hence, the proposed method permitted estimation of the degree of alcohol dependence reflected in real samples. In the future, a higher number of sera from control and alcohol dependent individuals, with known percentages of CDT glycoforms determined by using the HPLC reference method [4], will be used to perform a reliable validation of the established methodology for alcohol dependence diagnosis.

4. Conclusions

CE-TOF-MS and μLC-TOF-MS methods were developed for the separation and detection of Tf glycopeptide glycoforms. The use of a commercial surfactant called RapiGest® was necessary to increase the digestion yield and consequently, to increase the detection sensitivity of Tf tryptic peptides and glycopeptides. Using both methods, the detected glycopeptide glycoforms fitted perfectly with those intact glycoforms present in a non-pathological sample. However, CE-TOF-MS separations were poorer than with μLC-TOF-MS. This is probably because the surfactant interfered with the inner wall of the fused silica capillary in CE-TOF-MS. Therefore, the μLC-TOF-MS method was preferred to analyse serum samples. Prior to tryptic digestion and μLC-TOF-MS analysis, serum samples were subjected to immunoaffinity chromatography using a CNBr-sepharose column activated with a polyclonal anti-Tf antibody. The established IAC methodology permitted purification of Tf from serum samples. The glycopeptide glycoforms detected for both N_{611} and N_{413} glycopeptides in the different serum samples from alcohol dependent and teetotal individuals suggest that the mechanism causing abnormality in the glycosylation of Tf in cases of heavy alcohol consumption impedes the initiation of N-linked glycosylation, as happens in CDG type I. This was demonstrated by a decrease in the 2Ant2SiA glycoform and an increase in the non-glycosylated glycoform, corresponding to S4 and S0 intact Tf glycoforms, respectively. Furthermore, the decrease of the 2Ant/1SiA and 2Ant/0SiA glycoforms in samples from alcohol dependent individuals seems to indicate that the cleaving of sialic acids or partial carbohydrate chains does not occur with ethanol intake. Hence, once the developed methodology has been exhaustively validated with a higher number of samples, this methodology based on characterisation of glycopeptide glycoforms could be useful not only to further improve the diagnosis of alcohol dependence, but also to elucidate and increase our understanding of the mechanism behind Tf alteration in other diseases such as congenital disorders of glycosylation (CDG), especially CDGs type II.

Acknowledgements

Part of this study was supported by the Spanish Ministry of Science and Innovation (CTQ2011-27130). Albert Barroso thanks the *Generalitat de Catalunya* for a FI fellowship. The authors acknowledge Balagué Center S.A. for providing the serum samples from alcohol dependent individuals.

References

[1] R.T. MacGillivray, E. Mendez, S.K. Sinha, M.R. Sutton, J. Lineback-Zins, K. Brew, Proc. Natl. Acad. Sci. U.S.A 79 (1982) 2504–2508.
[2] M. Ridinger, P. Köhl, E. Gäbele, N. Wodarz, G. Schmitz, P. Kiefer, C. Hellerbrand, Exp. Mol. Pathol. 92 (2011) 50–53.
[3] W. Oberrauch, A.C. Bergman, A. Helander, Clin. Chim. Acta 395 (2008) 142–145.
[4] F. Bortolotti, G. De Paoli, F. Tagliaro, J. Chromatogr. B: Analyt. Technol. Biomed. Life Sci. 841 (2006) 96–109.
[5] L. Sturiale, R. Barone, D. Garozzo, J. Inherit. Metab. Dis. 34 (2011) 891–899.
[6] S. Wopereis, S. Grünewald, E. Morava, J.M. Penzien, P. Briones, M.T. García-Silva, P.N. Demacker, K.M. Huijben, R.A. Wevers, Clin. Chem. 49 (2003) 1839–1845.
[7] M. Binkhorst, S.B. Wortmann, S. Funke, T. Kozicz, R.A. Wevers, E. Morava, J. Inherit. Metab. Dis. 35 (2012) 399–405.
[8] C. Flahaut, J.C. Michalski, T. Danel, M.H. Humbert, A. Klein, Glycobiology 13 (2003) 191–198.
[9] N. Waszkiewicz, S.D. Szajda, A. Zalewska, A. Szulc, A. Kępka, A. Minarowska, M. Wojewódzka-Żelezniakowicz, B. Konarzewska, S. Chojnowska, J.R. Ladny, K. Zwierz, Folia Histochem. Cytobiol. 50 (2012) 1–11.
[10] V. Sanz-Nebot, E. Balaguer, F. Benavente, C. Neusüß, J. Barbosa, Electrophoresis 28 (2007) 1949–1957.
[11] M. Wuhrer, Glycoconj. J. 30 (2013) 11–22.
[12] M.V. Novotny, W.R. Alley, B.F. Mann, Glycoconj. J. 30 (2013) 89–117.
[13] B. Pérez, C. Medrano, M.J. Ecay, P. Ruiz-Sala, M. Martínez-Pardo, M. Ugarte, C. Pérez-Cerdá, J. Inherit. Metab. Dis. 36 (2013) 535–542.

[14] E. Landberg, E. Åström, B. Kågedal, P. Påhlsson, Clin. Chim. Acta 414 (2012) 58–64.
[15] C. Neusüss, U.M. Demelbauer, M. Pelzing, Electrophoresis 26 (2005) 1442–1450.
[16] E. Giménez, F. Benavente, J. Barbosa, V. Sanz-Nebot, Electrophoresis 29 (2008) 2161–2170.
[17] L.Z. Luo, H.W. Jin, H.Q. Huang, Rapid Commun. Mass Spectrom. 25 (2011) 1391–1398.
[18] N. Leymarie, J. Zaia, Anal. Chem. 84 (2012) 3040–3048.
[19] M. Pabst, F. Altmann, Proteomics 11 (2011) 631–643.
[20] M.E. del Castillo Busto, M. Montes-Bayón, E. Blanco-González, J. Meija, A. Sanz-Medel, Anal. Chem. 77 (2005) 5615–5621.
[21] E. Giménez, F. Benavente, C. de Bolós, E. Nicolás, J. Barbosa, V. Sanz-Nebot, J. Chromaotgr. A 1216 (2009) 2574–2582.
[22] E. Giménez, R. Ramos-Hernan, F. Benavente, J. Barbosa, V. Sanz-Nebot, Rapid Commun. Mass Spectrom. 25 (2011) 2307–2316.
[23] S. Medina-Casanellas, F. Benavente, J. Barbosa, V. Sanz-Nebot, Anal. Chim. Acta 717 (2012) 134–142.
[24] Z.M. Segu, L.A. Hammad, Y. Mechref, Rapid Commun. Mass Spectrom. 24 (2010) 3461–3468.
[25] D.J. Janecki, J.P. Reilly, Rapid Comun. Mass Spectrom. 19 (2005) 1268–1272.
[26] F. Benavente, B. Andón, E. Giménez, J. Barbosa, V. Sanz-Nebot, Electrophoresis 29 (2008) 2790–2800.
[27] R. Ullmer, A.M. Rizzi, J. Mass Spectrom. 44 (2009) 1596–1603.
[28] Y.Q. Yu, M. Gilar, J. Kaska, J.C. Gelber, Rapid Commun. Mass Spectrom. 19 (2005) 2331–2336.
[29] C.S. Olver, T.L. Webb, L.J. Long, H. Scherman, J.E. Prenni, Vet. Clin. Pathol. 39 (2010) 337–345.
[30] C. Magagnotti, I. Fermo, R.M. Carletti, M. Ferrari, A. Bachi, Clin. Chem. Lab. Med. 48 (2010) 531–535.
[31] P.F. Ruhn, S. Garver, D.S. Hage, J. Chromatogr. A 669 (1994) 9–19.

Talanta 160 (2016) 614–623

Contents lists available at ScienceDirect

Talanta

journal homepage: www.elsevier.com/locate/talanta

Classification of congenital disorders of glycosylation based on analysis of transferrin glycopeptides by capillary liquid chromatography-mass spectrometry

Albert Barroso, Estela Giménez, Fernando Benavente*, José Barbosa, Victoria Sanz-Nebot

Department of Chemical Engineering and Analytical Chemistry, University of Barcelona, Diagonal 645, 08028 Barcelona, Spain

ARTICLE INFO

Article history:
Received 30 April 2016
Received in revised form
22 July 2016
Accepted 24 July 2016
Available online 26 July 2016

Keywords:
Congenital disorders of glycosylation
Glycopeptide
Glycoforms
Mass spectrometry
Multivariate data analysis

ABSTRACT

In this work, we describe a multivariate data analysis approach for data exploration and classification of the complex and large data sets generated to study the alteration of human transferrin (Tf) N-glycopeptides in patients with congenital disorders of glycosylation (CDG). Tf from healthy individuals and two types of CDG patients (CDG-I and CDG-II) is purified by immunoextraction from serum samples before trypsin digestion and separation by capillary liquid chromatography mass spectrometry (CapLC-MS). Following a targeted data analysis approach, partial least squares discriminant analysis (PLS-DA) is applied to the relative abundance of Tf glycopeptide glycoforms obtained after integration of the extracted ion chromatograms of the different samples. The performance of PLS-DA for classification of the different samples and for providing a novel insight into Tf glycopeptide glycoforms alteration in CDGs is demonstrated. Only six out of fourteen of the detected glycoforms are enough for an accurate classification. This small glycoform set may be considered a sensitive and specific novel biomarker panel for CDGs.

1. Introduction

Protein glycosylation, which enhances the functional diversity of proteins and provides a highly distinct structure variation, is by far the most common and complex post-translational modification with more than half of all secretory and cellular proteins being glycosylated [1–3]. The lack or attachment of certain monosaccharides to core glycan or branches may result in the alteration of the normal function of the glycoprotein and the resulting transformation of cellular phenotypes is known to be involved in various biological or pathological processes [4].

This is the case of congenital disorders of glycosylation (CDGs), formerly known as carbohydrate-deficient glycoprotein syndromes. CDGs are a family of genetic defects caused by mutations in the genes coding for enzymes involved in the biosynthesis or remodelling of the oligosaccharide moieties of glycoconjugates [5–8]. Depending on the defective biosynthesis step, CDGs can be classified in two subgroups [9,10]. CDG type I (CDG-I) consists in defects in the glycan assembly and in the attachment of glycans to the nascent glycoprotein in the cytosol or the endoplasmic reticulum. This results in the lack of complete N-linked glycans on some glycosylation sites [8]. CDG type II (CDG-II) is the result of abnormal remodelling or processing of the glycan moieties in the Golgi network generally resulting in truncated or structural deficient carbohydrate chains [5,8,11]. Human transferrin (Tf) is a well-known biomarker of CDGs [5,12–14]. Tf (~80 kDa, ~5.8% carbohydrates) is an iron-binding serum transport glycoprotein with one O-glycan (one hexose unit at Ser32), and two complex N-glycans at Asn413 and Asn611 [15]. Due to the microheterogeneity associated with the complex N-glycans, Tf exists as a mixture of glycoforms which differ in composition, structure and charge. The main glycoform of Tf is the tetrasialoform (S4, 85% of the total amount of Tf in healthy individuals) that shows two disialylated biantennary glycans (2Ant2SiA+2Ant2SiA; antennae (Ant) and sialic acid (SiA)=N-acetylneuraminic acid). Other sialoforms such as pentasialo (S5, 2Ant/2SiA+3Ant/3SiA), hexasialo (S6, 3Ant/3SiA+3Ant/3SiA), trisialo (S3, 2Ant/2SiA+2Ant/1SiA) and disialo (S2, 2Ant/2SiA) are present at much lower concentrations in healthy individuals (S1, 2Ant/1SiA, and S0, no glycosylation, are almost inexistent) [16–18]. With regard to Tf glycosylation in CDG patients, there is a difference between CDG-I and CDG-II. CDG-I is mostly characterised by increased S2 and S0 and decreased S4 while CDG-II mostly shows increased S3, S1 and/or S0 and decreased S4 [8]. Nowadays, the diagnosis of CDGs is mainly based upon the glycoform pattern observed for intact Tf by

Abbreviations: CapLC-MS, capillary liquid chromatography mass spectrometry; Tf, human transferrin; CDG, congenital disorders of glycosylation; EIC, extracted ion chromatogram; PLS-DA, partial least squares discriminant analysis

* Corresponding author.
E-mail address: fbenavente@ub.edu (F. Benavente).

http://dx.doi.org/10.1016/j.talanta.2016.07.055
0039-9140/

isoelectric focusing (IEF) [7,19,20]. However, this method does not allow the unequivocal identification of Tf glycoforms, often lead to false negative results (25% of the identified CDG may show a normal profile of intact Tf by IEF) and sometimes it is difficult to discriminate between CDG-I and CDG-II, or from other glycosylation defects (galactosemia, fructose intolerance, alcohol abuse, etc.) [1,7]. In addition, the analysis of intact Tf is not enough to deeply investigate the mechanism altering glycosylation.

High performance separation techniques coupled to mass spectrometry (MS) are the most reliable way to obtain structural information about protein glycosylation as they allow fast and high sensitivity profiling and accurate characterisation of heterogeneous glycan structures [21–25]. In this regard, the enzymatic digestion of the glycoprotein and the detection of certain glycomarkers of lower molecular mass in the digests (glycopeptides or glycans) is a powerful alternative to detect glycoprotein glycoforms by MS instead of intact glycoproteins, whose sensitivity in MS tends to be lower [26–30]. Furthermore, the analysis of glycopeptides obtained after protein digestion provides information about the structure and composition of the glycans, as well as about the glycosylation sites and their degree of occupancy [31,32]. However, the analysis of glycopeptides of an enzymatic protein digest is a challenge, due to the complexity of the digests, which are mixtures of peptides and glycopeptides, and the microheterogeneity of the glycopeptides, which present several glycoforms [33,34]. As a result, very often, experiments involving the analysis of glycoconjugates result in complex and massive data sets which are easier to interpret using multivariate data analysis methods [35].

In this preliminary study, Tf tryptic digests from healthy individuals and CDG-I and CDG-II patients are analysed to investigate glycosylation alteration. Tf is purified by immunoextraction from serum samples before trypsin digestion and separation by capillary liquid chromatography mass spectrometry (CapLC-MS). Following a targeted data analysis approach, partial least squares discriminant analysis (PLS-DA) is applied to the relative abundance of Tf glycopeptide glycoforms obtained after integration of the extracted ion chromatograms of the different samples. The performance of PLS-DA for classification of the different samples and for providing a novel insight into how differently both N-glycosylation points of Tf can be affected in CDGs is demonstrated.

2. Materials and methods

2.1. Chemicals

All chemicals used in the preparation of buffers and solutions were of analytical reagent grade. Isopropanol (iPrOH), hydrochloric acid (HCl, glacial), formic acid (HFor, 98–100%) and glycine ($\geq$99.7%) were supplied by Merck (Darmstadt, Germany). CNBr-activated-Sepharose 4B was provided by GE Healthcare (Waukesha, WI, USA). Sodium chloride (NaCl, $\geq$99.5%), DL-Dithiothreitol (DTT, $\geq$99%), iodoacetamide (IAA), ammonium hydrogencarbonate, water (LC-MS grade) and acetonitrile (LC–MS grade) were supplied by Sigma-Aldrich (Madrid, Spain) and Tris(hydroxymethyl) aminomethane (TRIS, $\geq$99.5%) by J.T. Baker (Deventer, Holland). Trypsin (Sequencing grade modified) was provided by Promega (Madison, WI, USA). RapiGest® from Waters (Bedford, MA, USA) was used to facilitate the enzymatic digestion. Goat polyclonal antibody against Tf (immunogen affinity purified) was purchased from Abcam (Cambridge, UK). ESI low concentration (ESI-L) tuning mix was supplied by Agilent Technologies (Waldbronn, Germany) for tuning and calibration of the TOF mass spectrometer.

2.2. Tf standard and serum samples

Tf standard was purchased from Sigma-Aldrich (Madrid, Spain). Tf standards were prepared at 1500 μg mL^{-1} (~19 μM) in 50 mM NH_4HCO_3 buffer (pH 7.9).

A commercial lyophilised pool of sera with a non-pathological profile (control, SERODOS® plus) was purchased from Bio-Rad (Hercules, CA, USA). Serum control samples from young-adult healthy volunteers were obtained in the Department of Analytical Chemistry of the University of Barcelona [36]. Briefly, venous blood was collected in 9 mL Vacuette tubes (Greiner Bio-One, Frickenhausen, Germany) with Z serum separation clot activator, and then allowed to coagulate by leaving it undisturbed at room temperature for 9 h. Afterwards, the clot was kept at 4 °C for 12–16 h to improve the clot retraction. The supernatant serum was subsequently separated from the clot with a Pasteur pipette and centrifuged at 1200xg for 20 min at 4 °C. Clear serum was then separated and aliquoted to store in a freezer at −20 °C when not in use. Positive CDG-I and CDG-II serum samples from infants and young-adults were kindly provided by the Balagué Center S.A. (Barcelona, Spain), the Institute of Clinical Biochemistry (Hospital Clínic, University of Barcelona) and the Center for the Diagnosis of Molecular Diseases (Autonomous University of Madrid). The assay was approved by the Ethics Committee of the University of Barcelona.

In the last years, a novel nomenclature for the CDGs has been proposed, which connects the name of the defective gene followed by a common -CDG suffix [37]. Nine healthy controls (1–9), five CDG-I (PGM1-CDG (10); DPM1-CDG (11); RFT1-CDG (12); DPAGT1-CDG (13); PMM2-CDG (14)) and five CDG-II (ATP6V0A2-CDG (15); $CDGII_x$ (16); $CDGII_x$(sepsis) (17) and two samples of COG8-CDG (18–19)) samples were analysed for the calibration. The validation set contained nine samples (i–ix): three healthy controls (i–iii), three CDG-I (PMM2-CDG (iv), DPAGT1-CDG (v) and DPM1-CDG (vi)) and three CDG-II (ATP6V0A2-CDG (vii), CDGIIx (viii) and CDGIIx (sepsis) (ix)). The assay was approved by the Ethics Committee of the University of Barcelona.

2.3. Purification and digestion of Tf

Tf was purified from serum by immunoaffinity chromatography with a cyanogen-bromide sepharose column as described in a previous work [33]. Briefly, 50 μL of serum were diluted 1:4 in 10 mM Tris-HCl (pH 7.6). Before passing the sample through the column, a conditioning step was carried out consisting in two washes of 1 mL of 10 mM Tris-HCl (pH 7.6). Afterwards, the diluted serum was passed through the column ten times. After washing with 400 μL of 10 mM Tris-HCl and 0.5 M NaCl (pH 7.6), retained Tf was eluted with 250 μL of 100 mM glycine-HCl (pH 2.5). Eluted Tf was immediately neutralised with 0.5 M Tris pH ~11. Afterwards, the buffer was exchanged for 50 mM NH_4HCO_3 buffer by ultracentrifugation using Microcon YM-10 filters (M_r cut-off 10,000, Millipore, Bedford, MA, USA).

Tf standards and Tf purified from serum samples were reduced, alkylated and immediately subjected to enzymatic digestion in the presence of RapiGest® as explained in [33]. Briefly, 1.90 μL of 0.5 M DTT in 50 mM NH_4HCO_3 buffer was added to an aliquot of 50 μL of standard or purified Tf with 0.1% (w/v) RapiGest®. The mixture was incubated in a TS-100 Thermo-Shaker digester (Bio-San, Riga, Latvia) at 56 °C for 30 min and then alkylated with 50 mM IAA for 30 min at room temperature in the dark (5.25 μL of 0.73 M IAA). Excess reagent was removed by ultracentrifugation with Microcon YM-10 filters, washing 3 times with NH_4HCO_3 buffer. The final residue was reconstituted to the initial volume (50 μL) with NH_4HCO_3 buffer with 0.1% (w/v) RapiGest®. Trypsin in an enzyme to protein ratio of 1:40 (m/m) (considering that Tf concentration is

around 1500–3000 μg mL^{-1} in serum [38,39] from an adult) was added and the mixture was carefully vortexed and incubated overnight at 37 ºC. The surfactant was hydrolysed after incubating overnight as follows: HFor was added to the digest to a final concentration of 5% (v/v) and the mixture was incubated in the digester at 37 ºC for 30 min. Then, the solution was centrifuged for 10 min at 10,000*g* to separate RapiGest® residues. Finally, the supernatant was carefully collected and stored at −20 ºC until the analysis. pH measurements were performed with a Crison 2002 potentiometer and a Crison electrode 52-03 (Crison Instruments, Barcelona, Spain). Centrifugation procedures were carried out in a Mikro 20 centrifuge (Hettich, Tuttlingen, Germany) at room temperature.

2.4. CapLC-MS

CapLC-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 ESI interface (Agilent Technologies). LC and MS control, separation, data acquisition and processing were performed with MassHunter workstation software (Agilent Technologies). The oa-TOF mass spectrometer was tuned and calibrated following the manufacturer's instructions. Once a day, or even twice a day if working days were too long, a "Quick Tune" of the instrument was carried out in positive mode followed by a mass-axis calibration to ensure accurate mass assignments. MS measurement parameters were as indicated: capillary voltage 4000 V, drying gas (N_2) flow rate 4 L min^{-1}, drying gas temperature 200 ºC, nebuliser gas (N_2) 15 psig, fragmentor voltage 215 V, skimmer voltage 60 V, OCT 1 RF Vpp voltage 300 V. Data were collected in profile (continuum) at 1 spectrum s^{-1} (approx. 10,000 transients/spectrum) between *m/z* 100 and 3200 working in the highest resolution mode (4 GHz).

For the separation of the tryptic digests, a Zorbax 300SB-C18 column (3.5 μm particle diameter, 300 Å pore diameter, 150 × 0.3 mm $L_T$xid, Agilent Technologies) was used [33]. Experiments were performed at room temperature with gradient elution at a flow rate of 4 μL min^{-1}. Eluting solvents were A: water with 0.1% (v/v) of HFor, and B: acetonitrile with 0.1% (v/v) HFor. Solvents were degassed by sonication (10 min) before use. The optimum elution program was: solvent B from 10 to 60% (v/v) within 45 min as linear gradient, followed by cleaning and reequilibration steps of B: 60–100% (v/v) (5 min), 100% (v/v) (10 min), 100–10% (v/v) (5 min) and 10% (v/v) (10 min). Before analysis, samples were filtered using a 0.22 μm polyvinyldene difluoride centrifugal filter (Ultrafree-MC, Millipore, Bedford, MA, USA) centrifuging at 10,000 *g* for 4 min. Sample injection (0.15 μL) was performed with an autosampler refrigerated at 4 ºC.

2.5. Data analysis

LC-MS data collected for Tf purified from serum samples of the calibration and validation sets were processed to obtain the extracted ion chromatograms (EIC) corresponding to the N_{413} and N_{611} glycopeptide glycoforms (taking into account all the detected ions for each glycoform. In general: +2,+3,+4 and +5 ions, see Table 1). Glycoform peak areas were measured from the different EICs and were normalised (A_{norm}) using the peak area of Tf peptide 558–568, DYELLCLDGTR, as an internal standard, which had a similar peak area and eluted within the range of the glycoforms. Therefore, A_{norm} was a measurement of the relative abundance of the glycopeptide glycoforms.

Mean A_{norm} of each glycoform was compared for the group of controls and CDG samples. A single factor analysis of variance (ANOVA) was performed to compare the means setting the statistical significance at $P < 0.05$.

A_{norm} of the glycoforms were autoscaled (mean centered and scaled to unit standard deviation) and PCA was applied to explore the classes present in the data and the presence of outliers [40]. PLS-DA [41] was performed later to maximise class separation and rapidly classify the different samples in controls, CDG-I or CDG-II as well as to identify which glycoforms were the most significant to discriminate between these three classes taking into account the Variable Importance in the Projection (VIP) scores of the glycoforms in the PLS-DA model [42–46]. A (leave-one-out) cross validation of the PLS-DA model was performed during calibration [47,48]. In the prediction step, the model was used to classify the samples from the validation set.

ANOVA was performed with Microsoft Excel 2010 software (Microsoft Corporation, Redmon, WA). The rest of calculations were performed under MATLAB R2013a (The MathworksInc. Natick, MA, USA). SOLO (Version 8.1, student edition, Eigenvector Research Inc., Wenatchee, WA, USA) was used for PCA, PLS-DA and VIP calculations.

3. Results and discussion

In a previous paper, we established a method for the analysis of Tf glycopeptides by CapLC-MS, which included purification of Tf from serum by immunoaffinity chromatography before tryptic digestion. The method was applied to study the abnormal glycosylation of Tf in alcoholism, and it was useful to draw some conclusions about the mechanism involved [33]. In this paper, the same method was used to obtain CapLC-MS fingerprints of Tf glycopeptides from healthy controls and CDG samples.

Fig. 1 shows the extracted ion chromatograms (EICs) of the

Table 1
Theoretical molecular mass (M_r) and detected ions of the 14 glycopeptide glycoforms detected by CapLC-MS.

Glycopeptide	Glycoform	Theoretical M_r	Theoretical *m/z*
N_{413}	Pep	1475.7443	1476.7515 (+1), 738.8794 (+2)
	2Ant/0SiA	3098.3259	1550.1702 (+2), 1033.7826 (+3)
	2Ant/1SiA	3389.4213	1695.7179 (+2), 1130.8144 (+3), 848.3626 (+4)
	2Ant/2SiA	3680.5167	1841.2656 (+2), 1227.8432 (+3), 921.1365 (+4)
	2Ant/2SiAFuc	3826.5746	1276.5321 (+3), 957.6509 (+4)
	3Ant/2SiA	4045.6489	1349.5569 (+3), 1012.4195 (+4)
	3Ant/3SiA	4336.7443	1446.5887 (+3), 1085.1934 (+4)
N_{611}	Pep	2514.1169	1258.0657 (+2), 839.0462 (+3)
	2Ant/0SiA	4136.6985	1035.1819 (+4)
	2Ant/1SiA	4427.7939	1476.9386 (+3), 1107.9557 (+4), 886.5661 (+5)
	2Ant/2SiA	4718.8893	1573.9704 (+3), 1180.7296 (+4), 944.7851 (+5)
	2Ant/2SiAFuc	4864.9472	1622.6563 (+3), 1217.2441 (+4), 973.9967 (+5)
	3Ant/2SiA	5084.0215	1695.6811 (+3), 1272.0127 (+4)
	3Ant/3SiA	5375.1169	1792.7129 (+3), 1344.7865 (+4), 1076.0307 (+5)

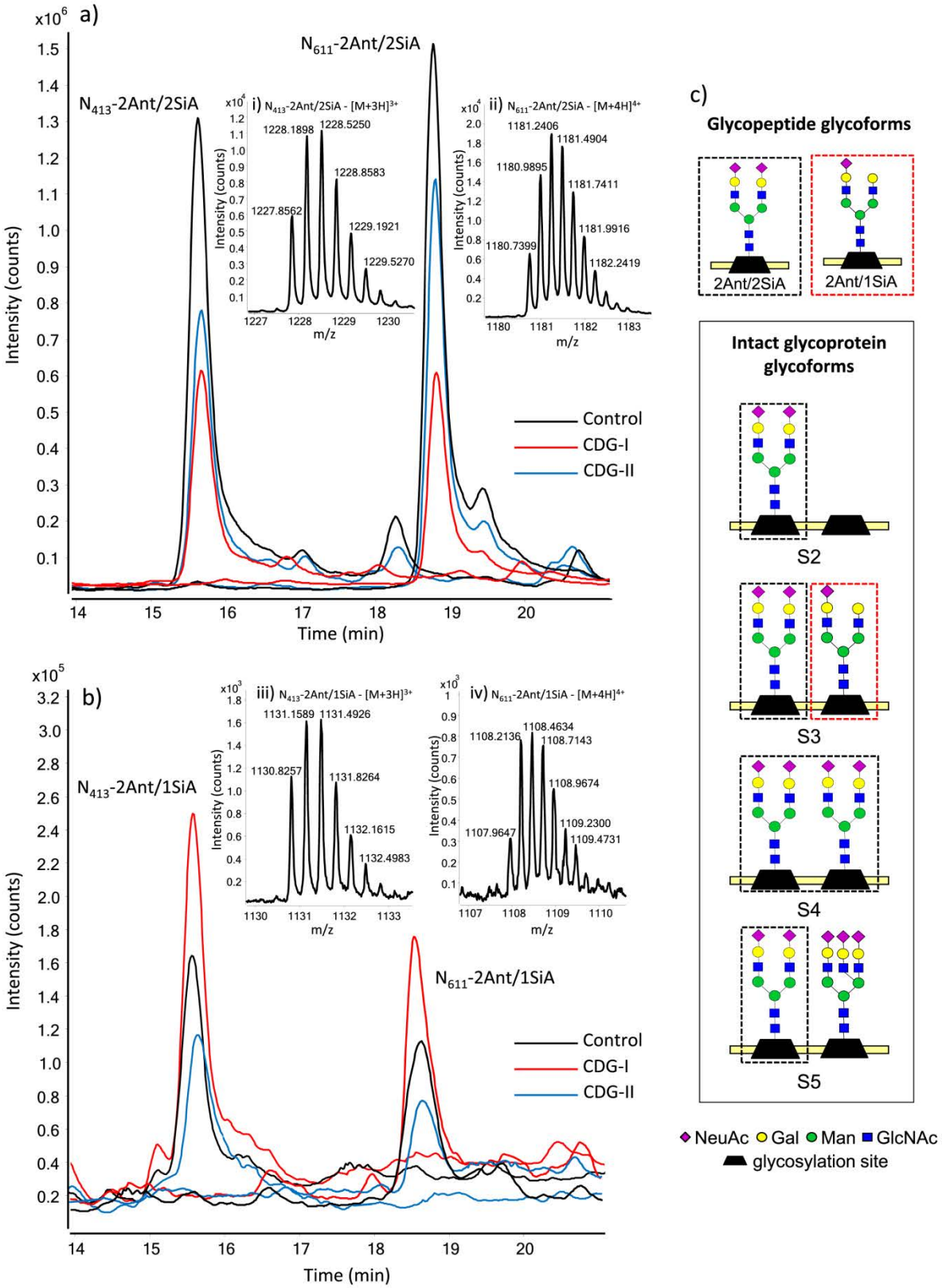


Fig. 1. a) Extracted ion chromatograms (EICs) for the N_{413}- and N_{611}-2Ant/2SiA glycoforms for control, CDG-I and CDG-II samples; b) Extracted ion chromatograms (EICs) for the N_{413}- and N_{611}-2Ant/1SiA glycoforms for control, CDG-I and CDG-II samples; i–iv) Mass spectra of the glycoforms in a) and b) for a control sample; c) Schematic representation of the intact Tf glycoforms from which the 2Ant/2SiA and 2Ant/1SiA glycopeptide glycoforms are derived.

2Ant/2SiA (Fig. 1a) and 2Ant/1SiA (Fig. 1b) glycoforms of N_{413} and N_{611} Tf glycopeptides for a control, a CDG-I and a CDG-II sample. As an example, the mass spectra of these glycoforms are also shown for the control sample (Fig. 1, i–iv). As can be observed, separation was excellent between both glycopeptides, but glycoforms of a certain glycopeptide could not be resolved by chromatography (compare Fig. 1a and b). However, as the different glycoforms had a different molecular mass (M_r) (Table 1), they were "resolved" by mass spectrometry. Differences between the abundance of the glycoforms shown in Fig. 1 for the control and CDG samples were evident, but statistical data analysis was necessary to draw some conclusions considering all the detected glycopeptide glycoforms in control and pathological samples.

3.1. Statistical data analysis

Table 1 shows the theoretical M_r of the 14 glycopeptide glycoforms detected by CapLC-MS (mass error < 5 ppm) and the observed ion charges to obtain the EICs. Glycoform peak areas were measured from the EICs and were normalised as explained in Section 2.5, before comparing the abundance of the different glycoforms between healthy controls and CDG patients, in order to minimise differences on total protein content, sample matrix and ionisation efficiency. Fig. 2a and b show bar graphs representing the mean A_{norm} of the detected glycoforms for both glycopeptides in controls, CDG-I and CDG-II samples. As can be observed, in general, the abundance of the different glycopeptide glycoforms

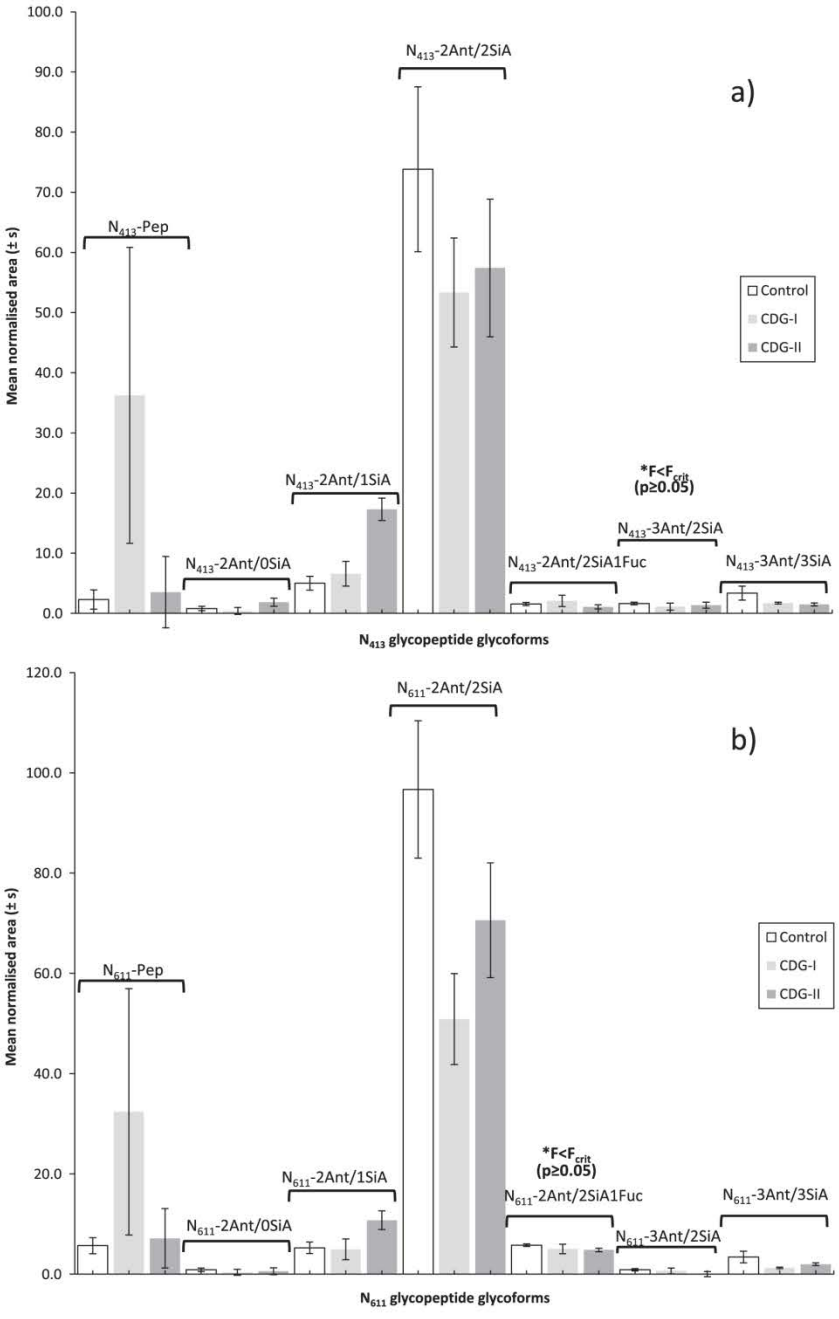


Fig. 2. Bar graphs of the mean of the normalised area (A_{norm}) and standard deviation of each detected glycoform of the a) N_{413} glycopeptide and the b) N_{611} glycopeptide for controls and CDG samples. A_{norm} of glycoforms marked with an asterisk were similar in both cases (ANOVA, $P > 0.05$: N_{413}-3Ant/2SiA and N_{611}-2Ant/2SiA1Fuc).

was in agreement with the abundance of the original intact glycoforms in controls and patients. As an example, note the small amount of N_{413}- and N_{611}-Pep in controls and CDG-II patients compared to CDG-I patients, where intact S2 and S0 glycoforms were much more abundant, or the high abundance of N_{413}- and N_{611}-2Ant/2SiA in the three groups of samples, because they could come from four different intact glycoforms (the most abundant S4, and three less abundant S5, S3 and S2, see Fig. 1c). In these two cases, it is also worth mentioning that the standard deviation of peak areas was especially high (see the error bars on the mean A_{norm} columns). This was probably due to the heterogeneity of the analysed samples (different individuals, mutation, etc.). It is well-known that the abundance of S2 and S0 intact glycoforms in CDG-I patients, and hence of N_{413}- and N_{611}-Pep, largely depends on the mutation causing CDG-I [8] (the five analysed samples were from patients with a different mutation). Moreover, the variation of N_{413}- and N_{611}-2Ant/2SiA abundances in the three groups of samples would be linked to the fact that they were from different individuals, and each individual has a natural variation in the abundance of the different glycoforms.

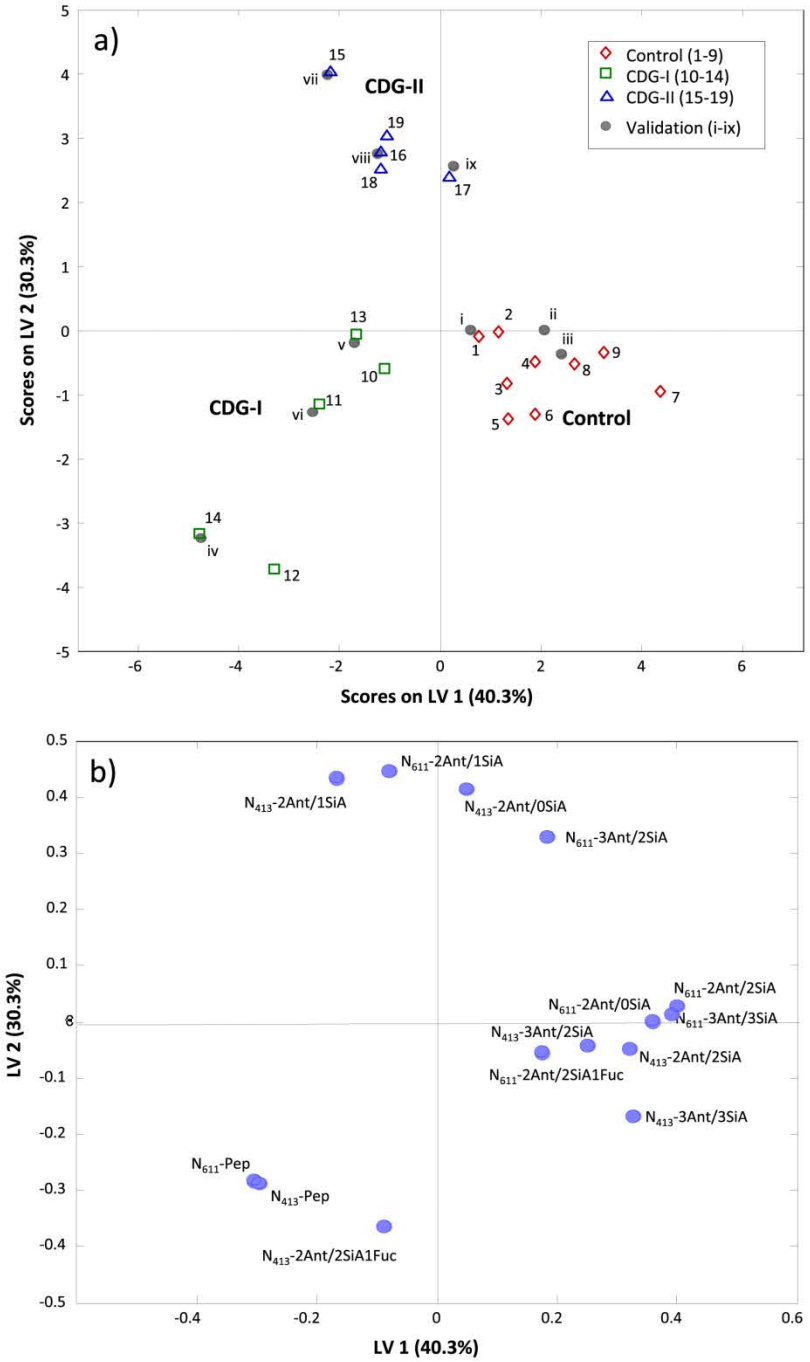


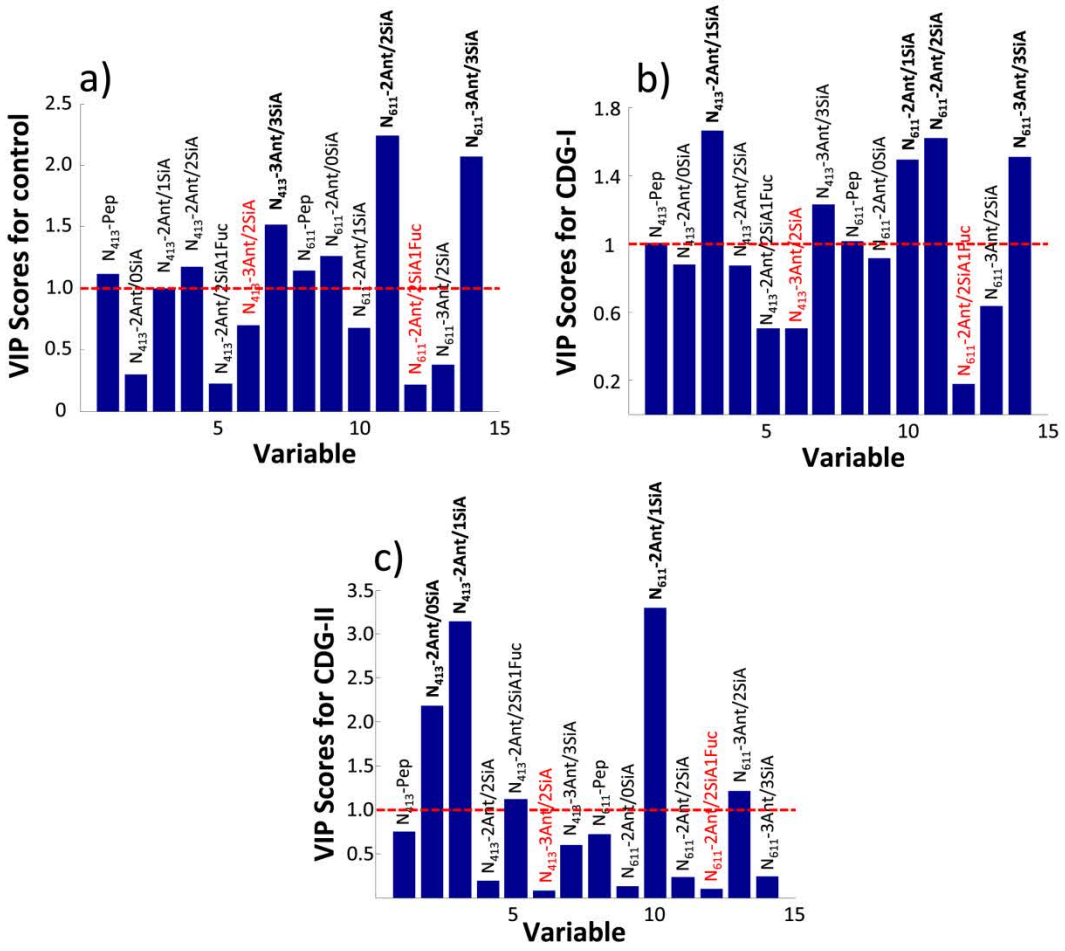
Fig. 3. a) Scores plot and b) loading plot of the PLS-DA model applied to the relative abundance of Tf glycopeptide glycoforms obtained after integration of the EICs of the different samples. Calibration set: nine healthy controls (1–9), five CDG-I (PGM1-CDG (10); DPM1-CDG (11); RFT1-CDG (12); DPAGT1-CDG (13); PMM2-CDG (14)) and five CDG-II (ATP6V0A2-CDG (15); $CDGII_x$ (16); $CDGII_x$(sepsis) (17) and two samples of COG8-CDG (18–19)). Validation set: three healthy controls (i–iii), three CDG-I (PMM2-CDG (iv), DPAGT1-CDG (v) and DPM1-CDG (vi)) and three CDG-II (ATP6V0A2-CDG (vii), CDGIIx (viii) and CDGIIx (sepsis) (ix)).

Despite the differences found in standard deviation in some cases, we found interesting to perform single factor ANOVA to compare the abundances of each glycoform in the three groups of samples. ANOVA revealed significant differences in the abundances of 12 glycoforms when the controls were compared with the patients [49]. As shown in Fig. 2, only the expression of 2 glycoforms was similar in all cases (see glycoforms marked with an asterisk, $P > 0.05$). The information provided by the ANOVA test was somewhat useful, because we established that for 12 glycoforms at least one of the mean abundances was different. As performing Student's *t*-tests to compare each pair of means was labour intensive and of limited usefulness, to further investigate if Tf glycopeptide glycoform fingerprints allowed differentiation between controls, CDG-I and CDG-II samples, it was preferred to use multivariate data analysis tools to take into account multicompound relationships and interactions.

3.2. Multivariate data analysis

First, we explored the data with PCA for the unsupervised identification of trends and clustering of the data, as well as outliers [40,50,51]. The scores plot for the first two principal components (a total of 70.7% of variance explained by the sum of PC 1 (41.4%) and PC 2) is provided as supplementary Fig. S1a. Despite the heterogeneity of the analysed samples three groups of samples were clearly separated, which demonstrated that the small variation due to the different individuals or the mutation (the subtypes within CDG-I or CDG-II) are barely significant compared to the differences between controls, CDG-I and CDG-II samples. The first principal component helped to differentiate between controls and CDG-I patients while the second separated these classes from the CDG-II patients. The loadings plot (supplementary Fig. S1b) showed interesting information about the glycoforms and the groups. However, PCA did not reveal their importance for differentiation between the three groups of samples. Therefore, we investigated a supervised multivariate data analysis method such as PLS-DA [41] to build a refined classification model with improved class separation that would also facilitate data interpretation and identification of potential and differential glycopeptide glycoform biomarkers of CDG-I and CDG-II.

The PLS-DA model was built considering the three classes of samples observed by PCA. As can be observed in the scores plot of Fig. 3a, a PLS-DA model with two latent variables (LV) allowed a perfect discrimination between the three groups of samples (70.6% of X-variance and 80.3% of Y-variance explained). Sensitivity, specificity and classification error in the calibration and (leave-one-out) cross-validation were excellent (see the figures of merit in supplementary Fig. S2). The loadings plot (Fig. 3b) was very similar to the one obtained before by PCA (supplementary Fig. S1b) and showed the contribution of the different variables to the different LV. As can be seen, in general, the glycoforms with the highest content of sialic acid, which were in the positive side of the x-axis (N_{413}-3Ant/3SiA, -3Ant/2SiA, -2Ant/2SiA, N_{611}-3Ant/3SiA, -2Ant/2SiA and -2Ant/2SiA1Fuc) were increased in controls and indeed, they had lower concentration in CDG-I and CDG-II patients. In contrast, the glycoforms with lower content of sialic acid and peptide forms were, in general, related to the CDG patients, in agreement with a higher abundance of carbohydrate deficient transferrin (S2, S1 and S0) or truncated glycopeptide glycoforms of S6, S5, S4 and S3. Specifically, the glycoforms N_{413}-2Ant/1SiA, -2Ant/0SiA, N_{611}-2Ant/1SiA and -3Ant/2SiA were elevated in the CDG-II patients, whereas, N_{413}-Pep and N_{611}-Pep seemed to be clearly characteristic of CDG-I patients when compared to CDG-II and control samples. It is well known that the lack of complete N-linked glycan and the increase of intact S2 and S0 glycoforms is typical in CDG-I patients, and may result in the increase of N_{413} and N_{611} peptide forms [8,31,52]. The presence of both peptide forms suggested that the abnormal glycosylation initiation affected equally both glycosylation sites. In contrast, sialic deficient glycopeptide glycoforms (N_{413}-2Ant/1SiA, -2Ant/0SiA, N_{611}-2Ant/1SiA and -3Ant/2SiA) were, as described by several authors, more characteristic in CDG-II [8,31,52]. We can add that the presence of a significant content of N_{413}-2Ant/0SiA glycoform in CDG-II and absence of N_{611}-2Ant/0SiA (it was in the controls region of the loading plot, Fig. 3b) denoted that the N_{413} glycosylation site was desialylated to a larger extent. This is probably due to the protein three-dimensional structure, which may affect the accessibility of the enzymes catalysing the alteration, seemingly resulting in each glycosylation site being altered in a different manner [53]. In addition to this qualitative information, the VIP scores [44–46] allowed now to quantify the influence of the different glycoforms on separation between control, CDG-I and CDG-II samples. The bar plots of Fig. 4 shows the VIP scores of the different glycoforms when considering separation of controls from CDG-I and CDG-II (4-a), CDG-I from controls and CDG-II (4-b) and CDG-II from controls and CDG-I (4-c). As can be observed, considering the three graphics, only two out of the fourteen selected glycoforms had all VIP values lower than 1 (N_{413}-3Ant/2SiA and N_{611}-2Ant/2SiA1Fuc, marked in red colour). Both glycoforms were already detected to be irrelevant to distinguish between control and CDG patients with the ANOVA. In contrast, the rest of glycoforms were more significant, especially the six glycoforms with VIP values higher than 1.5. These six glycoforms were enough to build a PLS-DA model with similar figures of merit than the model with 14 glycoforms (data not shown). As can be observed in Fig. 4a, N_{413}-3Ant/3SiA, N_{611}-2Ant/2SiA and N_{611}-3Ant/3SiA glycoforms had $VIP > 1.5$ and were the most significant to differentiate controls from CDG-I and CDG-II patients. Similarly, the glycoforms N_{413}-2Ant/1SiA, N_{611}-2Ant/2SiA, N_{611}-3Ant/3SiA and N_{611}-2Ant/1SiA were the most discriminating for CDG-I from controls and CDG-II (Fig. 4b). N_{413}- and N_{611}-Pep glycoforms were also relevant, but lower than expected ($1 < VIP < 1.5$, Fig. 4b), probably, as we indicated before, because the group of CDG-I samples was quite heterogeneous. Finally, N_{413}-2Ant/1SiA, N_{413}-2Ant/0SiA and N_{611}-2Ant/1SiA glycoforms were the most important for differentiation between CDG-II from controls and CDG-I (Fig. 4c). The table of Fig. 4d summarises the relationships between the twelve glycoforms with $VIP > 1$ for the different groups (the six glycoforms with $VIP > 1.5$ are highlighted in bold letters). As can be observed, in general, the glycoforms from both glycosylation sites were decisive for discrimination between the three groups. Furthermore, the significance and relationships between the different glycoforms agreed and reinforced the explanations given from the loadings plot. A new set of nine samples (3 controls (i–iii), 3 CDG-I (iv–vi) and 3 CDG-II (vi–ix)) were analysed and data processed following the same procedure in order to validate the PLS-DA model. Classification error was extremely low and class prediction was excellent (see the representation of the validation set in the scores plot of Fig. 3a, symbols are labelled from i–ix) suggesting that the PLS-DA model was also accurate regarding the validation set and reinforcing the excellent performance of the model for classification and the interpretation of the most discriminant glycoforms. Several authors have shown before that Tf glycosylation in CDGs can be characterised analysing the intact glycoprotein, glycopeptides or glycans by MS [31,53–55]. In this paper, we described a novel and simple multivariate data analysis method to evaluate systematically the complex information provided by Tf glycopeptide CapLC-MS fingerprints and classify controls, CDG-I and CDG-II samples. Moreover, this bottom-up approach complements the typical procedures currently applied for diagnosis of CDGs, which are based on comparison of the glycoform pattern observed for intact Tf by IEF or anion exchange



d) Tf glycopeptide glycoforms with VIP>1.0 in the PLS-DA model

CONTROL	CDG-I	CDG-II
N_{413}-Pep	N_{413}-Pep	
		N_{413}-2Ant/0SiA
N_{413}-2Ant/1SiA	**N_{413}-2Ant/1SiA**	**N_{413}-2Ant/1SiA**
N_{413}-2Ant/2SiA		
	N_{413}-2Ant/2SiA1Fuc	N_{413}-2Ant/2SiA1Fuc
N_{413}-3Ant/3SiA	N_{413}-3Ant/3SiA	
N_{611}-Pep	N_{611}-Pep	
N_{611}-2Ant/0SiA		
	N_{611}-2Ant/1SiA	**N_{611}-2Ant/1SiA**
N_{611}-2Ant/2SiA	**N_{611}-2Ant/2SiA**	
		N_{611}-3Ant/2SiA
N_{611}-3Ant/3SiA	**N_{611}-3Ant/3SiA**	

* In bold letters glycoforms with VIP>1.5

Fig. 4. VIP scores of the different glycoforms when considering the separation of a) controls from CDG-I and CDG-II, b) CDG-I from controls and CDG-II and c) CDG-II from controls and CDG-I; d) Table with the relationships between the twelve glycoforms with VIP > 1 for the different groups (the six glycoforms with VIP > 1.5 are highlighted in bold letters).

chromatography, that lack the unequivocal identification by MS [7,19,20].

4. Concluding remarks

We have demonstrated that Tf glycopeptide profiling by CapLC-MS combined with multivariate data analysis is an excellent approach to identify differences between healthy individuals, CDG-I and CDG-II patients, as well as for getting further insight on abnormal glycosylation. We have shown that only six glycopeptide glycoforms (N_{413}-3Ant/3SiA, -2Ant/1SiA, -2Ant/0SiA, N_{611}-3Ant/3SiA, -2Ant/2SiA and -2Ant/1SiA) are enough for an accurate and rapid classification, hence for diagnosis of CDG-I and CDG-II (probably even using an ordinary triple-quadrupole or quadrupole-ion trap mass spectrometer). In the future we will further validate this approach with larger sets of samples, including more genetic variants and samples from patients with secondary causes of underglycosylation (e.g. galactosemia or fructose intolerance). Furthermore, similar complementary studies will be performed using Tf glycan profiling by CapLC-MS to further the understanding of the structural changes involved in Tf abnormal glycosylation. The presented strategy shows a great potential to be applied for investigation of other diseases, based on bottom-up analysis of other glycoproteins or post-translationally modified protein biomarkers.

Acknowledgements

Part of this study was supported by the Spanish Ministry of Economy and Competitiveness (CTQ2011-27130 and CTQ2014-56777-R). Albert Barroso thanks the Spanish Ministry of Education, Culture and Sport for a FPU fellowship (FPU13/01596). The authors acknowledge Balagué Center S.A., Institute of Clinical Biochemistry (Hospital Clínic, University of Barcelona) and Centre for the Diagnosis of Molecular Diseases (Autonomous University of Madrid) for kindly providing the serum samples from healthy individuals and CDG patients.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.talanta.2016.07.055.

References

[1] B. Cylwik, M. Naklicki, L. Chrostek, E. Gruszewska, Congenital disorders of glycosylation. Part I. Defects of protein N-glycosylation, Acta Biochim. Pol. 60 (2013) 151–161.
[2] T. Hennet, J. Cabalzar, Congenital disorders of glycosylation: a concise chart of glycocalyx dysfunction, Trends Biochem. Sci. 40 (2015) 377–384.
[3] A.P. Corfield, M. Berry, Glycan variation and evolution in the eukaryotes, Trends Biochem. Sci. 40 (2015) 351–359.
[4] M. Theodore, E. Morava, Congenital disorders of glycosylation: sweet news, Curr. Opin. Pediatr. 23 (2011) 581–587.
[5] M. Guillard, Y. Wada, H. Hansikova, I. Yuasa, K. Vesela, N. Ondruskova, M. Kadoya, A. Janssen, L.P. Van den Heuvel, E. Morava, J. Zeman, R.A. Wevers, D.J. Lefeber, Transferrin mutations at the glycosylation site complicate diagnosis of congenital disorders of glycosylation type I, J. Inherit. Metab. Dis. 34 (2011) 901–906.
[6] B. Cylwik, K. Lipartowska, L. Chrostek, E. Gruszewska, Congenital disorders of glycosylation. Part II. Defects of protein o-glycosylation, Acta Biochim. Pol. 60 (2013) 361–368.
[7] J. Jaeken, Congenital disorders of glycosylation, Ann. N. Y. Acad. Sci. 1214 (2010) 190–198.
[8] K. Scott, T. Gadomski, T. Kozicz, E. Morava, Congenital disorders of glycosylation: new defects and still counting, J. Inherit. Metab. Dis. 37 (2014) 609–617.
[9] D. Krasnewich, Human glycosylation disorders, Cancer Biomark. 14 (2014) 3–16.
[10] L.A. Wolfe, D. Krasnewich, Congenital disorders of glycosylation and intellectual disability, Dev. Disabil. Res. Rev. 17 (2013) 211–225.
[11] B. Xia, W. Zhang, X. Li, R. Jiang, T. Harper, R. Liu, R.D. Cummings, M. He, Serum N-glycan and O-glycan analysis by mass spectrometry for diagnosis of congenital disorders of glycosylation, Anal. Biochem. 442 (2013) 178–185.
[12] T. Hennet, Diseases of glycosylation beyond classical congenital disorders of glycosylation, Biochim. Biophys. Acta 1820 (2012) 1306–1317.
[13] H.H. Freeze, Understanding human glycosylation disorders: biochemistry leads the charge, J. Biol. Chem. 288 (2013) 6936–6945.
[14] M. Mohamed, M. Guillard, S.B. Wortmann, S. Cirak, E. Marklova, H. Michelakakis, E. Korsch, M. Adamowicz, B. Koletzko, F.J. van Spronsen, K. E. Niezen-Koning, G. Matthijs, T. Gardeitchik, D. Kouwenberg, B.C. Lim, R. Zeevaert, R.A. Wevers, D.J. Lefeber, E. Morava, Clinical and diagnostic approach in unsolved CDG patients with a type 2 transferrin pattern, Biochim. Biophys. Acta 1812 (2011) 691–698.
[15] R.T. MacGillivray, E. Mendez, S.K. Sinha, M.R. Sutton, J. Lineback-Zins, K. Brew, The complete amino acid sequence of human serum transferrin, Proc. Natl. Acad. Sci. USA 79 (1982) 2504–2508.
[16] J.R. Delanghe, M.L. De Buyzere, Carbohydrate deficient transferrin and forensic medicine, Clin. Chim. Acta 406 (2009) 1–7.
[17] M. Ridinger, P. Köhl, E. Gäbele, N. Wodarz, G. Schmitz, P. Kiefer, C. Hellerbrand, Analysis of carbohydrate deficient transferrin serum levels during abstinence, Exp. Mol. Pathol. 92 (2012) 50–53.
[18] W. Oberrauch, A.C. Bergman, A. Helander, HPLC and mass spectrometric characterization of a candidate reference material for the alcohol biomarker carbohydrate-deficient transferrin (CDT), Clin. Chim. Acta 395 (2008) 142–145.
[19] R. Barone, A. Fiumara, J. Jaeken, Congenital disorders of glycosylation with emphasis on cerebellar involvement, Semin. Neurol. 34 (2014) 357–366.
[20] L. Sturiale, R. Barone, D. Garozzo, The impact of mass spectrometry in the diagnosis of congenital disorders of glycosylation, J. Inherit. Metab. Dis. 34 (2011) 891–899.
[21] M. Guillard, E. Morava, F.L. Van Delft, R. Hague, C. Körner, M. Adamowicz, R. A. Wevers, D.J. Lefeber, Plasma N-glycan profiling by mass spectrometry for congenital disorders of glycosylation type II, Clin. Chem. 57 (2011) 593–602.
[22] M.L. Hennrich, A.C. Gavin, Quantitative mass spectrometry of posttranslational modifications : keys to confidence, Sci. Signal 8 (2015) 1–6.
[23] M. Wuhrer, Glycomics using mass spectrometry, Glycoconj. J. 30 (2013) 11–22.
[24] S.B. Levery, 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.
[25] Y. Zhang, J. Jiao, P. Yang, H. Lu, Mass spectrometry-based N-glycoproteomics for cancer biomarker discovery, Clin. Proteom. 11 (2014) 18.
[26] 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.
[27] Z.R. Gregorich, Y. Ge, Top-down proteomics in health and disease: challenges and opportunities, Proteomics 14 (2014) 1195–1210.
[28] P.D. Compton, L. Zamdborg, P.M. Thomas, N.L. Kelleher, On the scalability and requirements of whole protein mass spectrometry, Anal. Chem. 83 (2011) 6868–6874.
[29] L.C. Gillet, A. Leitner, R. Aebersold, Mass Spectrometry applied to bottom-up proteomics: entering the high-throughput era for hypothesis testing, Annu. Rev. Anal. Chem. (2016), http://dx.doi.org/10.1146/annurev-anchem-071015-041535.
[30] K. Artemenko, J. Mi, J. Bergquist, Mass-spectrometry-based characterization of oxidations in proteins, Free Radic. Res. 49 (2015) 477–493.
[31] W.E. Heywood, P. Mills, S. Grunewald, V. Worthington, J. Jaeken, G. Carreno, H. Lemonde, P.T. Clayton, K. Mills, A new method for the rapid diagnosis of protein N-linked congenital disorders of glycosylation, J. Proteom. Res. 12 (2013) 3471–3479.
[32] J. Gloerich, R.A. Wevers, J.A. Smeitink, B.G. Van Engelen, L.P. Van Den Heuvel, Proteomics approaches to study genetic and metabolic disorders, J. Proteom. Res. 6 (2007) 506–512.
[33] 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.
[34] 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.
[35] N. Hashii, A. Harazono, 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.
[36] L. Pont, F. Benavente, J. Barbosa, V. Sanz-Nebot, Analysis of transthyretin in human serum by capillary zone electrophoresis electrospray ionization time-of-flight mass spectrometry. Application to familial amyloidotic polyneuropathy TYPE I (FAP-I), Electrophoresis 36 (2015) 1265–1273.
[37] J. Jaeken, T. Hennet, G. Matthijs, H.H. Freeze, CDG nomenclature: time for a change! Biochim. Biophys. Acta 1792 (2009) 825–826.

[38] R.F. Ritchie, G.E. Palomaki, L.M. Neveux, O. Navolotskaia, T.B. Ledue, W.Y. Craig, Reference distributions for the negative acute-phase serum proteins, albumin, transferrin and transthyretin: a practical, simple and clinically relevant approach in a large cohort, J. Clin. Lab. Anal. 13 (1999) 273–279.
[39] Y.N. Ordonez, R.F. Anton, W.C. Davis, Quantification of total serum transferrin and transferrin sialoforms in human serum; an alternative method for the determination of carbohydrate-deficient transferrin in clinical samples, Anal. Methods 6 (2014) 3967–3974.
[40] I.T. Joliffe, B.J. Morgan, Principal component analysis and exploratory factor analysis, Stat. Methods Med. Res. 1 (1992) 69–95.
[41] M. Barker, W. Rayens, Partial least squares for discrimination, J. Chemom. 17 (2003) 166–173.
[42] E. Ortiz-Villanueva, J. Jaumot, F. Benavente, B. Piña, V. Sanz-Nebot, R. Tauler, Combination of CE-MS and advanced chemometric methods for high-throughput metabolic profiling, Electrophoresis (2015), http://dx.doi.org/10.1002/elps.201500027.
[43] L. Pont, F. Benavente, J. Jaumot, R. Tauler, J. Alberch, S. Ginés, J. Barbosa, V. Sanz-Nebot, Metabolic profiling for the identification of Huntington biomarkers by on-line solid-phase extraction capillary electrophoresis mass spectrometry combined with advanced data analysis tools, Electrophoresis 37 (2016) 795–808.
[44] S. Wold, M. Sjöström, L. Eriksson, PLS-regression: a basic tool of chemometrics, Chemom. Intell. Lab. Syst. 58 (2001) 109–130.
[45] S. Wold, PLS-partial least squares projections to latent structures, in: A. Johansson, M. Cochi (Eds.), PLS-Partial Least Squares Proj. to Latent Struct, ESCOM Science Publishers, Leiden, 1993.
[46] S. Wold, PLS for multivariate linear modeling, in: H. van de Waterbeemd (Ed.), QSAR Chemom. Methods Mol. Des. Methods Princ. Med. Chem., Verlag Chemie, Weinheim, 1995, pp. 195–218.
[47] G.C. Cawley, N.L.C. Talbot, Efficient leave-one-out cross-validation of kernel fisher discriminant classifiers, Pattern Recognit. 36 (2003) 2585–2592.
[48] M. Stone, Cross-Validatory choice and assessment of statistical predictions, J. R. Stat. Soc. 36 (1974) 111–147.
[49] J.N. Miller, J.C. Miller, Statistics and Chemometrics for Analytical Chemistry, in: Stat. Chemom. Anal. Chem., 6th ed., 2010: pp. 56–58.
[50] I.T. Jolliffe, J. Cadima, Principal component analysis : a review and recent developments, Philos. Trans. R. Soc. A Math. Phys. Eng. Sci. (2016), http://dx.doi.org/10.1098/rsta.2015.0202.
[51] K.L. Sainani, Introduction to principal components analysis, PM R 6 (2014) 275–278.
[52] D.J. Lefeber, E. Morava, J. Jaeken, How to find and diagnose a CDG due to defective N-glycosylation, J. Inherit. Metab. Dis. 34 (2011) 849–852.
[53] Y. Wada, Mass spectrometry of transferrin and apolipoprotein C-III for diagnosis and screening of congenital disorder of glycosylation, Glycoconj. J. 33 (2016) 1–11.
[54] M. Van Scherpenzeel, G. Steenbergen, E. Morava, R.A. Wevers, D.J. Lefeber, High-resolution mass spectrometry glycoprofiling of intact transferrin for diagnosis and subtype identification in the congenital disorders of glycosylation, Transl. Res. 166 (2015) 639–649.
[55] V. Sanz-Nebot, E. Balaguer, F. Benavente, C. Neusüss, J. Barbosa, Characterization of transferrin glycoforms in human serum by CE-UV and CE-ESI-MS, Electrophoresis 28 (2007) 1949–1957.

Supplementary Figures

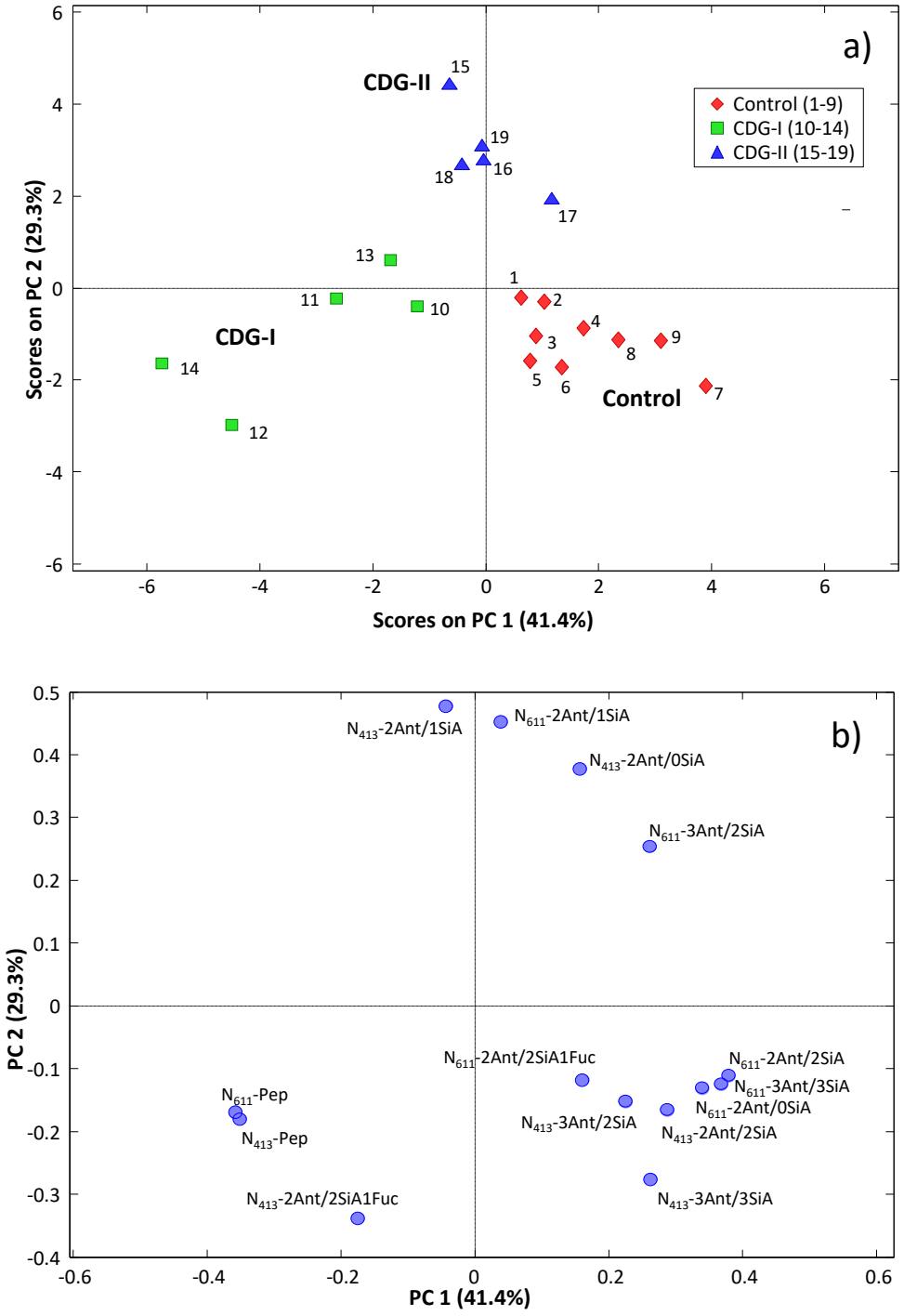


Figure S1: a) Scores plot and b) loadings plot of the PCA model applied to the relative abundance of Tf glycopeptide glycoforms obtained after integration of the EICs of the different samples. Calibration set: nine healthy controls (1-9), five CDG-I (PGM1-CDG (10); DPM1-CDG (11); RFT1-CDG (12); DPAGT1-CDG (13); PMM2-CDG (14)) and five CDG-II (ATP6V0A2-CDG (15); $CDGII_x$ (16); $CDGII_x$(sepsis) (17) and two samples of COG8-CDG (18-19)). Validation set: three healthy controls (i-iii), three CDG-I (PMM2-CDG (iv), DPAGT1-CDG (v) and DPM1-CDG (vi)) and three CDG-II (ATP6V0A2-CDG (vii), CDGIIx (viii) and CDGIIx (sepsis) (ix)).

Figure S2: Table with several quality parameters regarding the calibration and (leave-one-out) cross-validation for the PLS-DA model.

Modeled Class:	Control	CDG-I	CDG-II
Sensitivity (Cal):	1.000	1.000	1.000
Specificity (Cal):	1.000	1.000	1.000
Sensitivity (CV):	1.000	1.000	1.000
Sensitivity (CV):	1.000	1.000	1.000
Class. Err (Cal):	0	0	0
Class. Err (CV):	0	0	0
RMSEC:	0.209	0.224	0.172
RMSECV:	0.277	0.277	0.234
Bias:	-5.551×10^{-17}	0	-5.551×10^{-17}
CV Bias:	7.703×10^{-3}	8.519×10^{-3}	-1.622×10^{-2}
R^2 Cal:	0.823	0.739	0.845
R^2 CV:	0.705	0.623	0.725

Journal of Proteomics 134 (2016) 127–137

Contents lists available at ScienceDirect

Journal of Proteomics

journal homepage: www.elsevier.com/locate/jprot

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

A. Rosal-Vela [a,1], A. Barroso [b,1], E. Giménez [b], S. García-Rodríguez [a], V. Longobardo [c], J. Postigo [d], M. Iglesias [d], A. Lario [c], J. Merino [d], R. Merino [e], M. Zubiaur [a], V. Sanz-Nebot [b], J. Sancho [a,*]

[a] Instituto de Parasitología y Biomedicina "López-Neyra" (IPBLN), CSIC, Armilla, Granada, Spain
[b] Departament de Química Analítica, Universitat de Barcelona, Barcelona, Spain
[c] Unidad de Proteómica, IPBLN, CSIC, Armilla, Granada, Spain
[d] Facultad de Medicina, Universidad de Cantabria, Santander, Spain
[e] Instituto de Biomedicina y Biotecnología de Cantabria, CSIC, Santander, Spain

ARTICLE INFO

Article history:
Received 14 August 2015
Received in revised form 11 November 2015
Accepted 26 November 2015
Available online 27 November 2015

Keywords:
Arthritis
Inflammation
Protein species
Transferrin
Glycosylation
CD38
Citrullination

ABSTRACT

Collagen type II-induced arthritis (CIA) is an inflammatory and autoimmune disease. Spleen protein extracts were subjected to 2D-DiGE and MS-MALDI-TOF/TOF analysis to identify protein species that differ in abundance in CD38-KO versus B6 WT mice either with arthritis or with inflammation. Using multivariate analyses, in Col-II-immunized mice, 23 distinct spleen protein species were able to discriminate between WT and CD38-KO mice. Among them, several citrullinated proteins and multiple serotransferrin (Tf) species were identified. In contrast, in CFA/IFA-treated mice, the distinct protein profile, which discriminates between CD38-KO and WT mice, was unrelated with Tf, but not with citrullination. Unexpectedly, non-immunized CD38-KO mice showed a distinct proteome profile as compared with that in non-immunized WT mice, and again multiple protein species were identified as Tf. By using a μLC-TOF-MS method to separate and detect Tf glycopeptide glycoforms, increases in fucosylation and glycan branching was observed in sera from mice CIA$^+$ versus non-immunized, and between WT and CD38-KO with arthritis. Data on 2-DE Tf spots indicated differences in glycosylation related with NeuGc content. Thus, Tf changed significantly in its glycosylation pattern in arthritic mice. The MS data have been deposited to the ProteomeXchange Consortium with the dataset identifiers: PXD002644, PXD002643, PXD003183, and PXD003163.

Significance: 2-DE followed by μLC-TOF-MS could be implemented to identify Tf glycoforms linked to specific protein species, and correlate a particular Tf species to a function. To gain insight into the relationship between transferrin glycoforms and its biological function it is particularly interesting to study putative differences in the glycosylation pattern of Tf in specific tissues associated with the disease (i.e.: joints), or in specific compartments such as exosomes/microvesicles, which are highly enriched in Tf receptors.

1. Introduction

Mouse models of human diseases are quite interesting for functional proteomic studies, since a relative large number of cells with the same phenotype can be analyzed at the same time, without the constraints inherent to human studies. Furthermore, samples can be collected under standard operating protocols while the onset and progression of the disease can be closely monitored in well-defined experimental settings [1,2].

C57BL/6 wild-type mice (WT), which are H-2^b, develop arthritis in response to chicken collagen II (Col-II) that is milder than that in DBA/1 (H-2^q) mice immunized with bovine Col-II, but more chronic, with more pronounced and more persistent T-cell responses [3]. Moreover, B6 mice develop a sustained T-cell response to chicken collagen as well as to autologous (mouse) collagen [3]. CIA in B6 mice resembles human rheumatoid arthritis (RA) in terms of its disease course, histological findings, and in its response to commonly used anti-arthritic drugs [3,4]. Therefore, CIA in B6 mice provides a valuable model for assessing the role of specific genes involved in the induction and/or maintenance of arthritis and for evaluating the efficacy of novel drugs, particularly those targeted at T cells.

Abbreviations: CIA, collagen-type-II-induced arthritis; WT, wild-type; CFA, complete Freund's adjuvant; IFA, incomplete Freund's adjuvant; RA, rheumatoid arthritis.

* Corresponding author at: Departamento de Biología Celular e Inmunología, IPBLN-CSIC, Avenida del Conocimiento s/n, Parque Tecnológico de Ciencias de la Salud, 18016 Armilla, Granada, Spain.
E-mail address: granada@ipb.csic.es (J. Sancho).

[1] A. Rosal-Vela and A. Barroso equally contributed to this work.

http://dx.doi.org/10.1016/j.jprot.2015.11.023
1874-3919/

Mice deficient in CD38 (*CD38*$^{-/-}$ or CD38-KO) develop an attenuated CIA as compared with that in WT mice. Immunized CD38-KO mice produce high levels of circulating IgG_1 and low of IgG_{2a} anti-Col-II antibodies in association with reduced percentages of T-helper 1 (Th1) cells in the draining lymph nodes. Altogether, the results show that CD38 participates in the pathogenesis of CIA controlling the number of iNKT cells and promoting Th1 inflammatory responses [5].

In this study, the goal has been to identify proteins from the spleen that change in their abundance in CD38-KO versus WT mice with positive clinical and radiological scores of having developed arthritis (CIA$^+$), which may reflect their distinct response to an antigen-challenge that induces the development of an autoimmune disease. To this end, we have analyzed the spleen from CD38-KO and WT mice immunized with chicken Col-II/CFA (the CIA model described above), with CFA/IFA alone (as a model of chronic inflammation), or without any treatment. Samples were then compared by 2D-DiGE. This approach allows comparison of changes in protein abundance across multiple samples simultaneously with minimal gel-to-gel variation [6].This approach also allows the identification of multiple protein species [7], or proteoforms [8] of a given protein in a single analysis, and therefore, to focus the interest in fully characterize just the protein species that differ in abundance. In this study, by using a μLC-TOF-MS method to separate and detect Tf glycopeptide glycoforms from serum samples, differences in the pattern of serum Tf glycosylation between CD38-KO and WT mice were found in both non-immunized and Col-II-immunized mice, and between mice with arthritis or not within each mouse-type. This is in concordance with the alteration of the glycosylation pattern expected in inflammation, as it has been widely described an increase in fucosylation and glycan branching (number of antennae) in inflammatory diseases. Applying this technology to the Tf spots isolated by 2-DE preliminary data indicate that the most acidic glycoforms, those with more glycolylneuraminic acid (NeuGc) (e.g. 2Ant/3NeuGc), were more abundant in the acidic spots, while the less acidic glycoforms, those with less NeuGc (e.g. 2Ant/1NeuGc), were more abundant in the basic spots. In comparison to the traditional bottom-up LC–MS approach, the applied 2-DE followed by the tryptic digestion of selected spots and μLC-TOF-MS analysis may allow the quantification at the protein species level [9], and reveal protein species-specific regulation during arthritis and inflammation.

2. Material and methods

2.1. Mice

WT mice were purchased from Harlan Ibérica (Barcelona, Spain). Mice deficient in CD38 (CD38-KO) were backcrossed onto the B6 background for more than 12 generations, as described previously [10]. All studies with live animals were approved by the IPBLN and Universidad de Cantabria Institutional Laboratory Animal Care and Use Committees.

2.2. Induction and assessment of arthritis

For the induction of CIA, 8–12 weeks-old male mice were immunized as previously described [3,11].

2.3. Protein extraction from spleen preparations

Proteins were extracted from spleen by using the MicroRotofor Lysis Kit (for mammalian tissues and cells) (Bio-Rad, Ref #163–2141), following the manufacturer's instructions, which includes the use of mini-grinders for effective disruption of cells and tissues. The excess of salts and other contaminants were removed using the Bio-Rad's ReadyPrep 2-D cleanup kit. Samples were then resuspended in a DiGE-compatible buffer (7 M urea, 2 M thiourea, 4% CHAPS, 20 mM Tris, pH 8.5), quantified using the RC DC assay, and kept at −20 °C until further use.

2.4. Design of DiGE experiments

Unless otherwise indicated in each DiGE experiment conducted, four biological replicates of each condition were compared, comprising protein samples derived from four CD38-KO mice and four WT mice as previously described [5]. Each gel contained one CD38-KO protein sample, one WT protein sample, and a pooled internal standard. Three separate DiGE experiments were conducted, the first DiGE (DiGE #1), comprised two groups of mice, CD38-KO and WT, and two conditions per group (mice with arthritis: CIA$^+$, and mice with no arthritis: CIA$^-$), in mice that were all previously immunized with Col-II in CFA and IFA (arthritis model), as described previously [5]. DiGE #2 comprised two groups of mice, CD38-KO and WT, all treated with CFA/IFA (chronic inflammation model) as described [5]. DiGE #3 comprised two groups of non-immunized mice, CD38-KO and WT, which served as normal controls. Spleen samples were from the same mice in which their respective serum proteome profiles were previously analyzed using a similar approach [5].

Table 1
Protein species from spleen extracts that have shown differences in abundance by DiGE and that were identified by MS/MS.

Protein name	Accession number[a]	Spot number[b]
Aconitate hydratase, mitochondrial	Q99KI0	501
Serotransferrin	Q921I1	533, 537, 538, 539
Far upstream element-binding protein 1	Q91WJ8	554
Prelamin-A/C isoform A precursor[c]	gi|162287370	572
Fibroblast growth factor 22[c]	gi|161760667	572
Fibroblast growth factor 22[c]	gi|12963627	572
Stress-induced-phosphoprotein 1	Q60864	614
Fibrinogen alpha chain	E9PV24	633
Heterogeneous nuclear ribonucleoprotein L	Q8R081	638
Fibrinogen beta chain	Q8K0E8	692
Catalase	P24270	697
Coronin-1	gi|4895037	732
Vimentin	P20152	738
Protein disulfide-isomerase A6 precursor	gi|58037267	770
Alpha-enolase	P17182	778
Beta-enolase	P21550	778
Gamma-enolase	P17183	778
Actin, cytoplasmic 1	P60710	898
Beta-actin-like protein 2	Q8BFZ3	898
F-actin-capping protein subunit alpha-1	P47753	898
Voltage-dependent anion selective channel protein 2	Q60930	906
Tropomyosin alpha-1 chain	P58771	972
Proteasome subunit alpha type-1	Q9R1P4	981
Carbonic anhydrase 2	P00920	982
Carbonic anhydrase 1	P13634	982
Mitochondrial peptide methionine sulfoxide reductase[c]	Q9D6Y7	983
Proteasome subunit alpha type-6	Q9QUM9	1001
Pyridoxine-5′-phosphate oxidase	Q91XF0	1001
Proteasome subunit beta type-10	O35955	1049
Growth factor receptor-bound protein 2	Q60631	1049
Flavin reductase	Q923D2	1063
ATP synthase subunit d, mitochondrial	Q9DCX2	1097
Ferritin light chain 1	P29391	1103
Ferritin light chain 2	P49945	1103
Low molecular weight phosphotyrosine protein phosphatase[c]	Q9D358	1104
Peptidyl-prolyl cis-trans isomerase A	P17742	1136
Actin-related protein 2/3 complex subunit 5-like protein	Q9D898	1147
E3 ubiquitin-protein ligase RNF181	Q9CY62	1150
Nucleoside diphosphate kinase A	P15532	1171
Nucleoside diphosphate kinase B	Q01768	1171
Ubiquitin-conjugating enzyme E2 N	P61089	1173
Protein S100-A9	P31725	1184
N-acyl-aromatic-L-amino acid amidohydrolase	Q91XE4	1302
Protein S100-A8	P27005	1313

[a] UniprotKB/Swiss-Prot or NCBI Accession Number.
[b] Spots are named as indicated on the 2DE gel shown in Fig. 1.
[c] These proteins were identified by PMF.

2.5. DiGE labeling and two-dimensional gel electrophoresis

Samples were aliquoted at 45 μg, and the pooled internal standard was made with 23 μg of each of the sixteen test samples combined. The proteins were labeled with 400 pmol (in 1 μL of anhydrous DMF) of CyDye per 50 μg of protein as per the manufacturer's instructions (GE Healthcare). After labeling, the appropriate samples were combined for each gel. Each combined sample (~50 μl) was made up to 200 μL with Readyprep Rehydration/Sample buffer (8 M urea, 2% CHAPS, 50 mM dithiothreitol (DTT), 0.2% (w/v) Bio-Lyte® 3/10 ampholytes, and Bromophenol Blue (trace)).

2-DE was carried out using the Protean IEF cell and Criterion electrophoresis cell systems (Bio-Rad. Hercules, CA, USA) as previously described [12], with the following modifications: 1) First-dimension IPG strips (Bio-Rad: 11 cm, linear pH 3–10 gradient); 2) Active in-gel rehydration at 50 V, 12 h at 20 °C; 3) The IPG strips were focused in a one-step procedure, at 8000 V for a total of 35,000 Vh at 20 °C with a current limit of 50 μA/strip.

After electrophoresis, one of the gels was pre-scanned using the Typhoon 9400 variable mode imager at each of the appropriate CyDye excitation wavelengths (Cy3 (532 nm), Cy5 (633 nm), Cy2 (488 nm)), in order to determine the appropriate laser intensity for each CyDye. Thereafter, each of the analytical gels was scanned at this optimum laser intensity at a resolution of 100 μm. Gels were then fixed and stained with SYPRO Ruby (Bio-Rad) and re-scanned using the 488 nm laser. Scanned images were analyzed using the DeCyder7.0 software (GE Healthcare) using the Differential In-gel Analysis (DIA) module to detect and normalize the protein spots. Standard was used to normalize gels by calculating the standardized abundance of each spot, i.e., the ratio of either Cy3 or Cy5 signal to that of Cy2.

2.6. Protein identification by MALDI-TOF/TOF MS/MS

In-gel digestion of proteins has been described previously [13]. A set of protein spots were identified by MS/MS using a 4800 MALDI-TOF/TOF Analyzer (AB SCIEX) in automatic mode with the settings described previously [5]. Protein identification was assigned by peptide mass fingerprinting and confirmed by MS/MS analysis of at least three peptides in each sample. Mascot 2.0 search engine (Matrix Science) was used for protein identification running on GPS software (Applied Biosystems) against the SwissProt *Mus musculus* database (uniprot_sprot_26042011.fasta). The search setting allowed one missed cleavage with the selected trypsin enzyme, a MS/MS fragment tolerance of 0.2 Da and a precursor mass tolerance of 100 ppm.

Other spots were identified by MS/MS using a MALDI TOF/TOF UltrafleXtreme (Bruker) in manual mode as previously described [5]. Fragment selection criteria were a minimum S/N ratio of 15, a maximum number of peaks set at 200. For each precursor selected for MS/MS analysis, fragment mass values in the range from 13 Da to 4 Da below precursor mass were used to peptide identification.

Protein identification was assigned by peptide mass fingerprinting and confirmed by MS/MS analysis of 5 peptides. Mascot Server 2.4 (Matrix Science) and ProteinScape 3.1 (Bruker) were used for protein identification against the SwissProt *Mus musculus* database (SwissProt_2015_06.fasta and NCBInr_20150409.fasta). The search setting allowed two missed cleavage with the selected trypsin enzyme, fixed modification was cysteine carbamidomethylation and variable modifications was methionine oxidation, a MS/MS fragment tolerance of 0.5 Da and a precursor mass tolerance of 50 ppm, unless otherwise indicated.

The MS spectra of the identified proteins were further examined in order to detect the presence of citrullinated proteins. Protein citrullination (deimination) is the enzymatic conversion of peptidyl-arginine residues to peptidyl-citrulline, mediated by the family of calcium-dependent peptidylarginine deiminases (PADs) [14]. The search setting for this PTM with MASCOT was performed as in the previous paragraph, including as variable modification the deamination of arginine, with the following considerations [15]: a) for one citrullinated arginine, the peptide theoretical mass increase is 0.98 Da and the modified peptide, losing one amino group, becomes more acidic; b) citrullinated arginine residues are not likely to be cleaved by trypsin, so that a minimum number of one missed cleavage must be specified; c) a peptide that includes a C-terminal citrullinated arginine must be rejected; d) citrullinated peptides generate an unusual isotopic mass cluster as compared with that of unmodified peptides.

2.7. Purification of serotransferrin from mouse serum samples by immunoaffinity chromatography (IAC)

In order to isolate Tf from the rest of serum proteins, an immunoaffinity purification was carried out using a cyanogen-bromide Sepharose column where a polyclonal antibody against human transferrin was bond as explained in a previous work [16]. The immunoaffinity

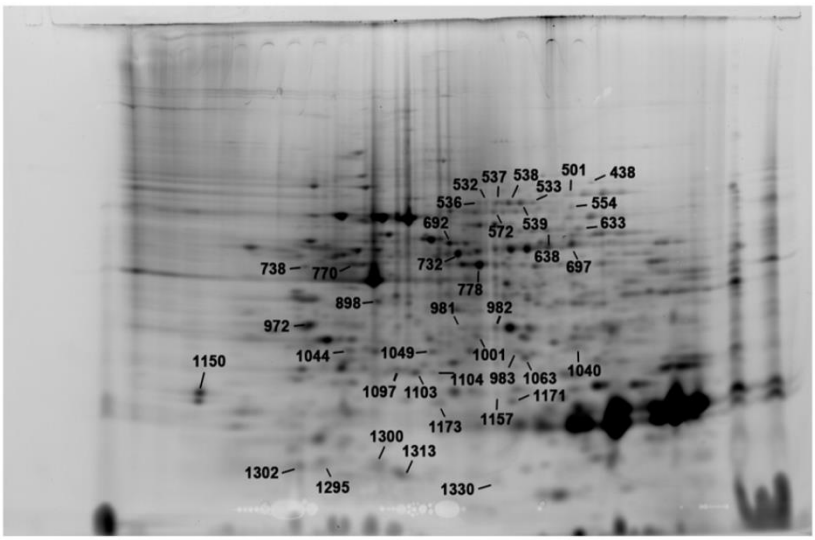


Fig. 1. 2D-gel image of the protein spots from a mouse spleen extract. The image corresponds to the Sypro Ruby staining of proteins. The total amount of protein loaded was of 150 μg. The numbers indicate the position of the proteins identified by MS/MS, which showed differences in protein abundance by DiGE analysis. The information about the identified proteins is in Tables 1, 2, and 3 of Ref [18].

procedure consisted in: first, a conditioning step with two washes of 10 mM Tris–HCl. Second, the serum sample was diluted 1/16 in 10 mM Tris–HCl (pH 7.6) in order to improve antigen–antibody interaction, and passed through the column ten times. After washing with 10 mM Tris–HCl and 0.5 M NaCl (pH 7.6) to remove the non-specifically retained fraction (washing fraction), retained Tf was eluted with 100 mM glycine-HCl (pH 2.5). Eluted Tf was immediately neutralized with 0.5 M Tris. Afterwards, glycine-HCl buffer was exchanged for 50 mM NH_4HCO_3 buffer (pH 7.9) by ultracentrifugation using Microcon YM-10 (MW cut-off 10 kDa, Millipore, Bedford, MA, USA). Finally, Tf was digested as explained in Section 2.8.

2.8. Serotransferrin in-solution digestion

Tf previously purified by immunoaffinity chromatography as described above, or a commercially available mouse apoTf (Sigma, reference: T0523), which was used as standard, was reduced, alkylated and immediately subjected to enzymatic digestion in the presence of RapiGest® as explained in a previous work [16]. 0.5 M DTT in 50 mM NH_4HCO_3 buffer was added to an aliquot of isolated Tf with 0.1% (w/v) RapiGest®. The mixture was incubated in a TS-100 Thermo-Shaker digester (Bio-San, Riga, Latvia) at 56 °C for 30 min and then alkylated with 50 mM IAA for 30 min at room temperature in the dark. Excess of reagents were removed by ultracentrifugation with Microcon YM-10 washing 3 times with NH_4HCO_3 buffer. The final residue was recovered from the upper reservoir by being centrifuged upside down in a new vial, and reconstituted to the initial volume with NH_4HCO_3 buffer with 0.1% RapiGest®. Trypsin was added and the mixture was carefully vortexed and incubated overnight at 37 °C. After the overnight incubation the surfactant was hydrolyzed to avoid MS incompatibilities. Formic acid (HFor) was added to the digest to a final concentration of 5% (v/v) and the mixture was incubated in the digester at 37 °C for 30 min. Then, the

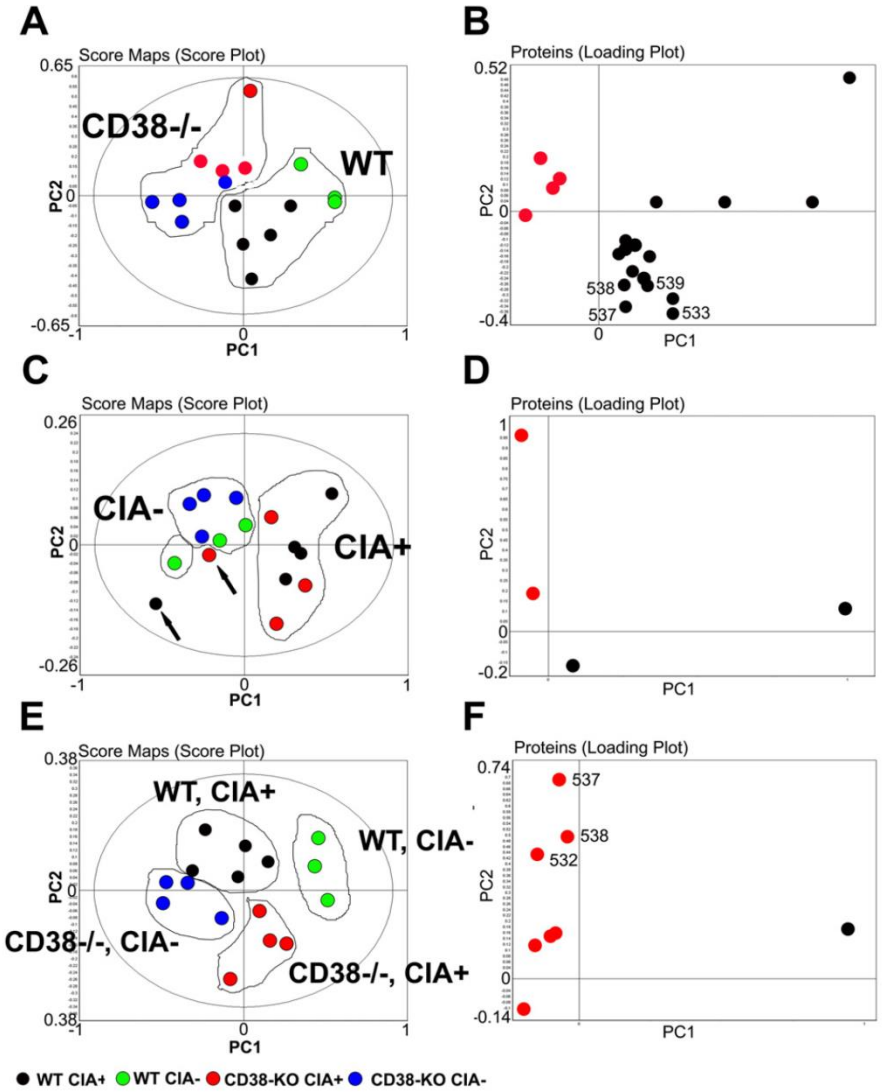


Fig. 2. Principal component analysis (PCA) on the protein species from the spleen with a distinct profile in Col-II-immunized mice. A, Score plot of a spot map of the 16 mice color coded according to the legend, projected onto the first two principal components. PCA included the 20 proteins that showed statistical significance by the 2-ANOVA-mouse test. The areas highlight the segregation of WT and CD38-KO mice according to mouse-type. Most mice clustered in the same area according to their mouse status. B, Loading plot of the 20 proteins that resulted statistically significant in the 2-ANOVA-mouse test, projected onto the first two principal components. The red and black circles highlight the proteins that reached statistical significance in CD38-KO and WT mice, respectively. The plot shows that the proteins clustered in distinct areas according to its differential protein abundance in CD38-KO and WT mice. C, and D, PCA on spleen protein species with a distinct profile in CIA^+ versus CIA^- mice, according to the 2-ANOVA-arthritis test. The areas highlight clustering of mice in two distinct areas, according to their arthritis condition. E, The score plot of the distribution of the 16 mice according to the variance of proteins that were statistically significant by 2-ANOVA-interaction. The areas highlight the clustering of mice in 4 different subareas depending upon whether they showed signs of arthritis or not, and they were WT or CD38-KO. F, Loading plot of the 8 proteins that resulted statistically significant by 2-ANOVA-interaction.

solution was centrifuged for 10 min at 12,000 rpm to separate the RapiGest® residues. A residual fraction of RapiGest® still remained in the solution but it was completely compatible with LC–MS analysis as indicated by the manufacturer. Finally, the supernatant was carefully collected and stored at −20 °C until its analysis.

2.9. Serotransferrin in-gel digestion

Each gel spot was subjected to the following in-gel digestion procedure: the spots were washed with 50% ACN, 100% ACN and 50 mM NH_4HCO_3. Afterwards, the spots were dried with 100% ACN and the solvent was then evaporated in a vacuum concentrator (SpeedVacTM, Thermo Fisher Scientific, Waltham, MA). The reduction of Tf was performed adding 10 mM DTT in 50 mM NH_4HCO_3 with 0.1% RapiGest® to the spots and incubating in the digester at 56 °C for 45 min. Then, the alkylation of Tf was carried out adding 55 mM of IAA in 50 mM NH_4HCO_3 with 0.1% RapiGest® to the spots and leaving in the dark for 30 min. The spots were again washed with 50% ACN, 100% ACN and 50 mM NH_4HCO_3 and dried with 100% ACN. Then, the solvent was again evaporated. 0.1% of RapiGest® was added to the spots and the mixture was incubated at 37 °C for 10 min. The solvent was once more evaporated in the vacuum concentrator and trypsin was finally added to the spots until reswollen. The mixture was incubated in ice for 45 min. Then, 50 mM NH_4HCO_3 was added to completely cover the spots and the mixture was incubated at 37 °C overnight. After the incubation, the supernatant was collected. This supernatant contains

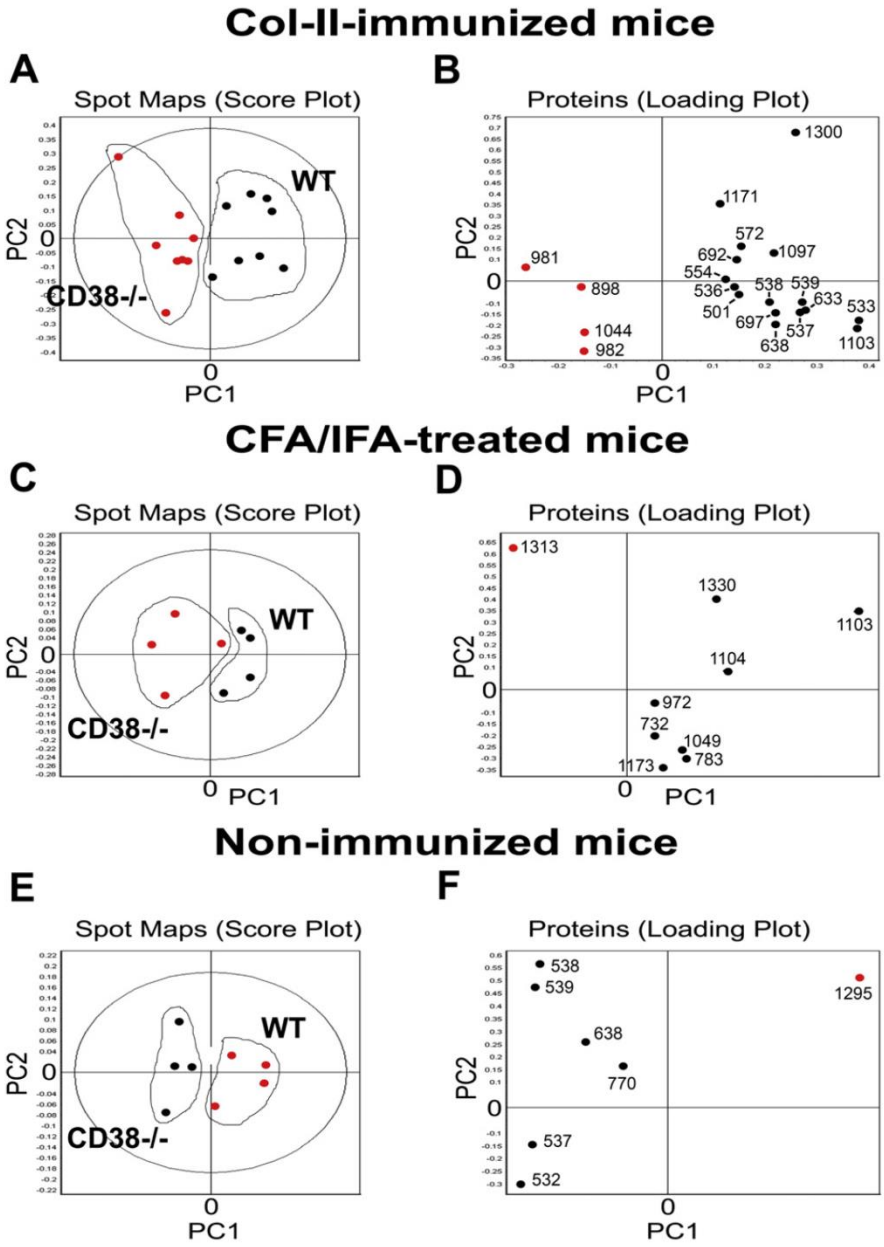


Fig. 3. A and B, PCA analysis of the same 16 mice as in Fig. 2A, and B, which included the 20 protein species that were statistically significant by average ratio (*t*-test). The areas highlight the segregation of mice and proteins according to mouse type. C and D, PCA on spleen protein species with a distinct profile in CFA/IFA-treated CD38-KO versus WT mice, according to their differences in protein abundance (log of average ratio, *t*-test ≤0.05). Red circles, CD38-KO; black circles, WT. The areas highlight clustering of CD38-KO and WT mice in two distinct areas. E and F, PCA on spleen protein species with a distinct profile in non-immunized CD38-KO versus WT mice, according to their differences in protein abundance (log of average ratio, *t*-test ≤0.05). Red circles, CD38-KO; black circles, WT. The areas highlight clustering of CD38-KO and WT mice in two distinct areas.

RapiGest® which had to be hydrolyzed to avoid MS interferences. To ensure the complete removal of RapiGest® the spots were washed twice with water and the supernatants were collected and mixed with the first supernatant. The mixture was then hydrolyzed as explained in Section 2.8. Afterwards, the spots were washed twice with 100% ACN and both supernatants were collected and mixed with the solution resulting from the surfactant hydrolyzation. Then, the solvent was evaporated and, finally, the sample was reconstituted in 50 mM NH_4HCO_3 and stored at −20 °C until its analysis.

2.10. μLC-TOF-MS

The μLC-TOF-MS experiments were performed in the 1200 series capillary liquid chromatography system coupled to a 6220 oa-TOF LC/MS mass spectrometer with an orthogonal G1385-44,300 interface (Agilent Technologies). Data were collected in positive mode between m/z 100 and 3200 working in the highest resolution mode (4 GHz). For the separation of the tryptic digests, a Zorbax 300SB-C18 column (3.5 μm particle diameter, 300 Å pore diameter, 150 × 0.3 mm L_T x id, Agilent Technologies) was used. Chromatographic conditions were the same as those used in a previous work [16]. Sample injection was performed with an autosampler refrigerated at 4 °C and the injection volume was 1 μL when analyzing Tf isolated from serum samples and 5 μL when analyzing Tf in-gel digests.

2.11. Data access

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium [17] via the PRIDE partner repository with the following dataset identifiers: PXD002644 for proteins identified by 4800 MALDI-TOF/TOF Analyzer (AB SCIEX), PXD002643 for proteins identified by MALDI TOF/TOF UltrafleXtreme, Bruker, PXD003183 for the identified citrullinated proteins, and PXD003163 for the μLC-TOF-MS data. Supporting data are shown in ref. [18].

2.12. Statistical analyses

Protein relative abundance across all samples and statistical analyses were performed using the Biological Variation Analysis (BVA) module of the DeCyder software. Data were filtered by retaining spots that were present in all gels. Comparison of protein abundance in serum samples were calculated based on the relative change of sample respect to its in-gel internal standard. Comparison is done by Student's t-test and for calculations it was used the log standardized protein abundance. The degree of difference between protein groups was expressed as average ratio.

3. Results and discussion

3.1. A multi-protein species signature discriminates CIA^+ mice from CIA^- mice, and WT from CD38-KO within each condition

In WT and CD38-KO mice the incidence of CIA is about 70% of the Col-II-immunized mice, despite the fact that all of them show anti-Col-II autoantibody response. Therefore, it was of interest to study whether there were differences in spleen protein profiles between CIA^+ and CIA^- mice that all were immunized with Col-II/CFA and, therefore, subjected to same inflammatory challenge induced by CFA. On the other hand, the clinical course of arthritis in CD38-KO mice is milder than that in WT mice [11], which opens the question of whether putative differences in spleen protein profiles between these mice may reflect a distinct immune response of CD38-KO mice to Col-II/CFA immunization. The goal was to identify proteins which may be post-translationally modified, or altered in their spleen concentration levels in response to Col-II/CFA immunization. For the complete list of identified protein species that showed differences in abundance by DiGE, see Table 1, and for their positions on the 2D-gels, see Fig. 1. See also Tables 1, 2 and 3 in Ref [18].

In the first experiment (DiGE #1), 2 groups of samples were analyzed by 2D-DiGE, mouse type: CD38-KO versus WT mice, and two conditions in each group: arthritis (CIA^+) versus no arthritis (CIA^-). 2-way ANOVA-mouse test for the independent parameter mouse type

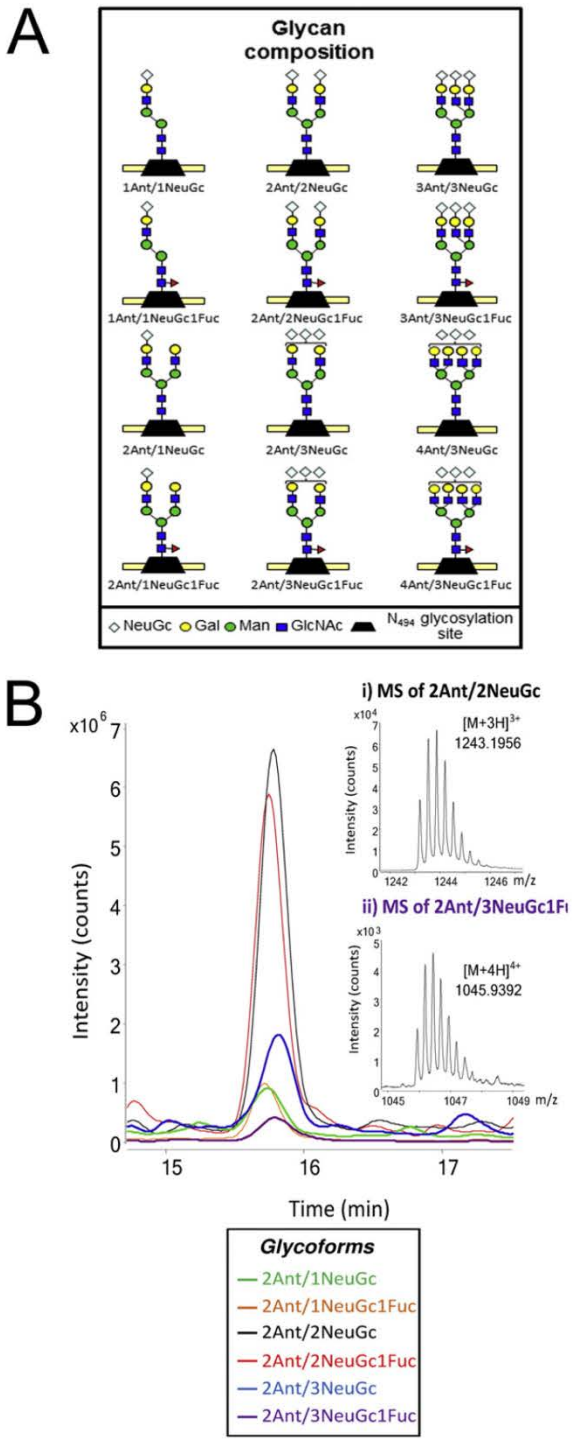


Fig. 4. A, Glycan composition for each detected glycopeptide of mouse Tf digested with trypsin. B, Extracted ion chromatograms (EICs) for the most abundant glycopeptide glycoforms of a Tf standard sample and mass spectrum of the 2Ant/2NeuGc glycoform (i) and the 2Ant/3NeuGc1Fuc glycoform (ii).

identified 20 protein spots that showed statistically significant differences ($P \leq 0.05$) (see Table 4 in Ref [18]), while the 2-ANOVA-arthritis test identified 4 protein spots that showed statistically significant differences (see Table 5 in Ref [18]). When the two parameters were considered together, 8 spots showed a 2-ANOVA-interaction $P < 0.05$ (see Table 6 in Ref [18]). Several of these spots were identified by MS/MS as serotransferrin (Tf), and appeared to be strongly associated with the WT mice group (spots #533, #537, #538, and #539), or with the interaction of arthritis and mouse conditions (spots #532, #537, and #539).

To further select the proteins that contributed better to discriminate between CIA^+ and CIA^- mice, a multivariate statistical approach was used: Principal Component Analysis (PCA). Fig. 2A shows a score plot of the distribution of the 16 mice, according to the variance of the 20 proteins that were statistically significant by 2-ANOVA-mouse. 6 out of 8 WT mice clustered in the lower left quadrant, indicating limited variance within this group, independently that they showed clinical signs of arthritis or not. Likewise, most CD38-KO mice clustered within or closer to the upper right quadrant. In addition, the Loading Plot of the 20 proteins that resulted statistically significant in the 2-ANOVA-mouse test shows that the proteins clustered in distinct areas according to its differential abundance in WT versus CD38-KO mice (Fig. 2B).

The score plot of the distribution of the 16 mice, according to the variance of the 4 proteins that were statistically significant by 2-ANOVA-arthritis showed that 7 out of the 9 mice with arthritis clustered together in the two right quadrants, whereas the other two mice (WT #13 and CD38-KO #16) clustered in the lower left quadrant, closer to the mice with no clinical signs of arthritis (Fig. 2C). In addition, the Loading Plot of the 4 proteins that resulted statistically significant in the 2-ANOVA-arthritis test showed that the protein species clustered in distinct areas according to their differential abundance in CIA^+ versus CIA^- mice (Fig. 2D).

The score plot of the distribution of the 16 mice according to the variance of the 8 proteins that were statistically significant by 2-ANOVA-interaction showed that the 16 mice clustered according to their condition (CIA^+ versus CIA^-), and within each condition they clustered according to mouse type (WT versus CD38-KO). Thus, mice clustered in 4 different subareas depending upon whether they showed signs of arthritis or not, and they were WT or CD38-KO (Fig. 2E, and F). These data suggested that the response to Col-II immunization was different between WT and CD38-KO mice, independently that they were going to develop arthritis or not.

3.2. Identification of spleen protein species distinct to CD38-KO mice in response to Col-II immunization

To better distinguish the response of CD38-KO to Col-II immunization or to CFA/IFA treatment PCAs were performed with those protein species which showed statistical differences in the protein abundance based in their average ratios (*t*-test ≤0.05). In Col-II-immunized mice, the segregation between WT and CD38-KO mice was even more patent than that using the 2-ANOVA-mouse test (Fig. 3A vs Fig. 2A). A close examination of the protein species that were shared by both type of analyses showed a general coincidence, with only three protein species difference (see Tables 4 and 7 in Ref [18]). Therefore, the data showed that these 23 protein species that differed in their abundance between WT and CD38-KO mice were useful to discriminate between them, which are consistent with a different immune-response to Col-II immunization [5,11]. To note is that 6 out of these 23 protein species (spots #533, #537, #538, #539, #697, and #1103) correspond to Tf, catalase, and ferritin, which are proteins related to iron transport and metabolism, processes that are altered in RA patients, and other autoimmune diseases [19,20].

CD38-KO mice have higher NAD^+ levels than WT mice and are protected against obesity and metabolic syndrome [21]. Moreover, CD38 regulates global protein acetylation through changes in NAD^+ levels and sirtuin activity [22,23]. In this sense, spots #501, #906, #983, and #1097 (aconitase, VDAC2, PMSR, and ATPase subunit d) correspond to enzymes that are involved in mitochondrial metabolism, and are substrates of the mitochondrial NAD^+-dependent deacetylase, SIRT3 [24]. Likewise, spot #697 that corresponds to Catalase is also a substrate of SIRT3, despite the fact that is a major component of peroxisomes [24]. These enzymes and at least another 9 identified proteins are potential substrates of another sirtuin, SIRT5, which is a weak deacetylase and a potent desuccinylase and malonylase [25]. SIRT3 and SIRT5 regulate diverse metabolic pathways [25–28]. It is likely that these and other sirtuins are hyperactive in CD38-KO mice by virtue of the increased levels of NAD^+, in particular under chronic inflammatory or stress conditions [22]. Therefore, it is feasible that the decreased abundance of these spots in CD38-KO relative to WT mice is due to massive sirtuin-mediated lysine deacetylation, and desuccinylation, and not to changes in protein concentration.

3.3. Protein species changes observed with inflammation or under steady-state conditions

The immunization process with Col-II involves the use of CFA, which contains *Mycobacterium tuberculosis* extracts and mineral oil that are strongly pro-inflammatory in normal mice [29,30]. Therefore, it was uncertain whether the observed differences in protein abundance between Col-II-immunized CD38-KO and WT mice were the consequence of the specific immune response to collagen, or rather they reflected differences in the inflammatory response to the adjuvant. To distinguish between these possibilities, 2D-DiGE analysis was performed in spleen extracts from 4 WT and 4 CD38-KO mice that were challenged with CFA at day 0, and with IFA at day 14, and then taken spleen samples at 6 weeks of the CFA injection (DiGE #2). None of the adjuvant-only animals ever developed signs of arthritis (data not shown). As shown in Table 8 of Ref [18], spot #1313, corresponding to S100-A8, showed increased abundance in CFA/IFA-treated CD38-KO mice as compared

Table 2
Normalized peak area and %RSD values of Tf glycopeptides glycoforms detected in a standard sample of 1500 μg mL^{-1}.

Glycoform	M_{theo}	M_{exp}	Error (ppm)	Charges	A_{norm} [a]	%RSD
1Ant/1NeuGc	2892.2138	2892.2443	10.5	+2, +3	5.0	7.4
1Ant/1NeuGc1Fuc	3038.2717	3038.3080	11.9	+2	1.6	7.7
2Ant/1NeuGc	3419.3988	3419.4394	11.9	+2, +3, +4, +5	12.5	5.8
2Ant/1NeuGc1Fuc	3565.4568	3565.4955	10.9	+2, +3, +4	6.6	6.3
2Ant/2NeuGc	3726.4892	3726.5352	12.3	+2, +3, +4, +5	88.2	2.6
2Ant/2NeuGc1Fuc	3872.5471	3872.5958	12.6	+2, +3, +4, +5	77.7	2.6
2Ant/3NeuGc	4033.5795	4033.6386	14.6	+2, +3, +4	28.7	6.6
2Ant/3NeuGc1Fuc	4179.6374	4179.6860	11.6	+2, +3, +4	30.8	5.6
3Ant/3NeuGc	4398.7117	4398.7553	9.9	+3, +4	6.9	4.2
3Ant/3NeuGc1Fuc	4544.7696	4544.8172	10.5	+3, +4	5.9	2.0
4Ant/3NeuGc	4763.8439	4763.7646	16.7	+3	4.3	9.2
4Ant/3NeuGc1Fuc	4909.9018	4909.9535	10.5	+3	2.6	6.8

[a] A_{norm}: normalized peak areas were calculated as: (Glycoform peak area/peptide 354–364 (CDEWSIISEGK) peak area) × 100.

with WT mice, while 8 additional spots showed decreased abundance in those mice. PCA (Fig. 3C, and D) of these proteins indicated that CD38-KO and WT mice could be segregated in response to CFA/IFA treatment, which suggested differences in the inflammatory response to CFA/IFA injection elicited by CD38-KO versus WT mice.

The spleen proteomic profiles of non-immunized CD38-KO and WT mice were also analyzed by 2D-DiGE (DiGE #3). 7 protein species showed statistically significant differences in protein abundance. 4 of them were identified as Tf, and all of them showed decreased abundance in CD38-KO versus WT mice (see Table 9 of Ref [18]). In addition, PCA showed that these 7 protein species were useful to discriminate non-immunized CD38-KO from WT mice (Fig. 3E and F).

3.4. Identification of citrullinated peptides linked to specific protein species in CIA$^+$, or CFA-treated mice

Some of the protein spots identified by MS in WT mice with arthritis or treated with CFA such as fibrinogen, vimentin, hnRNPL, alpha-enolase, and FuBP-1 are known RA autoantigens, whose antigenicities were highly dependent on citrullination, a PTM that occurs on arginine residues that are deiminated by specific PADs to form citrulline residues [15]. In CIA$^+$ mice the development of autoantibodies against citrullinated filaggrin and citrullinated fibrinogen has been shown [31]. Moreover, immunoblotting and MS analysis of CIA joint tissue lysates identified citrullinated ATP synthase subunit beta, actin, tropomyosin alpha-1 chain, and S100A9 [32]. In our study 13 out of 40 spots analyzed for citrullination scored positively for protein species with 1 or more citrullinated peptides (see Table 10 in Ref [18]), including Fibrinogen α chain, Tropomyosin alpha-3 chain, Coronin-1A, Actin, and F-actin-capping protein subunit alpha-1, which are structural proteins. Citrullination is known to alter the structural properties of keratin and MBP, and thus citrullination may play an important role in modulating the structural properties of such proteins [14]. Other citrullinated proteins such as Catalase, alpha-Enolase, Stress-induced phosphoprotein-1, and the two different Proteasome subunits might be involved in anti-inflammatory and anti-oxidative stress responses. Spot #537, one of the more acidic Tf species was also citrullinated. Regardless of what their targets are, antibodies against citrullinated proteins may contribute to the pathogenesis of the disease [31]. It would be interesting to determine whether the CIA$^+$ sera specifically recognize these citrullinated proteins on 2-DE protein maps.

3.5. Analysis of the glycopeptide glycoforms of serotransferrin standard by μLC-TOF-MS

As the results obtained by 2D-DiGE analysis of the protein extract from spleen demonstrated, several Tf protein species seemed to be among the altered proteins that could be able to discriminate between WT and CD38-KO both in non-immunized mice and in Col-II-immunized mice, we focused our study on the analysis of Tf. The protein species of Tf identified by 2D-DiGE differed in their p*I*s, as judged by their different mobility on 2D gels, which could be indicative of changes in the glycosylation pattern of this protein. Hence, a more exhaustive analysis of this protein was mandatory to further correlate changes in Tf glycosylation in CD38-KO and Collagen type II-induced arthritis (CIA$^+$) mice.

As it has been extensively proved in the last years, MS is the most reliable way to obtain structural information about protein glycosylation as it allows fast and high sensitivity profiling and accurate characterization of heterogeneous glycan structures [33–37].The analysis of Tf glycoforms by MS could be approached either by analyzing the intact glycoprotein or by analyzing the glycans or glycopeptides obtained by enzymatic digestion [38,39]. Glycopeptide analysis is usually the best alternative as it shows better sensitivity compared to intact protein analysis and provides information about glycosylation sites and their degree of occupation in contrast to the analysis of glycans. In the particular case of mouse Tf, the analysis of glycopeptides or glycans would seemingly provide the same information as mouse Tf has only one glycosylation point, thus, only one glycopeptide would be obtained and all the glycans released would only come from this glycosylation point. However, glycopeptide analysis is still preferred as characterization of glycans is more tedious due to the need of a derivatization step to enhance their ionization yield in MS.

In this work, a previous developed methodology for the analysis of human serotransferrin glycopeptide glycoforms by μLC-TOF-MS was used to characterize the glycosylation pattern of mouse Tf glycopeptide [16]. In order to establish the methodology, an in-solution digestion of

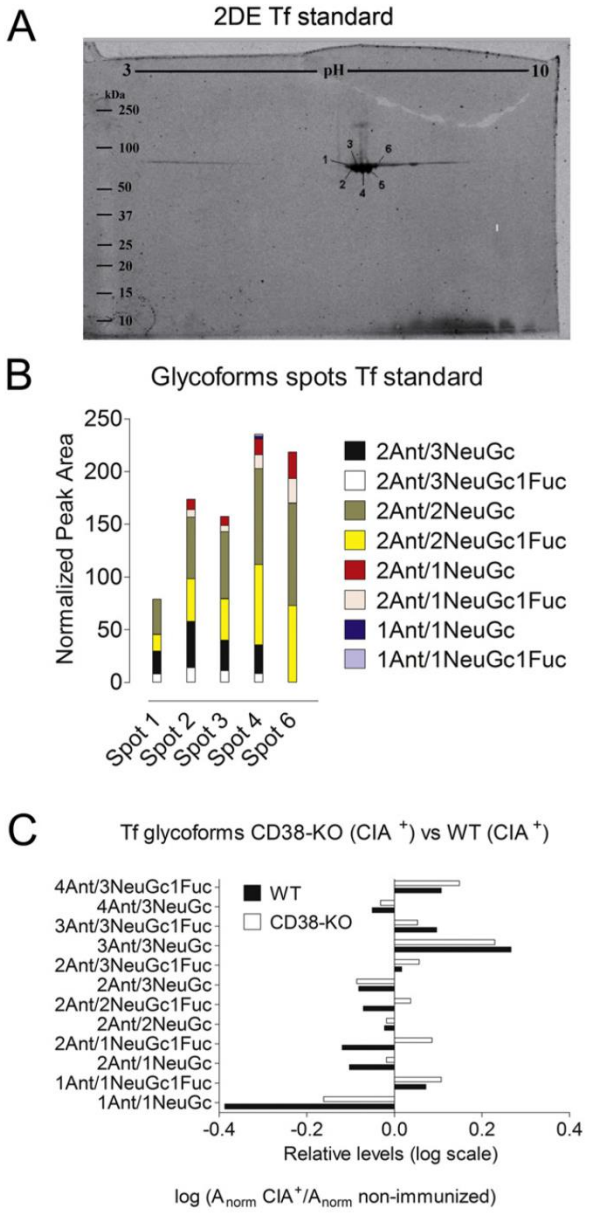


Fig. 5. A, 2D-gel image of a Tf standard sample. The images correspond to the Sypro Ruby staining of proteins. The total amount of protein loaded was of 10 μg of Tf. The number of the spots goes from one to six, being the spot number one the most acidic (lowest p*I*) and the spot number six the most basic (highest p*I*). B, Bar graph of the normalized peak areas of the main glycoforms for the Tf glycopeptide detected in the spots of a Tf standard subjected to a 2-DE separation and in-gel tryptic digestion. C, Bar graph of the logarithm in base ten of the quotient between the normalized peak areas of the detected glycoforms for the Tf glycopeptide in CIA$^+$ mice serum and in non-immunized mice serum ($\log \frac{\text{Anorm CIA+}}{\text{Anorm non-immunized}}$) for the WT mice (black bars) and the CD38-KO mice (white bars).

Tf standard at a concentration of 1500 μg mL^{-1} was carried out as explained in Section 2.8. The resulting glycopeptide and peptides were analyzed by μLC-TOF-MS. The amino acid coverage of the protein sequence was around 91% and all detected peptides are shown in Table 11 of Ref [18]. As reported in [16] for human Tf, the use of RapiGest® surfactant was needed to obtain quantitative digestions of mouse Tf, and thus, to correctly detect and characterize the distinct glycoforms of its glycopeptide. Fig. 4A shows all the *N*-glycopeptide glycoforms detected, and Table 2 shows their normalized area (A_{norm}; calculated as $\frac{\text{Glycopeptide glycoform peak area}}{\text{Reference peptide peak area}} \times 100$; reference peptide: 354–364, CDEWSIISEGK), and their relative standard deviation (%RSD) values. As it can be observed, several glycoforms were detected; being the biantennary structures the most abundant ones. As reported by Coddeville et al. [40], the terminal sialic acid present in all the detected glycoforms of mouse Tf was the NeuGc and no glycoforms containing an acetylneuraminic acid (NeuAc) were observed. Furthermore, Tf glycans seemed to be quite fucosylated, as glycopeptide glycoforms with one fucose unit were rather abundant too, although they were always less intense than the non-fucosylated counterpart, except for the glycoform 2Ant/3NeuGc1Fuc. As an example of peak shape, retention times and intensity of the observed glycoforms, Fig. 4B shows the extraction ion chromatograms (EIC) for each of the main detected glycoforms along with the mass spectrum of the 2Ant/2NeuGc glycoform (i) and 2Ant/3NeuGc1Fuc glycoform (ii).

3.6. Analysis of the Tf glycopeptide glycoforms in 2-DE spots from the Tf standard or spleen protein extracts

As previously stated, with the aim of analyzing the Tf glycopeptide glycoforms in protein extracts from spleen separated by 2D-gels, an in-gel digestion of Tf standard was carried out to establish the in-gel digestion methodology and later to evaluate whether the same glycoforms were observed in the standard and in the spleen protein extract. Following the methodology described in Sections 2.5 and 2.9, a 2D-gel was loaded with 10 μg of Tf standard and the six spots obtained after the electrophoretic separation (see Fig. 5A) were subjected to in-gel digestion. Tf glycopeptide and peptides obtained were finally analyzed by μLC-TOF-MS. Table 3 shows the main glycopeptide glycoforms detected along with their normalized area (A_{norm}) and %RSD values. Despite the increase in the injection volume from 1 μL to 5 μL, sensitivity was rather low, and tri- and tetra-antennary glycoforms and also some minor abundant biantennary glycoforms were not detected in any spot.

Fig. 5B depicts the bar graphs with the normalized area of each glycopeptide glycoform in each spot. If we consider the position of the different spots in the 2D-gel, the spot with the lowest p*I* (most acidic) is number 1 and the spot with the highest p*I* (most basic) is number 6. Supposing that distinct electrophoretic mobility observed between spots is due to differences in the glycosylation of Tf protein species, only the NeuGc could be responsible of the different p*I*s as the rest of monosaccharides are neutral. In Fig. 5A it can be observed that the most acidic glycoforms, those with increased NeuGc (e.g. 2Ant/3NeuGc), were more abundant in the acidic spots (1–3, low p*I*), while the less acidic glycoforms, those with less NeuGc (e.g. 2Ant/1NeuGc), were more abundant in the basic spots (4–6, high p*I*). Even though these results are just tentative as some glycoforms detected in Tf standard digested in solution (see Table 2) were not detected in any spot. This is likely due to the relatively less sensitivity of the technique (2-DE followed by the tryptic digestion of selected spots and subsequent μLC-TOF-MS analysis) as compared with that performed with the whole protein in solution. In fact, the total amount of Tf loaded in the gel was lower than the amount injected in solution, or even worse the protein was separated in 6 different spots. This is also why the signals of the glycopeptide glycoforms were even lower in the spots from spleen protein extracts, where the amount of protein per Tf spot was significantly lower than that in the spots from the Tf standard. Despite that, the abundance and composition of the glycoforms detected in spots #532, and #537 were quite similar to those detected in spots #2, and #3 of the Tf standard. These spots showed an equivalent position on the 2D-gels, which was indicative of similar p*I*s.

Table 12 of Ref [18] shows the glycopeptide glycoforms detected in the spots of spleen protein extracts, along with their normalized area and their %RSD values. It is important to note that it was not possible to obtain information about the glycoform composition of the major Tf spot from spleen, spot #538, and the most abundant Tf spot from the Tf standard, spot #5, despite the fact of using a surfactant during the trypsin digestion, which clearly implements the yield of glycopeptides in human Tf [16], and mouse Tf (this work). We are tempting to speculate that additional PTMs nearby the *N*-glycosylation site may impede the proper trypsin digestion of the Tf glycopeptide, resulting in a lower yield of glycoforms as compared with other less abundant spots. Alternatively, the additional PTM may have a negative influence on the ionization process of the glycopeptide, which without any further modification usually shows a significantly lower MS signal strength than its unglycosylated counterpart.

3.7. Analysis of Tf glycopeptide glycoforms in serum samples from CIA^+ and non-immunized mice

As the main function of Tf is the transport of iron through the blood plasma, the glycosylation pattern of serum Tf should also show differences between non-immunized and CIA^+ mice. Hence, an immunopurification of this protein present in serum followed by a tryptic digestion and a μLC-TOF-MS detection could prove to be an excellent alternative to the analysis of Tf in protein extracts from spleen. In order to analyze Tf from serum samples the immunoaffinity procedure described previously [16] was carried out. Even though the antibody used was raised against human Tf, the retention and elution of mouse Tf was still possible as both proteins share approximately 72% of the peptide sequence. After immunopurification, Tf was subjected to enzymatic digestion and the resulting glycopeptide and peptides were analyzed by μLC-TOF-MS. Four serum samples were analyzed following the established methodology: non-immunized and CIA^+ WT sera versus non-immunized and CIA^+ CD38-KO sera. Fig. 1 in Ref [18] shows the

Table 3
Normalized peak area and %RSD of Tf glycopeptide glycoforms detected in the spots of a Tf standard subjected to 2-DE separation and in-gel tryptic digestion.

Glycoforms	Spot 1		Spot 2		Spot 3		Spot 4		Spot 6	
	A_{norm}[a]	%RSD	A_{norm}[a]	%RSD	A_{norm}[a]	%RSD	A_{norm}[a]	%RSD	A_{norm}[a]	%RSD
1Ant/1NeuGc	–	–	–	–	–	–	2.5	6.1	–	–
1Ant/1NeuGc1Fuc	–	–	–	–	–	–	2.0	14.5	–	–
2Ant/1NeuGc	–	–	9.8	15.6	8.4	8.5	15.3	5.5	25.3	6.4
2Ant/1NeuGc1Fuc	–	–	7.3	14.0	6.0	8.1	13.1	6.7	23.3	10.2
2Ant/2NeuGc	33.6	13.2	58.2	4.3	63.4	1.1	90.8	2.1	97.2	4.0
2Ant/2NeuGc1Fuc	16.1	16.4	40.6	7.2	39.7	8.7	76.3	3.6	72.0	6.1
2Ant/3NeuGc	21.4	12.1	43.5	8.6	28.4	8.3	26.9	1.8	–	–
2Ant/3NeuGc1Fuc	8.0	10.4	14.1	11.9	11.4	6.9	8.5	3.7	–	–

[a] A_{norm}: normalized peak areas were calculated as: (Glycoform peak area / peptide 354–364 (CDEWSIISEGK) peak area) × 100.

Table 4
Normalized peak area and %RSD of Tf glycopeptide glycoforms detected in four serum samples: WT non-immunized, WT CIA$^+$, CD38-KO non-immunized and CD38-KO CIA$^+$.

Glycoforms	WT non-immunized		WT CIA$^+$		CD38-KO non-immunized		CD38-KO CIA$^+$	
	A_{norm}	%RSD	A_{norm}	%RSD	A_{norm}	%RSD	A_{norm}	%RSD
1Ant/1NeuGc	7.2	9.7	3.0	2.4	4.9	5.0	3.4	2.8
1Ant/1NeuGc1Fuc	1.3	9.9	1.6	4.5	1.9	4.3	2.4	5.6
2Ant/1NeuGc	20.4	0.9	16.1	5.9	20.7	1.6	19.9	4.6
2Ant/1NeuGc1Fuc	6.0	4.3	4.6	1.2	6.9	4.5	8.4	4.8
2Ant/2NeuGc	142.6	1.5	135.4	2.6	153.6	2.1	147.6	0.2
2Ant/2NeuGc1Fuc	71.5	3.1	60.8	3.2	58.9	1.2	64.5	1.0
2Ant/3NeuGc	46.1	1.2	38.3	3.5	42.9	2.2	35.1	4.6
2Ant/3NeuGc1Fuc	3.9	5.0	4.1	6.2	4.0	7.4	4.5	10.0
3Ant/3NeuGc	2.2	10.1	4.1	5.0	2.7	2.5	4.6	7.1
3Ant/3NeuGc1Fuc	0.5	9.0	0.7	4.7	0.6	12.2	0.6	12.9
4Ant/3NeuGc	4.3	4.6	3.9	12.2	5.6	1.6	5.2	6.0
4Ant/3NeuGc1Fuc	2.4	7.6	3.0	9.7	2.4	5.7	3.4	5.9

A_{norm}: normalized peak areas were calculated as: (Glycoform peak area / peptide 354–364 (CDEWSIISEGK) peak area) × 100.

EICs of the glycoforms detected in the Tf isolated from a WT non-immunized serum (Fig. 1A in Ref [18]) compared to the Tf standard onto a 2D-gel (spot #3, Fig. 1B in Ref [18]) and Tf from a spleen extract on a 2D-gel (spot #537 equivalent p*I* to spot #3, Fig. 1C in Ref [18]). As can be observed, the intensity of the peaks was at least tenfold higher in the serum compared to the spots, allowing the detection of an increased number of glycopeptide glycoforms. This demonstrates the ability of this methodology to correctly isolate and digest Tf and detect its glycopeptide glycoforms, compared to the analysis by 2-DE followed by the tryptic digestion of selected spots and subsequent μLC-TOF-MS analysis. Table 4 shows the detected glycoforms for each serum. As can be observed, the most abundant glycoforms in Tf standard were the most abundant as well in WT non-immunized serum and the same glycoforms were detected in both samples (compare Tables 2 and 4), which was to be expected as both samples are supposed to show similar composition of intact Tf glycoforms as the non-immunized WT serum can be regarded as a serum control. To clearly observe possible differences in Tf glycoforms between samples, Fig. 5C shows a bar graph with the logarithm in base 10 (log) of the A_{norm} for each glycopeptide glycoform in CIA$^+$ samples normalized to the A_{norm} of the same glycopeptide glycoform in non-immunized samples ($\log \frac{Anorm\ CIA+}{Anorm\ non-immunized}$) for both types of mice, WT and CD38-KO. As can be observed, with the exception of the 4Ant/3NeuGc glycoform, tri- and tetra-antennary glycoforms (4Ant/3NeuGc1Fuc, 3Ant/3NeuGc1Fuc and 3Ant/3NeuGc) increased in CIA$^+$ samples, whereas most of biantennary glycoforms decreased compared to the non-immunized samples. Moreover, most of fucosylated glycans seemed to also increase in arthritis-positive samples. This is in concordance with the alteration of the glycosylation pattern expected in inflammation, as it has been widely described an increase in fucosylation and glycan branching (number of antennae) in inflammatory diseases [41,42]. Additionally, in both cases a considerably reduction in the relative abundance of the glycoform 1Ant/1NeuGc was clearly observed. Another promising result is that in more than half of the detected glycoforms, the increase or decrease of the relative abundance of a given glycoform (e.g. 3Ant/3NeuGc1Fuc, 1Ant/1NeuGc or 2Ant/1NeuGc in Fig. 5B) was less pronounced in CD38-KO CIA$^+$ serum, which seems to be in concordance with the fact that mice deficient in CD38 develop an attenuated CIA$^+$ as compared with that in WT mice.

Our findings extend the results of others who described, by using crossed immune isoelectric focusing (CIEF) technique, a significant shift to highly sialylated human Tf (a pool fraction of 5 to 8-sialotransferrin species) in RA patients, with the concomitant decrease in the percentages of 4-sialotransferrin and low sialylated transferrin (0 to 3-sialotransferrin) fractions [43]. This shift in Tf microheterogeneity positively correlated with increased disease activity, whereas there was an inverse correlation between total Tf serum levels and parameters of disease activity, suggesting that changes in Tf synthetic rates and changes in Tf glycosylation induced in the inflammatory/autoimmune response are regulated independently [43]. Likewise, in our study spleen Tf levels measured by Western blotting were significantly lower in CIA$^+$ than in CIA$^-$ mice (0.3756 (n = 9) vs 0.6761 (n = 7), $P = 0.0052$ Mann Whitney test; median Tf densitometric values normalized by Cyclophilin A levels, data not shown).

4. Concluding remarks

Although these results need to be validated with a higher number of samples, this methodology could be implemented to identify Tf glycoforms specifically associated with either arthritis or inflammation, or to assess the contribution of a given gene or protein, such as CD38, to the aggravation or amelioration of these diseases. In this assessment it could be particularly interesting to study putative differences in the glycosylation pattern of serotransferrin in specific tissues associated with the disease (i.e.: joints), or in specific compartments such as exosomes/microvesicles, which are highly enriched in Tf receptors.

Disclosure

The authors have declared no conflict of interest.

Acknowledgments

We thank Dr. Frances E. Lund (Department of Microbiology, University of Alabama at Birmingham, Birmingham, Alabama, USA) for the gift of the CD38$^{-/-}$ mice and helpful discussions; to Pilar Navarro-Cuesta for her technical assistance.

The proteomic analyses were performed in the proteomics facilities of IPBLN-CSIC (Granada) and UCO-SCAI (Córdoba). Both facilities belong to ProteoRed, PRB2-ISCIII, supported by grant PT13/0001.

This work was supported in part by the European Commission in collaboration with the following Funding Agencies:

Ministerio de Economía y Competitividad (MINECO) del Gobierno de España grant number SAF2011-27261 (to J.S.), grant CTQ2011-27130, awarded to V. S-N; grant number SAF2011-22463 (to R.M.; cofunded by the European Regional Development Fund), and grant number SAF2012-34059 (to J.M.; cofunded by the European Regional Development Fund); Consejería de Innovación, Ciencia y Empresa de la Junta de Andalucía, grant number P08-CTS-04046 (to J.S.). From the MINECO del Gobierno de España, grant number IPT2011-1527-010000, associated to Fibrostatin SL, to J.M. A.R.V. was supported by a fellowship-contract from Consejería de Innovación, Ciencia y Empresa de la Junta de Andalucía. S.G.R. was supported by a JAEDoc contract from CSIC and a MINECO contract. A.B. was supported by a FPU fellowship from the Ministry of Education, Culture and Sport.

MS data are available via *ProteomeXchange with identifiers*: PXD002644, PXD002643, PXD003183, and PXD003163. Additional supporting data are shown in ref. [18].

References

[1] J.R. Whiteaker, C. Lin, J. Kennedy, L. Hou, M. Trute, I. Sokal, et al., A targeted proteomics-based pipeline for verification of biomarkers in plasma, Nat. Biotechnol. 29 (2011) 625–634.
[2] C. von Toerne, M. Kahle, A. Schäfer, R. Ispiryan, M. Blindert, M. Hrabe De Angelis, et al., Apoe, Mbl2, and Psp plasma protein levels correlate with diabetic phenotype in NZO mice—an optimized rapid workflow for SRM-based quantification, J. Proteome Res. 12 (2013) 1331–1343.
[3] J.J. Inglis, G. Criado, M. Medghalchi, M. Andrews, A. Sandison, M. Feldmann, et al., 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) R113.
[4] C.A. Notley, J.J. Inglis, S. Alzabin, F.E. McCann, K.E. McNamee, R.O. Williams, Blockade of tumor necrosis factor in collagen-induced arthritis reveals a novel immunoregulatory pathway for Th1 and Th17 cells, J. Exp. Med. 205 (2008) 2491–2497.
[5] A. Rosal-Vela, S. Garcia-Rodriguez, J. Postigo, M. Iglesias, V. Longobardo, A. Lario, et al., Distinct serum proteome profiles associated with collagen-induced arthritis and complete Freund's adjuvant-induced inflammation in CD38(−/−) mice: the discriminative power of protein species or proteoforms, Proteomics 15 (2015) 3382–3393.
[6] A. Hannigan, R. Burchmore, J.B. Wilson, The optimization of protocols for proteome difference gel electrophoresis (DiGE) analysis of preneoplastic skin, J. Proteome Res. 6 (2007) 3422–3432.
[7] P. Jungblut, B. Thiede, U. Zimny-Arndt, E.C. Muller, C. Scheler, B. Wittmann-Liebold, et al., Resolution power of two-dimensional electrophoresis and identification of proteins from gels, Electrophoresis 17 (1996) 839–847.
[8] L.M. Smith, N.L. Kelleher, Consortium for Top Down P, Proteoform: a single term describing protein complexity, Nat. Methods 10 (2013) 186–187.
[9] P.R. Jungblut, H.G. Holzhutter, R. Apweiler, H. Schluter, The speciation of the proteome, Chem. Cent. J. 2 (2008) 16.
[10] D.A. Cockayne, T. Muchamuel, J.C. Grimaldi, H. Muller-Steffner, T.D. Randall, F.E. Lund, et al., Mice deficient for the ecto-nicotinamide adenine dinucleotide glycohydrolase CD38 exhibit altered humoral immune responses, Blood 92 (1998) 1324–1333.
[11] J. Postigo, M. Iglesias, D. Cerezo-Wallis, A. Rosal-Vela, S. Garcia-Rodriguez, M. Zubiaur, et al., Mice deficient in CD38 develop an attenuated form of collagen type II-induced arthritis, PLoS ONE 7 (2012), e33534.
[12] E.J. Pavon, P. Munoz, A. Lario, V. Longobardo, M. Carrascal, J. Abian, et al., Proteomic analysis of plasma from patients with systemic lupus erythematosus: increased presence of haptoglobin alpha2 polypeptide chains over the alpha1 isoforms, Proteomics 6 (Suppl 1) (2006) S282–S292.
[13] E.J. Pavon, S. Garcia-Rodriguez, E. Zumaquero, R. Perandres-Lopez, A. Rosal-Vela, A. Lario, et al., Increased expression and phosphorylation of the two S100A9 isoforms in mononuclear cells from patients with systemic lupus erythematosus: a proteomic signature for circulating low-density granulocytes, J. Proteomics 75 (2012) 1778–1791.
[14] E.R. Vossenaar, A.J. Zendman, W.J. van Venrooij, G.J. Pruijn, PAD, a growing family of citrullinating enzymes: genes, features and involvement in disease, BioEssays 25 (2003) 1106–1118.
[15] V. Goeb, M. Thomas-L'Otellier, R. Daveau, R. Charlionet, P. Fardellone, X. Le Loet, et al., Candidate autoantigens identified by mass spectrometry in early rheumatoid arthritis are chaperones and citrullinated glycolytic enzymes, Arthritis Res. Ther. 11 (2009) R38.
[16] A. Barroso, E. Gimenez, 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.
[17] J.A. Vizcaino, E.W. Deutsch, R. Wang, A. Csordas, F. Reisinger, D. Rios, et al., ProteomeXchange provides globally coordinated proteomics data submission and dissemination, Nat. Biotechnol. 32 (2014) 223–226.
[18] A. Rosal-Vela, A. Barroso, E. Giménez, S. García-Rodríguez, V. Longobardo, J. Postigo, et al., Supporting data for the MS identification of multiple transferrin species – glycopeptide glycoforms – in spleen and serum from WT and CD38$^{-/-}$ mice with collagen-induced arthritis, Data in Brief, 2015 ("accepted for publication").
[19] G. Weiss, L.T. Goodnough, Anemia of chronic disease, N. Engl. J. Med. 352 (2005) 1011–1023.
[20] D. Domanski, G.V. Cohen Freue, L. Sojo, M.A. Kuzyk, L. Ratkay, C.E. Parker, et al., The use of multiplexed MRM for the discovery of biomarkers to differentiate iron-deficiency anemia from anemia of inflammation, J. Proteomics 75 (2012) 3514–3528.
[21] M.T. Barbosa, S.M. Soares, C.M. Novak, D. Sinclair, J.A. Levine, P. Aksoy, et al., The enzyme CD38 (a NAD glycohydrolase, EC 3.2.2.5) is necessary for the development of diet-induced obesity, FASEB J. 21 (2007) 3629–3639.
[22] E.N. Chini, CD38 as a regulator of cellular NAD: a novel potential pharmacological target for metabolic conditions, Curr. Pharm. Des. 15 (2009) 57–63.
[23] C. Escande, V. Nin, N.L. Price, V. Capellini, A.P. Gomes, M.T. Barbosa, et al., Flavonoid apigenin is an inhibitor of the NAD+ ase CD38: implications for cellular NAD+ metabolism, protein acetylation, and treatment of metabolic syndrome, Diabetes 62 (2013) 1084–1093.
[24] M.J. Rardin, J.C. Newman, J.M. Held, M.P. Cusack, D.J. Sorensen, B. Li, et al., Label-free quantitative proteomics of the lysine acetylome in mitochondria identifies substrates of SIRT3 in metabolic pathways, Proc. Natl. Acad. Sci. 110 (2013) 6601–6606.
[25] M.J. Rardin, W. He, Y. Nishida, J.C. Newman, C. Carrico, S.R. Danielson, et al., SIRT5 regulates the mitochondrial lysine succinylome and metabolic networks, Cell Metab. 18 (2013) 920–933.
[26] M.J. Rardin, J.M. Held, B.W. Gibson, Targeted quantitation of acetylated lysine peptides by selected reaction monitoring mass spectrometry, Methods Mol. Biol. 1077 (2013) 121–131.
[27] G. Colak, O. Pougovkina, L. Dai, M. Tan, H. Te Brinke, H. Huang, et al., Proteomic and biochemical studies of lysine malonylation suggests its malonic aciduria-associated regulatory role in mitochondrial function and fatty acid oxidation, Mol. Cell. Proteomics 14 (2015) 3056–3071.
[28] J. Park, Y. Chen, D.X. Tishkoff, C. Peng, M. Tan, L. Dai, et al., SIRT5-mediated lysine desuccinylation impacts diverse metabolic pathways, Mol. Cell 50 (2013) 919–930.
[29] E. Schurgers, F. Mertens, J.A.J. Vanoirbeek, S. Put, T. Mitera, E.D. Langhe, et al., Pulmonary inflammation in mice with collagen-induced arthritis is conditioned by complete Freund's adjuvant and regulated by endogenous IFN-γ, Eur. J. Immunol. 42 (2012) 3223–3234.
[30] A. Billiau, P. Matthys, Modes of action of Freund's adjuvants in experimental models of autoimmune diseases, J. Leukoc. Biol. 70 (2001) 849–860.
[31] K.A. Kuhn, L. Kulik, B. Tomooka, K.J. Braschler, W.P. Arend, W.H. Robinson, et al., Antibodies against citrullinated proteins enhance tissue injury in experimental autoimmune arthritis, J. Clin. Invest. 116 (2006) 961–973.
[32] B. Kidd, P. Ho, O. Sharpe, X. Zhao, B. Tomooka, J. Kanter, et al., Epitope spreading to citrullinated antigens in mouse models of autoimmune arthritis and demyelination, Arthritis Res. Ther. 10 (2008) R119.
[33] W.R. Alley Jr., M.V. Novotny, Structural glycomic analyses at high sensitivity: a decade of progress, Annu. Rev. Anal. Chem. 6 (2013) 237–265.
[34] M. Wuhrer, Glycomics using mass spectrometry, Glycoconj. J. 30 (2013) 11–22.
[35] S.B. Levery, C. Steentoft, A. Halim, Y. Narimatsu, H. Clausen, S.Y. Vakhrushev, Advances in mass spectrometry driven O-glycoproteomics, Biochim. Biophys. Acta 2015 (1850) 33–42.
[36] M.J. Kailemia, L.R. Ruhaak, C.B. Lebrilla, I.J. Amster, Oligosaccharide analysis by mass spectrometry: a review of recent developments, Anal. Chem. 86 (2014) 196–212.
[37] Y. Zhang, J. Jiao, P. Yang, H. Lu, Mass spectrometry-based N-glycoproteomics for cancer biomarker discovery, Clin. Proteomics 11 (2014) 18.
[38] M.V. Novotny, W.R. Alley Jr., B.F. Mann, Analytical glycobiology at high sensitivity: current approaches and directions, Glycoconj. J. 30 (2013) 89–117.
[39] K. Marino, 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.
[40] 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(alpha 2–3)Gal(beta 1–3)[Neu5Gc(alpha 2–6)]GlcNAc(beta 1–2)Man, Biochim. Biophys. Acta 1475 (2000) 321–328.
[41] D.H. Dube, C.R. Bertozzi, Glycans in cancer and inflammation—potential for therapeutics and diagnostics, Nat. Rev. Drug Discov. 4 (2005) 477–488.
[42] S. Albrecht, L. Unwin, M. Muniyappa, P.M. Rudd, Glycosylation as a marker for inflammatory arthritis, Cancer Biomark. 14 (2014) 17–28.
[43] R.A. Feelders, G. Vreugdenhil, G. de Jong, A.J. Swaak, H.G. van Eijk, Transferrin microheterogeneity in rheumatoid arthritis. Relation with disease activity and anemia of chronic disease, Rheumatol. Int. 12 (1992) 195–199.

Data in Brief

Data in Brief 6 (2016) 587–602

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data article

Supporting data for the MS identification of distinct transferrin glycopeptide glycoforms and citrullinated peptides associated with inflammation or autoimmunity

A. Rosal-Vela [a,1], A. Barroso [b,1], E. Giménez [b], S. García-Rodríguez [b], V. Longobardo [c], J. Postigo [d], M. Iglesias [d], A. Lario [c], J. Merino [d], R. Merino [e], M. Zubiaur [a], V. Sanz-Nebot [b], J. Sancho [a,*]

[a] Instituto de Parasitología y Biomedicina "López-Neyra" (IPBLN), CSIC, Armilla, Granada, Spain
[b] Departament de Química Analítica, Universitat de Barcelona, Spain
[c] Unidad de Proteómica, IPBLN, CSIC, Armilla, Granada, Spain
[d] Facultad de Medicina, Universidad de Cantabria, Santander, Spain
[e] Instituto de Biomedicina y Biotecnología de Cantabria, CSIC, Santander, Spain

ARTICLE INFO

Article history:
Received 14 December 2015
Received in revised form
17 December 2015
Accepted 23 December 2015
Available online 11 January 2016

Keywords:
Arthritis
Inflammation
Protein species
Transferrin
Glycosylation
CD38
Citrullination

ABSTRACT

This data article presents the results of all the statistical analyses applied to the relative intensities of the detected 2D-DiGE protein spots for each of the 3 performed DiGE experiments. The data reveals specific subsets of protein spots with significant differences between WT and CD38-deficient mice with either Collagen-induced arthritis (CIA), or with chronic inflammation induced by CFA, or under steady-state conditions. This article also shows the MS data analyses that allowed the identification of the protein species which serve to discriminate the different experimental groups used in this study. Moreover, the article presents MS data on the citrullinated peptides linked to specific protein species that were generated in CIA^+ or CFA-treated mice. Lastly, this data article provides MS data on the efficiency of the analyses of the transferrin (Tf) glycopeptide glycosylation pattern in spleen and serum from CIA^+ mice and normal controls. The data supplied in this work is related to the research article entitled "identification of multiple transferrin species in spleen and serum from mice with collagen-induced arthritis which may

DOI of original article: http://dx.doi.org/10.1016/j.jprot.2015.11.023
* Correspondence to: Instituto de Parasitología y Biomedicina López-Neyra, CSIC, Avenida del Conocimiento s/n, Parque Tecnológico de la Salud (PTS), 18016 Armilla (Granada), Spain. Tel.: +34 958181664; fax: +34 958181632.
E-mail address: granada@ipb.csic.es (J. Sancho).
[1] A. Rosal-Vela and A. Barroso equally contributed to this work.

http://dx.doi.org/10.1016/j.dib.2015.12.045

reflect changes in transferrin glycosylation associated with disease activity: the role of CD38" [1]. All mass spectrometry data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with identifiers PRIDE: PXD002644, PRIDE: PXD002643, PRIDE: PXD003183 and PRIDE: PXD003163.

Specifications table

Subject area	*Biology*
More specific subject area	*Proteomics and glycoproteomics*
Type of data	*Tables, figures and raw data*
How data was acquired	Scanned 2D-DiGE images were analyzed using the DeCyder7.0 software (GE Healthcare) using the Differential In-gel Analysis (DIA) module to detect and normalize the protein spots. Protein relative abundance across all samples and statistical analyses were performed using the Biological Variation Analysis (BVA) module of the DeCyder software. *MS data for protein identification was acquired using a MALDI TOF/TOF UltrafleXtreme (Bruker), or a* 4800 MALDI-TOF/TOF Analyzer (AB SCIEX). *μLC–TOF–MS data for the analysis of the glycopeptides glycoforms of Tf was acquired with a 1200 series capillary liquid chromatography system (Agilent Technologies) coupled to a 6220 oa-TOF LC/MS mass spectrometer with an orthogonal G1385-44300 interface (Agilent Technologies).*
Data format	*Analyzed (excel files and word tables) and raw data*
Experimental factors	*Mice with Collagen-induced arthritis, or with chronic inflammation, or with no treatment. Protein extraction and/or purification from spleen or serum samples. CyDye labeling. 2-D gel electrophoresis.*
Experimental features	*Protein extracts from mice subjected to different experimental conditions were analyzed by 2D-DiGE, and protein species that differed in abundance were identified by MS/MS. PTMs such as citrullination of the identified proteins, or glycosylation of Tf species were further analyzed by MS.*
Data source location	*UB: Barcelona; UCO: Córdoba; IPBLN: Granada.*
Data accessibility	*Data is within this article. Data also available at the ProteomeXchange Consortium via the PRIDE partner repository, PRIDE: PXD002644, PRIDE: PXD002643, PRIDE: PXD003183 and PRIDE: PXD003163.*

Value of the data

- Application of μLC–TOF–MS for characterization of multiple glycopeptide glycoforms from mouse transferrin.
- Investigation of altered transferrin glycopeptide glycosylation patterns in inflammatory and/or autoimmune diseases.
- Mass spectrometry approach to identify new citrullinated peptides in mice with arthritis (CIA model).
- Properly described approach for 2D-DiGE analysis to identify protein species that differ in abundance due to certain pathologies.
- Basis for the study of altered protein species associated with inflammatory processes or arthritis in humans.

1. Data

Fig. 1 shows the extracted ion chromatograms (EICs) obtained by μLC–TOF–MS for the most abundant glycopeptide glycoforms of Tf isolated from WT non-immunized serum, Tf standard in a 2D gel, and Tf from a spleen extract in a 2D gel. Tables 1 and 2 in excel format show the list of the protein species identified by MS/MS, displaying the sequence of matched and fragmented peptides of a given protein. Table 3 shows the list of protein species identified by PMF. Tables 4–9, include the results of all the statistical analyses applied to the relative intensities of the detected 2D-DiGE protein spots for each of the 3 performed DiGE experiments. Table 10 shows the identities of the citrullinated peptides linked to specific protein species in CIA^{+}, or CFA-treated mice. Table 11 shows the peptide coverage of mouse Tf standard digested with trypsin, and Table 12 shows the normalized peak area and %RSD of Tf glycopeptide glycoforms detected by μLC–TOF–MS in the spots of spleen protein extracts subjected to 2D electrophoretic separation and in-gel tryptic digestion.

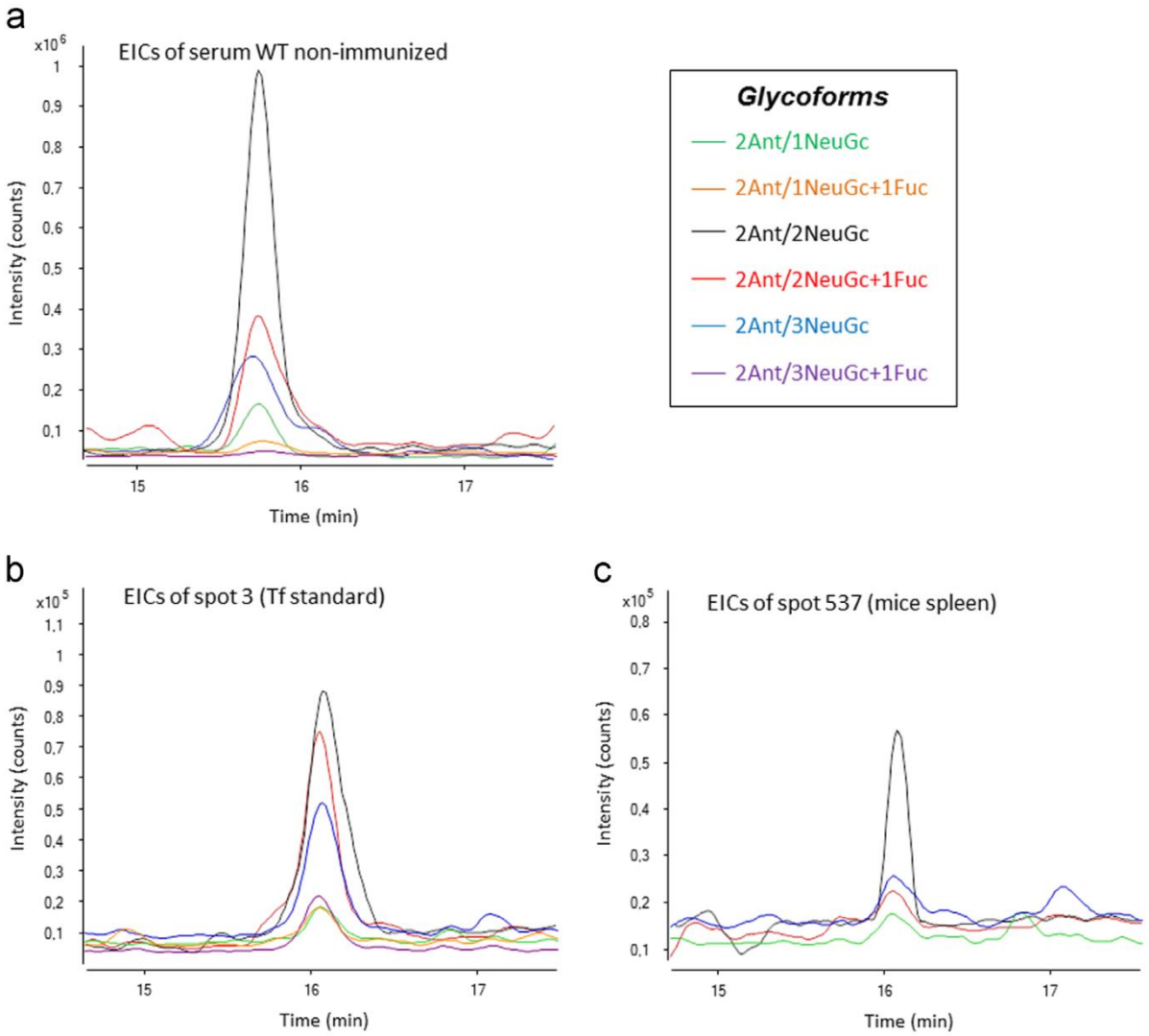


Fig. 1. Extracted ion chromatograms (EICs) for the most abundant glycopeptide glycoforms of Tf isolated from (a) WT non-immunized serum, (b) Tf standard in a 2D gel (spot 3), and (c) Tf from a spleen extract in a 2D gel (spot 537 equivalent *pI* to spot 3) by μLC–TOF–MS.

Table 1
Protein spots from mouse spleen identified by MS/MS using the 4800 MALDI-TOF/TOF (AB Sciex).

Spot number[a]	Identification	Accession	Accension number[b]	Protein score	Protein score CI%[c]	Pep. count	Total ion score[d]	Total % ion CI	MW (theorical)[e]	IEP (theorical)[e]
533	Serotransferrin	TRFE_MOUSE	Q921I1	119	100	11	69	100.000	78840.5	6.94
537	Serotransferrin	TRFE_MOUSE	Q921I1	217	100	18	109	100.000	78840.5	6.94
539	Serotransferrin	TRFE_MOUSE	Q921I1	461	100	25	266	100.000	78840.5	6.94
614	Stress-induced-phosphoprotein 1	STIP1_MOUSE	Q60864	277	100	19	146	100	63,170	6.40
692	Fibrinogen beta chain	FIBB_MOUSE	Q8K0E8	800	100	25	578	100	55401.9	6.68
697	Catalase	CATA_MOUSE	P24270	179	100	10	131	100	60012.7	7.72
778	Alpha enolase	ENOA_MOUSE	P17182	891	100	27	594	100	47453.3	6.37
	Beta enolase	ENOB_MOUSE	P21550	250	100	6	224	100	47337.4	6.73
	Gamma enolase	ENOG_MOUSE	P17183	94	100	5	75	100	47609.1	4.99
906	Voltage-dependent anion selective channel protein 2	VDAC2_MOUSE	Q60930	179	100	11	95	100	32339.80	7
981	Proteasome subunit alpha type-1	PSA1_MOUSE	Q9R1P4	426	100	10	350	100.000	29812.9	6
982	Carbonic anhydrase 2	CAH2_MOUSE	P00920	478	100	17	295	100.000	29128.5	6.49
	Carbonic anhydrase 1	CAH1_MOUSE	P13634	203	100	7	155	100.000	28360.2	6.44
1001	Proteasome subunit alpha type-6	PSA6_MOUSE	Q9QUM9	335	100	10	264	100.000	27,811	6.34
	Pyridoxine-5′-phosphate oxidase	PNPO_MOUSE	Q91XF0	81	99.999	5	57	99.999	30437.1	8.46
1063	Flavin reductase	BLVRB_MOUSE	Q923D2	777	100	16	583	100.000	22297.4	6.49
1097	ATP synthase subunit d, mitochondrial	ATP5H_MOUSE	Q9DCX2	243	100	10	156	100.000	18794.6	5.52
1103	Ferritin light chain 1	FRIL1_MOUSE	P29391	585	100	14	435	100.000	20846.5	5.66
	Ferritin light chain 2	FRIL2_MOUSE	P49945	276	100	9	202	100.000	20886.9	6.39
1136	Peptidyl-prolyl cis-trans isomerase A	PPIA_MOUSE	P17742	340	100	10	249	100	18130.9000	8
1147	Actin-related protein 2/3 complex subunit 5-like protein	ARP5L_MOUSE	Q9D898	165	100	6	116	100	17026.80	6.32
1171	Nucleoside diphosphate kinase A	NDKA_MOUSE	P15532	654	100	13	518	100.000	17310.9	6.84
	Nucleoside diphosphate kinase B	NDKB_MOUSE	Q01768	330	100	7	278	100.000	17,466	6.97
1184	Protein S100-A9	S10A9_MOUSE	P31725	95	100	1	89	100	13,211	6.64
1313	Protein S100-A8	S10A8_MOUSE	P27005	146	100	3	119	100.000	10345.1	5.43

The sequence of matched and fragmented peptides of the identified proteins, plus the ion scores and confidence intervals of the fragmented peptides can be found in the online version of this article (Table 1, .xlsx file) as supplementary material.

[a] Spots are named as indicated on the 2-DE gel shown in Fig. 1 in Ref [1].

[b] UniProtKB/Swiss-Prot accession number, MASCOT protein score, protein score confidence interval (C.I. %), total ion score, and total ion score confidence interval (C.I. %) are reported for the combined search of MALDI-TOF/TOF MS and MS/MS data (GPS Software, Applied Biosystems).

[c] For protein scores, only confidence intervals above 99% were considered as statistically significant.

[d] For total ion scores, only confidence intervals above 95% were considered as statistically significant.

[e] Theoretical molecular weights and isoelectric points are given for each protein.

Table 2
Protein spots identified by MS/MS using the MALDI-TOF/TOF UltrafleXtreme (Bruker).

Spot number[a]	Identification	Accession	Accession number[b]	Protein score	Sequence coverage (%)	No. of peptides	MW (theorical)[c]	IEP (theorical)[c]
501	Aconitate hydratase, mitochondrial	ACON_MOUSE	Q99KI0	93.05	3.60	2	85,400	8.08
538	Serotransferrin	TRFE_MOUSE	Q92111	96.65	4	3	78,841	6.94
539	Serotransferrin	TRFE_MOUSE	Q92111	190.90	5.9	4	78,841	6.94
554	Far upstream element-binding protein 1	FUBP1_MOUSE	Q91WJ8	32.75	2.00	1	68,668	7.74
633	Fibrinogen alpha chain	FIBA_MOUSE	E9PV24	123.51	4.20	3	88,117	5.77
638	Heterogeneous nuclear ribonucleoprotein L	HNRPL_MOUSE	Q8R081	69.55	4.60	2	64,550	8.33
732	Coronin-1	CORO1A_MOUSE	gi|4895037	102.3	2.2	1	51,627	6.05
738	Vimentin	VIME_MOUSE	P20152	286.83	16.10	5	53,712	5.06
770	Protein disulfide-isomerase A6 precursor	PDIA6_MOUSE	gi|58037267	117.78	10.6	3	49,058	5.05
898	Actin, cytoplasmic 1	ACTB_MOUSE	P60710	177.47	9.10	2	42,052	5.29
	Beta-actin-like protein 2	ACTBL_MOUSE	Q8BFZ3	129.62	9.00	2	42,319	5.30
	F-actin-capping protein subunit alpha-1	CAZA1_MOUSE	P47753	92.04	9.80	2	33,090	5.34
972	Tropomyosin alpha-1 chain	TPM1_MOUSE	P58771	59.03	4.90	1	32,718	4.69
981	Proteasome subunit alpha type-1	PSA1_MOUSE	Q9R1P4	133.78	13.7	3	29,813	6.00
1001	Proteasome subunit alpha type-6	PSA6_MOUSE	Q9QUM9	177.79	16.3	4	27,811	6.34
1049	Proteasome subunit beta type-10	PSB10_MOUSE	O35955	109.4	8.80	3	29,330	6.40
	Growth factor receptor-bound protein 2	GRB2_MOUSE	Q60631	103.47	11.50	3	25,336	5.89
1103	Ferritin light chain 1	FRIL1_MOUSE	P29391	48.36	4.4	1	20,847	5.66
1150	E3 ubiquitin-protein ligase RNF181	RN181_MOUSE	Q9CY62	26.01	4.80	1	19,487	5.65
1173	Ubiquitin-conjugating enzyme E2 N	UBE2N_MOUSE	P61089	104.79	19.70	2	17,184	6.13
1302	N-acyl-aromatic-L-amino acid amidohydrolase (carboxylate-forming)	ACY3_MOUSE	Q91XE4	23.11	2.50	1	35,720	5.30

The sequence of matched and fragmented peptides plus the ion scores and confidence intervals of the fragmented peptides can be found in the online version of this article (Table 2, .xlsx file) as supplementary material.

[a] Spots are named as indicated on the 2-DE gel shown in Fig. 1 in Ref [1].

[b] UniProtKB/Swiss-Prot or NCBI accession number.

[c] Theoretical molecular weights and isoelectric points are given for each protein.

Table 3
Protein spots identified by PMF using the MALDI-TOF/TOF UltrafleXtreme (Bruker).

Spot number[a]	Protein name	Accession number[b]	MW (theorical)[c]	IEP (theorical)[c]	Score	Expect	Sequence coverage	Queries matched	Queries searched
572	Prelamin-A/C isoform A precursor (MS)	gi|62287370	74,478	6.54	77	0.0034	29	18	73
	Prelamin-A/C isoform C (MS)	gi|61760667	65,464	6.37	69	0.02	33	17	73
	Fibroblast growth factor 22 (MS)	gi|12963627	18,972	11.73	72	0.011	61	10	73
983	Mitochondrial peptide methionine sulfoxide reductase	Q9D6Y7	26,200	8.6	76	0.00042	34.3	7	37
1104	Low molecular weight phosphotyrosine protein phosphatase	Q9D358	18,636	6.30	58.5	2.40E−02	31.6	5	35

[a] Spots are named as indicated on the 2-DE gel shown in Fig. 1 in Ref [1].
[b] UniProtKB/Swiss-Prot or NCBI accession number.
[c] Theoretical molecular weights and isoelectric points are given for each protein.

Table 4
Spleen protein species that differ in abundance by 2-ANOVA-Mouse in Col II immunized CD38 KO mice versus B6 WT mice.

DeCyder spot no.	Protein name[a]	*P* value (2-ANOVA-Mouse)
B6 WT:		
538	Serotransferrin	3.99E−04
633	Fibrinogen alpha chain	6.11E−04
692	Fibrinogen beta chain	1.12E−03
1097	ATP synthase subunit d, mitochondrial	1.30E−03
1302	N-acyl-aromatic-L-amino acid amidohydrolase (carboxylate-forming)	1.75E−03
539	Serotransferrin	3.53E−03
533	Serotransferrin	8.87E−03
554	Far upstream element-binding protein 1	9.53E−03
537	Serotransferrin	1.61E−02
501	Aconitate hydratase, mitochondrial	1.78E−02
1171	Nucleoside diphosphate kinase A Nucleoside diphosphate kinase B	0.0195
1103	Ferritin light chain 1 Ferritin light chain 2	2.06E−02
1300	Not identified	0.023
638	Heterogeneous nuclear ribonucleoprotein L	2.89E−02
572	Prelamin-A/C isoform A precursor (PMF) Prelamin-A/C isoform C (PMF) Fibroblast growth factor 22 (PMF)	3.59E−02
1313	Protein S100-A8	0.05
CD38 KO:		
898	Actin, cytoplasmic 2 Beta-actin-like protein 2 F-actin-capping protein subunit alpha-1	6.70E−03
982	Carbonic anhydrase 2 Carbonic anhydrase 1	8.98E−03
1044	Not identified	0.0202
981	Proteasome subunit alpha type-1	2.39E−02

[a] Protein name according to UniProt, or to NCBI.

Table 5
Spleen protein species that differ in abundance by 2-ANOVA-Arthritis test in Col II-immunized CIA^+ versus CIA^- mice.

DeCyder spot no.	Protein name[a]	*P* value (2-ANOVA-Arthritis)
In CIA^+:		
438	Not identified	0.0162
1150	E3 ubiquitin-protein ligase RNF181	0.0414
In CIA^-:		
1157	Not identified	0.021
778	Alpha-enolase Beta-enolase Gamma-enolase	0.0365

[a] Protein Name according to UniProt, or to NCBI.

Table 6
Spleen protein species that differ in abundance by 2-ANOVA-Interaction in two groups of Col.II-immunized mice (CD38 KO and B6 WT) with two conditions: CIA^+ and CIA^-).

DeCyder spot no.	Protein name[a]	*P* value(2-ANOVA-Mouse)	*P* value(2-ANOVA-Arthritis)	*P* value(2-ANOVA-Interaction)
1302	N-acyl-aromatic-L-amino acid amidohydrolase (carboxylate-forming)	1.75E−03	0.535	1.27E−04
532	Not identified	0.065	0.585	5.29E−04
538	Serotransferrin	3.99E−04	0.255	8.40E−03
982	Carbonic anhydrase 2 Carbonic anhydrase 1	8.98E−03	0.771	8.99E−03
1001	Proteasome subunit alpha type-6 Pyridoxine-5'-phosphate oxidase	0.878	0.303	9.65E−03
983	Mitochondrial peptide methionine sulfoxide reductase (PMF)	0.545	0.849	3.57E−02
1063	Flavin reductase	0.707	0.779	0.0362
537	Serotransferrin	1.61E−02	0.306	4.07E−02

[a] Protein name according to UniProt, or to NCBI.

2. Experimental design, materials and methods

2.1. Mice

WT mice were purchased from Harlan Ibérica (Barcelona, Spain). Mice deficient in CD38 (CD38-KO) were backcrossed onto the B6 background for more than 12 generations, as described previously [3]. All studies with live animals were approved by the IPBLN and Universidad de Cantabria Institutional Laboratory Animal Care and Use Committees.

2.2. Induction and assessment of arthritis

For the induction of CIA, 8–12 weeks-old male mice were immunized as previously described [4,5].

2.3. Protein extraction from spleen preparations

Proteins were extracted from spleen by using the MicroRotofor Lysis Kit (for mammalian tissues and cells) (Bio-Rad, Ref 163-2141), following the manufacturer's instructions, which includes the use of mini-grinders for effective disruption of cells and tissues. The excess of salts and other contaminants were removed using the Bio-Rad's ReadyPrep 2-D cleanup kit. Samples were then resuspended in a DIGE-compatible buffer (7 M urea, 2 M thiourea, 4% CHAPS, 20 mM Tris, pH 8.5), quantified using the RC DC assay, and kept at −20 °C until further use.

2.4. Design of DiGE experiments

Unless otherwise indicated in each DiGE experiment conducted, four biological replicates of each condition were compared, comprising protein samples derived from four CD38-KO mice and four WT mice as previously described [1,6].

2.5. DiGE labeling and two-dimensional gel electrophoresis

Samples were aliquoted at 45 μg, and the pooled internal standard was made with 23 μg of each of the sixteen test samples combined. The proteins were labeled with 400 pmol (in 1 μL of anhydrous DMF) of CyDye per 50 μg of protein as per the manufacturer's instructions (GE Healthcare). After labeling, the appropriate samples were combined for each gel. Each combined sample (~50 μL) was

Table 7
Differences in spleen protein species abundance compared between Col II immunized CD38 KO mice (test group) versus Col II-immunized B6 WT mice (control group).

DeCyder spot no.	Protein name[a]	Average ratio[b]	*P* value (*t*-test)
		Decreased abundance	
633	Fibrinogen alpha chain	−1.35	1.67E−04
692	Fibrinogen beta chain	−1.19	4.53E−04
1097	ATP synthase subunit d, mitochondrial	−1.25	4.95E−04
538	Serotransferrin	−1.23	1.35E−03
539	Serotransferrin	−1.29	2.25E−03
533	Serotransferrin	−1.39	4.17E−03
554	Far upstream element-binding protein 1	−1.17	5.26E−03
501	Aconitate hydratase, mitochondrial	−1.17	9.81E−03
1103	Ferritin light chain 1 Ferritin light chain 2	−1.37	0.0124
638	Heterogeneous nuclear ribonucleoprotein L	−1.23	0.0142
537	Serotransferrin	−1.25	0.0208
572	Prelamin-A/C isoform A precursor (MS) Prelamin-A/C isoform C (MS) Fibroblast growth factor 22 (MS)	−1.18	0.0209
1171	Nucleoside diphosphate kinase A Nucleoside diphosphate kinase B	−1.17	0.0258
697	Catalase	−1.16	0.0453
536	Not identified	−1.14	0.0475
1300	Not identified	−1.33	0.0493
		Increased abundance	
898	Actin, cytoplasmic 2 Beta-actin-like protein 2 F-actin-capping protein subunit alpha-1	1.19	4.37E−03
1044	Not identified	1.25	0.0163
981	Proteasome subunit alpha type-1	1.27	0.0194
982	Carbonic anhydrase 2 Carbonic anhydrase 1	1.16	0.0457

[a] Protein name according to UniProt, or to NCBI.
[b] Negative average ratios indicate decreased protein abundance, while positive ratios indicate increased protein abundance in CIA^+ CD38 KO relative to that in CIA^+ B6 WT mice.

made up to 200 μL with Readyprep Rehydration/Sample buffer (8 M urea, 2% CHAPS, 50 mM dithiothreitol (DTT), 0.2% (w/v) Bio-Lyte® 3/10 ampholytes, and Bromophenol Blue (trace)).

2-DE was carried out using the Protean IEF cell and Criterion electrophoresis cell systems (Bio-Rad, Hercules, CA, USA) as previously described [7], with the following modifications: (1) First-dimension IPG strips (Bio-Rad: 11 cm, linear pH 3-10 gradient); (2) Active in-gel rehydration at 50 V, 12 h at 20 °C; (3) The IPG strips were focused in a one-step procedure, at 8000 V for a total of 35,000 Vh at 20 °C with a current limit of 50 μA/strip.

Table 8
Chronic inflammation model. Differences in spleen protein species abundance compared between CFA/IFA-treated CD38 KO mice (test group) and CFA/IFA-treated B6 WT mice (control group).

DeCyder spot no.	Protein name[a]	Average ratio[b]	*P* value (*t*-test)
		Decreased abundance	
1049	Proteasome subunit beta type-10 Growth factor receptor-bound protein 2	−1.18	4.06E−03
1330	Not identified	−1.31	0.0106
972	Tropomyosin alpha-1 chain	−1.11	0.0163
1173	Ubiquitin-conjugating enzyme E2 N	−1.13	0.0217
1103	Ferritin light chain 1 Ferritin light chain 2	−1.62	0.0343
738	Vimentin	−1.22	0.0351
1104	Low molecular weight phosphotyrosine protein phosphatase (PMF)	−1.28	0.0384
732	Coronin-1	−1.1	0.0434
		Increased abundance	
1313	Protein S100-A8	1.37	0.0236

[a] Protein name according to UniProt, or to NCBI.
[b] Negative average ratios indicate decreased protein abundance, while positive ratios indicate increased protein abundance in CFA/IFA-treated CD38 KO relative to that in CFA/IFA-treated B6 WT mice.

Table 9
Non-immunized control mice. Differences in spleen protein species abundance compared between non-immunized CD38 KO mice (test group) and non-immunized B6 WT mice (control group).

DeCyder spot no.	**Protein name[a]**	**Average ratio[b]**	***P* value (*t*-test)**
		Decreased abundance	
538	Serotransferrin	−1.26	4.04E−03
539	Serotransferrin	−1.23	0.0185
537	Serotransferrin	−1.22	0.0221
638	Heterogeneous nuclear ribonucleoprotein L	−1.13	0.0266
770	Protein disulfide-isomerase A6	−1.11	0.0437
532	Not identified	−1.2	0.0477
		Increased abundance	
1295	Not identified	1.33	0.0162

[a] Protein name according to UniProt, or to NCBI.
[b] Negative average ratios indicate decreased protein abundance, while positive ratios indicate increased protein abundance in CFA/IFA-treated CD38 KO relative to that in CFA/IFA-treated B6 WT mice.

After electrophoresis, one of the gels was pre-scanned using the Typhoon 9400 variable mode imager at each of the appropriate CyDye excitation wavelengths (Cy3 (532 nm), Cy5 (633 nm), Cy2 (488 nm)), in order to determine the appropriate laser intensity for each CyDye. Thereafter, each of the analytical gels was scanned at this optimum laser intensity at a resolution of 100 μm. Gels were then fixed and stained with SYPRO Ruby (Bio-Rad) and re-scanned using the 488 nm laser. Scanned images were analyzed using

Table 10a
Citrullinated protein species and peptides[a] detected in spleen from collagen-induced arthritis, or CFA-treated mice. TOF/TOF 4800.

Spot number	Identification	Accesion	Accension number	Protein score	Protein score CI%	Pep. count	Total ion score	Total % ion CI	MW (theoretical)	IEP (theoretical)
537	Serotransferrin	TRFE_MOUSE	Q921I1	172	100	20	107	100.00	78840.50	6.94
614	Stress-induced-phosphoprotein 1	STIP1_MOUSE	Q60864	238	100	27	145	100.00	63169.60	6.40
697	Catalase	CATA_MOUSE	Q8C6E3	174	100	15	131	100.00	60082.80	7.73
778	Alpha enolase	ENOA_MOUSE	P17182	940	100	33	675	100.00	47453.30	6.37
	Beta enolase	ENOB_MOUSE	P21550	260	100	12	224	100.00	47337.40	6.73
	Gamma enolase	ENOG_MOUSE	P17183	102	100	10	75	100.00	47609.10	4.99
981	Proteasome subunit alpha type-1	PSA1_MOUSE	Q9R1P4	415	100	14	350	100.00	29812.90	6.00
	Proteasome subunit alpha type	PSMA1_MOUSE	Q8BTU5	399	100	12	349	100.00	29732.80	5.78
1001	Proteasome subunit alpha type-6	PSA6_MOUSE	Q9QUM9	325	100	13	264	100.00	27811.00	6.34
	Pyridoxine-5'-phosphate oxidase	PNPO_MOUSE	Q91XF0	81	99.999	5	57	100.00	30437.10	8.46
1103	Ferritin light chain 2	FRIL2_MOUSE	P49945	251	100	10	202	100.00	20886.90	6.39
1136	Peptidyl-prolyl cis–trans isomerase A	PPIA_MOUSE	P17742	320	100	11	249	100.00	18130.90	7.74
1171	Nucleoside diphosphate kinase A	NDKA_MOUSE	P15532	726	100	17	591	100.00	17310.90	6.84
	Nucleoside diphosphate kinase B	NDKB_MOUSE	Q01768	321	100	9	278	100.00	17466.00	6.97

[a] The sequence of matched citrullinated peptides of a given protein, and the positions of the deiminated arginines are shown in online version of this article as Supplementary material (Table 10, xlsx. File).

Table 10b
Citrullinated protein species and peptides[a] detected in spleen from collagen-induced arthritis, or CFA-treated mice. TOF/TOF UltrafleXtreme.

Spot number	Protein name	Accession	Accesion number	MW (theoretical)	IEP (theoretical)	Score	Sequence coverage (%)	Queries matched	Queries searched
633	Fibrinogen alpha chain	FIBA_MOUSE	E9PV24	87.40	5.78	101.00	17.00	15	26
732	Coronin-1A	COR1A_MOUSE	O89053	51.00	6.04	83.40	36.00	16	75
898	F-actin-capping protein subunit alpha-1	CAZA1_MOUSE	P47753	32.90	5.34	77.90	43.00	12	65
972	Tropomyosin alpha-3 chain	TPM3_MOUSE	P21107	33.00	4.68	60.50	38.60	10	40

[a] The sequence of matched citrullinated peptides of a given protein, and the positions of the deiminated arginines are shown in online version of this article as Supplementary material (Table 10, .xlsx file).

the DeCyder7.0 software (GE Healthcare) using the Differential In-gel Analysis (DIA) module to detect and normalize the protein spots. Standard was used to normalize gels by calculating the standardized abundance of each spot, i.e., the ratio of either Cy3 or Cy5 signal to that of Cy2.

2.6. Protein identification by MALDI-TOF/TOF MS/MS

In-gel digestion of proteins has been described previously [8]. A set of protein spots were identified by MS/MS using a 4800 MALDI-TOF/TOF Analyzer (AB SCIEX) in automatic mode with the settings described previously [6]. Protein identification was assigned by peptide mass fingerprinting and confirmed by MS/MS analysis of at least three peptides in each sample. Mascot 2.0 search engine (Matrixscience) was used for protein identification running on GPS software (Applied Biosystems) against the SwissProt *Mus musculus* database (uniprot_sprot_26042011.fasta). The search setting allowed one missed cleavage with the selected trypsin enzyme, a MS/MS fragment tolerance of 0.2 Da and a precursor mass tolerance of 100 ppm.

Other spots were identified by MS/MS using a MALDI TOF/TOF UltrafleXtreme (Bruker) in manual mode as previously described [6]. Fragment selection criteria were a minimum S/N ratio of 15, a maximum number of peaks set at 200. For each precursor selected for MS/MS analysis, fragment mass values in the range from 13 Da to 4 Da below precursor mass were used to peptide identification.

Protein identification was assigned by peptide mass fingerprinting and confirmed by MS/MS analysis of 5 peptides. Mascot Server 2.4 (Matrixscience) and ProteinScape 3.1 (Bruker) were used for protein identification against the SwissProt *Mus musculus* database (SwissProt_2015_06.fasta and NCBInr_20150409.fasta). The search setting allowed two missed cleavage with the selected trypsin enzyme, fixed modification was cysteine carbamidomethylation and variable modification was methionine oxidation, a MS/MS fragment tolerance of 0.5 Da and a precursor mass tolerance of 50 ppm, unless otherwise indicated.

The MS spectra of the identified proteins were further examined in order to detect the presence of citrullinated proteins. Protein citrullination (o deimination) is the enzymatic conversion of peptidyl-arginine residues to peptidyl-citruline, mediated by the family of calcium-dependent peptidylarginine deiminases (PADs) [9]. The search setting for this PTM with MASCOT was performed as in the previous paragraph, including as variable modification the deamination of arginine, with the following considerations [10]: (a) for one citrullinated arginine, the peptide theoretical mass increase is 0.98 Da and the modified peptide, losing one amino group, becomes more acidic; (b) citrullinated arginine residues are not likely to be cleaved by trypsin, so that a minimum number of one missed cleavage must be specified; (c) a peptide that includes a C-terminal citrullinated arginine must be rejected; (d) citrullinated peptides generate an unusual isotopic mass cluster as compared with that of unmodified peptides.

Table 11
Detected peptides in a tryptic digest of standard mTf analyzed by μLC–MS–TOF.

Detected peptides in mTf standard	
VPDK	✓
TVK	✓
WCAVSEHENTK	✓
CISFR	✓
DHMK	✓
TVLPPDGPR	✓
LACVK	✓
K	✓
TSYPDCIK	✓
AISASEADAMTLDGGWVYDA GLTPNNLKPVAAEFYGSVEH PQTYYYAVAVVK	X
K	✓
GTDFQLNQLEGK	✓
K	✓
SCHTGLGR	✓
SAGWVIPIGLLFCK	✓
LSEPR	✓
SPLEK	✓
AVSSFFSGSCVPCADPVAFP K	✓
LCQLCPGCGCSSTQPFFGYV GAFK	✓
CLK	✓
DGGGDVAFVK	✓
HTTIFEVLPEK	✓
ADR	✓
DQYELLCLDNTR	✓
KPVDQYEDCYLAR	✓
IPSHAVVAR	✓
K	✓
NNGK	X
EDLIWEILK	✓
VAQEHFGK	✓
GK	✓
SK	✓
DFQLFSSPLGK	✓
DLLFK	✓
DSAFGLLR	✓
VPPR	✓
MDYR	✓
LYLGHNYVTAIR	✓
NQQEGVCPEGSIDNSPVK	✓
WCALSHLER	✓
TK	✓
CDEWSIISEGK	✓
IECESAETTEDCIEK	✓
IVNGEADAMTLDGGHAYIAGQCGLVPVMAEYYESSNCAIPSQQGIFPK	✓
GYYAVAVVK	✓
ASDTSITWNNLK	✓
GK	✓
K	✓
SCHTGVDR	✓
TAGWNIPMGMLYNR	✓
INHCK	✓
FDEFFSQGCAPGYEK	✓
CAPNNK	✓
EEYNGYTGAFR	✓
CLVEK	✓
GDVAFVK	✓
HQTVLDNTEGK	✓
NPAEWAK	✓
NLK	✓
QEDFELLCPDGTR	✓

Table 11 (*continued*)

Detected peptides in mTf standard	
KPVK	✓
DFASCHLAQAPNHVVVSR	✓
K	✓
EK	✓
AAR	✓
VK	✓
AVLTSQETLFGGSDCTGNFC LFK	✓
STTK	✓
DLLFR	✓
DDTK	X
CFVK	✓
LPEGTTPEK	✓
YLGAEYMQSVGNMR	✓
K	✓
CSTSR	✓
LLEACTFHK	✓
H	✓
Total number of amino acids	678
Number of amino acids detected	618
Coverage (%)	91

Table 12
Normalized peak area and %RSD of Tf glycopeptide glycoforms detected in the spots of spleen protein extracts subjected to 2D electrophoretic separation and in-gel tryptic digestion.

Glycoforms	Spot 532		Spot 533		Spot 536		Spot 537		Spot 539	
	A_{norm}*	%RSD	A_{norm}*	%RSD	A_{norm}*	%RSD	A_{norm}**	%RSD	A_{norm}*	%RSD
2Ant/1NeuGc	7.7	6.1	–	–	–	–	17.0	18.1	15.4	6.4
2Ant/1NeuGc1Fuc	5.5	1.2	–	–	–	–	–	–	–	–
2Ant/2NeuGc	54.7	3.7	56.9	7.1	54.9	7.9	76.5	5.5	137.8	0.8
2Ant/2NeuGc1Fuc	29.8	6.7	25.9	7.5	–	–	16.2	20.7	26.1	17.1
2Ant/3NeuGc	17.7	5.5	–	–	–	–	29.8	15.6	8.7	4.9
2Ant/3NeuGc1Fuc	2.6	3.3	–	–	–	–	–	–	–	–

* A_{norm}: normalized peak areas were calculated as: (Glycoform peak area/peptide 354–364 (CDEWSIISEGK) peak area) × 100.

2.7. μLC–TOF–MS

The μLC–TOF–MS experiments were performed in a 1200 series capillary liquid chromatography system coupled to a 6220 oa-TOF mass spectrometer with an orthogonal G1385–44300 interface (Agilent Technologies). LC and MS control, separation, data acquisition and processing were performed using MassHunter workstation software (Agilent Technologies). The oa-TOF mass spectrometer was tuned and calibrated following the manufacturer's instructions. Once a day, or even twice a day when required, a “Quick Tune” of the instrument was carried out in positive mode followed by a mass-axis calibration to ensure accurate mass assignments. In order to enhance detection sensitivity of glycopeptides, no internal recalibration was used [11]. MS measurement parameters were as described in a previous work [12]: capillary voltage 4000 V, drying gas (N_2) temperature 200 °C, drying gasflow rate 4 L min^{-1},nebulizer gas (N_2) 15 psig, fragmentor voltage 215 V, skimmer voltage 60 V, OCT 1 RF Vpp voltage 300 V. Data were collected in profile (continuum) at 1 spectrum s^{-1} (approx. 10,000 transients/spectrum) between m/z 100 and 3200, working in the highest resolution mode (4 GHz). For separation, a Zorbax 300SB-C18 column (3.5 m particle diameter, 300 Å pore

diameter, 150 mM × 0.3 mm LT × id, Agilent Technologies) was used. Experiments were performed at room temperature with gradient elution at a flow rate of 4 μL min^{-1}. Eluting solvents were A: water with 0.1% (v/v) formic acid, and B: acetonitrile with 0.1% (v/v) formic acid. Solvents were degassed for 10 min by sonication before use. The optimum elution program was: solvent B from 10% to 60% (v/v) within 45 min as linear gradient, followed by cleaning and re-equilibration steps of B: 60% to 100% (v/v) (5 min), 100% (v/v) (10 min), 100% to 10% (v/v) (5 min) and 10% (v/v) (10 min). Before analysis, samples were filtered using a 0.22 μm polyvinylidene difluoride centrifugal filter (Ultrafree-MC, Millipore, Bedford, MA, USA) at 12,000 rpm for 4 min. Sample injection was performed with an autosampler refrigerated at 4 °C and the injection volume was 1 μL when analyzing Tf isolated from serum samples and digested with trypsin, and 5 μL when analyzing Tf in-gel digests.

2.8. μLC–TOF–MS data analysis

Prior to data analysis, a database with the exact monoisotopic mass of the different glycopeptide glycoforms of mouse Tf was created using Excel. To calculate the monoisotopic mass of each glycopeptide glycoform, it was necessary to calculate the elemental composition of all the glycopeptides taking into account the peptide and glycan contribution. First, the peptide sequence of mouse Tf was obtained from UniProt Knowledgebase (Q921I1), which also includes information about which cysteines and asparagines are involved in disulfide bonds and in *N*-glycosylation points, respectively. Afterwards, the theoretical sequence of each peptide and glycopeptide that would be obtained after tryptic digestion is obtained using the proteomic tool *PeptideMass* from the Expasy bioinformatics resource program. Subsequently, using the *ProtParam* tool from Expasy the elemental composition of the peptide sequence of the glycopeptide is obtained. Furthermore, the elemental composition of each glycan is calculated as the sum of the elemental composition of each monosaccharide that forms the glycan. *Ion source webpage* was used to obtain the elemental composition of each monosaccharide. Finally, the elemental composition of the peptide is added to obtain the molecular formula of each possible glycopeptides glycoform and thus, the monoisotopic mass with four decimals. Afterwards, the mass-to-charge values (m/z) for each glycopeptide glycoform are calculated up to a z value of 5 considering proton adducts (i.e. $[M+H]^{+}$, $[M+2H]^{2+}$, $[M+3H]^{3+}$, $[M+4H]^{4+}$ and $[M+5H]^{5+}$)

Finally, the data analysis is carried out using the software MassHunter Qualitative (Agilent Technologies). All the previously calculated m/z values for each glycopeptide glycoform are extracted together to obtain an extracted ion chromatogram (EIC) of that glycopeptide specie, as can be observed in Fig. 1, which shows the EIC for some glycopeptide glycoforms in three different samples. If more than one of the extracted masses is detected in one chromatographic peak of the EIC, the presence of the corresponding glycopeptide glycoforms can be confirmed.

Tables 1 and 2 can be found in the online version of this article (.xlsx files). They show the list of protein species identified by MS/MS, displaying the sequence of matched and fragmented peptides of a given protein. Ion scores and confidence intervals of the fragmented peptides are also shown.

Acknowledgments

We thank to Dr. Frances E. Lund (Department of Microbiology, University of Alabama at Birmingham, Birmingham, Alabama, USA) for the gift of the CD38-/- mice and helpful discussions; to Pilar Navarro-Cuesta for her technical assistance.

The proteomic analyses were performed in the proteomics facilities of IPBLN-CSIC (Granada) and UCO-SCAI (Córdoba). Both facilities belong to ProteoRed, PRB2-ISCIII, supported by grant PT13/0001.

This work was supported in part by the European Commission in collaboration with the following Funding agencies:

Ministerio de Economía y Competitividad (MINECO) del Gobierno de España Grant number SAF2011-27261 (awarded to J.S.), Grant number CTQ2011-27130 (awarded to V. S-N), Grant number SAF2011-22463 (to R.M.; cofunded by the European Regional Development Fund), and Grant number SAF2012-34059 (to J.M.; cofunded by the European Regional Development Fund); Consejería de

Innovación, Ciencia y Empresa de la Junta de Andalucía, Grant number P08-CTS-04046 (awarded to J. S.). From the MINECO del Gobierno de España, Grant number IPT2011-1527-010000, associated to Fibrostatin SL (awarded to J.M). A.R.V. was supported by a felllowship-contract from Consejería de Innovación, Ciencia y Empresa de la Junta de Andalucía. S.G.R. was supported by a JAEDoc contract from CSIC and a MINECO contract. A.B. was supported by a FPU fellowship from the Ministry of Education, Culture and Sport.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.dib.2015.12.045.

References

[1] A. Rosal-Vela, et al., Identification of multiple transferrin species in 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 (2015), http://dx.doi.org/10.1016/j.jprot.2015.11.023.
[2] J.A. Vizcaino, et al., ProteomeXchange provides globally coordinated proteomics data submission and dissemination, Nat. Biotechnol. 32 (3) (2014) 223–226.
[3] D.A. Cockayne, et al., Mice deficient for the ecto-nicotinamide adenine dinucleotide glycohydrolase CD38 exhibit altered humoral immune responses, Blood 92 (4) (1998) 1324–1333.
[4] J.J. Inglis, et al., 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 (5) (2007) R113.
[5] J. Postigo, et al., Mice deficient in cd38 develop an attenuated form of collagen type II-induced arthritis, PLoS One 7 (3) (2012) e33534.
[6] A. Rosal-Vela, et al., Distinct serum proteome profiles associated with collagen-induced arthritis and complete Freund's adjuvant-induced inflammation in CD38(-/-) mice: the discriminative power of protein species or proteoforms, Proteomics 15 (19) (2015) 3382–3393.
[7] E.J. Pavon, et al., Proteomic analysis of plasma from patients with systemic lupus erythematosus: increased presence of haptoglobin alpha2 polypeptide chains over the alpha1 isoforms, Proteomics 6 (1) (2006) S282–S292.
[8] E.J. Pavon, et al., Increased expression and phosphorylation of the two S100A9 isoforms in mononuclear cells from patients with systemic lupus erythematosus: a proteomic signature for circulating low-density granulocytes, J. Proteom. 75 (6) (2012) 1778–1791.
[9] E.R. Vossenaar, et al., PAD, a growing family of citrullinating enzymes: genes, features and involvement in disease, Bioessays 25 (11) (2003) 1106–1118.
[10] V. Goeb, et al., Candidate autoantigens identified by mass spectrometry in early rheumatoid arthritis are chaperones and citrullinated glycolytic enzymes, Arthritis Res. Ther. 11 (2) (2009) R38.
[11] E. Gimenez, et al., Capillary electrophoresis time-of-flight mass spectrometry for a confident elucidation of a glycopeptide map of recombinant human erythropoietin, Rapid Commun. Mass Spectrom. 25 (16) (2011) 2307–2316.
[12] A. Barroso, et al., 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.

Chapter 3. Glycosylation study of glycoproteins by capillary electrophoresis mass spectrometry

Even though the analysis of the glycosylation of proteins by **CapLC-MS** provided good results and ended up in three publications, there was still one small drawback, the separation of the different **glycoforms**, at the **glycopeptide** level, was not possible with the conventional stationary phase used in these works. Glycoform separation can provide substantial advantages such as improved sensitivity and increased reliability in the identification of the detected glycoforms. In this regard, **CE-MS** can be considered an alluring complementary technique to CapLC-MS for the separation and characterization of glycoprotein digests following a typical bottom-up strategy. CE-MS is especially interesting in glycosylation studies largely due to the fact that glycopeptide glycoforms that differ in the number of sialic acids can be separated in a relative easy manner. Additionally, CE-MS presents other remarked characteristics, such as minimal sample and reagent consumption, high selectivity, resolution and efficiency, short analysis times, ease of use and potential for automation.

However, certain glycoproteins, such as hTf, have tightly folded structures, mainly due to compacted tertiary or quaternary structures with high number of intraprotein connections such as disulfide or hydrogen bonds. These proteins are **proteolytic resistant**, and in order to obtain proper digestion yields, a **surfactant** is mandatory. However, these surfactants might interfere in the analysis by CE-MS. Hence, the complete removal of these surfactants might be necessary to properly use CE-MS for the analysis and characterization of protein glycosylation and, in this way, establish an alternative and complementary methodology to CapLC-MS.

Another great advantage of choosing CE as separation technique is that the electrophoretic migration behavior of ions can be easily predicted using the **classical semiempirical relationships** between electrophoretic mobility (m_e) and charge-to-mass ratio (q/M). Prediction of electrophoretic migration behavior is an appealing modelization tool to speed-up method development in CE-MS, as well as to refine the structural assignments based on the measured molecular mass (M).

The work carried out in both aforementioned topics has led to two new publications, which are listed below:

- **Publication 3.1:** 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. Albert Barroso, Estela Giménez, Fernando Benavente, José Barbosa, Victoria Sanz-Nebot. *Electrophoresis* 2016, **37**, 987-997.
- **Publication 3.2:** Modelling the electrophoretic migration behaviour of peptides and glycopeptides from glycoprotein digests in capillary electrophoresis-mass spectrometry. Albert Barroso, Estela Giménez, Fernando Benavente, José Barbosa, Victoria Sanz-Nebot. *Analytica Chimica Acta* 2015, **854**, 169-177.

Electrophoresis 2016, *37*, 987–997

Albert Barroso
Estela Giménez
Fernando Benavente
José Barbosa
Victoria Sanz-Nebot

Department of Analytical Chemistry, University of Barcelona, Barcelona, Spain

Received June 1, 2015
Revised July 21, 2015
Accepted August 7, 2015

Research Article

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

Certain glycoproteins are rather difficult to digest due to compacted tertiary or quaternary structures. In a previous study, a capillary LC coupled to TOF-MS (μLC-TOF-MS) method was developed for the detection and characterization of the glycopeptide glycoforms of human transferrin (Tf), a proteolytic resistant glycoprotein, in serum samples. After immunoaffinity purification, Tf was digested with trypsin in the presence of RapiGest® and μLC-TOF-MS analyses permitted to detect the N_{413} and N_{611} glycopeptide glycoforms. Conversely, the use of this surfactant, albeit mandatory to quantitatively digest the isolated Tf, proved detrimental to CE-TOF-MS analysis due to its interaction with the inner surface of the silica capillary walls. As CE is usually regarded as an interesting alternative to other separation techniques (low consumption of reagents, excellent separation efficiency, and reduced analysis times), in this work, the undesirable interferences of the surfactant have been removed to allow the correct separation and detection of Tf glycoforms by CE-TOF-MS. Moreover, the digestion protocol described by the RapiGest® manufacturer has been modified to minimize desialylation of Tf glycopeptide glycoforms. The new developed CE-TOF-MS methodology has been then compared with the former μLC-TOF-MS by means of sensitivity and separation efficiency of Tf glycopeptide glycoforms in the standard glycoprotein. Additionally, Tf glycopeptide glycoforms from serum of healthy volunteers and patients with congenital disorders of glycosylation have also been analyzed following the developed methodology.

Keywords:
Capillary electrophoresis / Glycoforms / Glycopeptides / Mass spectrometry / Transferrin
DOI 10.1002/elps.201500255

1 Introduction

MS is nowadays the most reliable way to obtain structural information about protein glycosylation as it allows fast and high sensitivity profiling and accurate characterization of heterogeneous glycan structures [1–6]. Glycoproteomic studies are usually performed by the analysis of the glycopeptides or glycans after enzymatic digestion or by chemical release [2,4,7–9]. However, in contrast to glycans, glycopeptides obtained after the glycoprotein enzymatic digestion are a desirable alternative as they not only provide information about the structure and composition of the oligosaccharides, but also about glycosylation sites and their degree of occupancy.

Capillary LC coupled to MS detection (μLC-MS) is widely used for the detection and characterization of peptides, glycopeptides, or glycans. C8 and C18 stationary phases have been extensively used for some years in glycoproteomics. However, due to their inability to separate different glycopeptide glycoforms [10] and the recent technological advances regarding LC stationary phases, they have been relegated to a second place, especially in the glycomics field. Other

Correspondence: Estela Giménez, Department of Analytical Chemistry, University of Barcelona, Diagonal 645, 08028 Barcelona, Spain
E-mail: estelagimenez@ub.edu
Fax: +34-934021233

Abbreviations: bAGP, bovine α_1-acid-glycoprotein; **CDG**, congenital disorders of glycosylation; **EIE**, extracted ion electropherogram; **IAC**, immunoaffinity chromatography; **PGC**, porous graphite carbon; **Tf**, human transferrin; **ZIC-HILIC**, zwitterionic-hydrophilic interaction LC

Colour Online: See the article online to view Figs. 1–5 in colour.

stationary phases such as zwitterionic-hydrophilic interaction LC (ZIC-HILIC) or porous graphite carbon (PGC) have aroused great interest [11–14]. Notwithstanding their efficiency in glycan analysis, when glycopeptides are to be analysed both stationary phases show certain limitations. ZIC-HILIC analyses usually require mobile phases with high ionic strength, which might lead to hindered ionization efficiency in ESI and, thus, reduced sensitivity. Additionally, glycopeptides strongly interact with this type of stationary phase which may result in larger analysis times [2, 12]. As far as PGC is concern, it has been reported that this stationary phase can be useful in the analysis of smaller glycopeptides, but glycopeptides with larger peptide fraction are not detected or their sensitivity is considerably lower [7]. Consequently, CE-MS has become an attractive alternative to LC-MS for separation and characterization of protein digests in the typical bottom-up strategies applied in proteomic studies [15–18]. Additionally, the recent development of novel nano-ESI sheathless interfaces is revitalizing the interest in CE-MS as a high-separation efficiency [17–20]. Several authors have demonstrated the excellent performance of CE-MS for the analysis of glycopeptides from glycoprotein digests [20–25], largely due to the fact that the different glycopeptide glycoforms are easily separated in CE-MS, especially if the number of sialic acids is different. Moreover, the reagent consumption is greatly reduced and the analysis time is pretty low.

Some proteins are quite difficult to digest due to compact tertiary or quaternary structures with high number of intraprotein connections, such as disulfide or hydrogen bonds, leading to tightly folded proteins. To increase the digestion yield of these proteins, alternative methods such as denaturalization or microwave assisted digestion, have been proposed by some authors [26–29]. However, in the case of glycoproteins, the effectiveness of the enzymatic digestion could be even more compounded due to the steric hindrance of the attached glycans [28]. Recently, the use of acid-labile anionic surfactants, such as Rapigest®, has raised great interest in proteomic studies due to its high efficiency when digesting proteolytic resistant proteins and its compatibility with MS detection [10, 30–32].

Human transferrin (Tf), an iron-binding serum transport glycoprotein, is one of such compact glycoproteins, as it has more than 650 amino acids and about 20 disulfide bonds. When this compacted glycoprotein is to be digested, the use of a surfactant is mandatory to correctly denaturalize the glycoprotein and obtain quantitative digestions [10]. However, as it has been reported in a previous work, when Rapigest® is used, even though it is hydrolyzed at the end of the digestion, surfactant residuals seem to interact with the inner surface of the silica capillary walls, and thus, altering the electrophoretic separation and detection by CE-MS [10].

In this work, an improved methodology has been developed to allow the correct separation and detection of Tf glycoforms by CE-TOF-MS, removing undesirable interferences of the surfactant. Moreover, the digestion protocol described by the Rapigest® manufacturer has been modified to minimize desialylation of Tf glycopeptide glycoforms. Finally, the developed CE-TOF-MS methodology has been applied to the analysis of Tf glycopeptides in serum control and in serum samples from patients with congenital disorders of glycosylation (CDG).

2 Materials and methods

2.1 Chemicals

All chemicals used in the preparation of buffers and solutions were of analytical reagent grade. Isopropanol (iPrOH), hydrochloric acid (HCl, glacial), acetic acid (HAc, glacial), formic acid (HFor, 98–100%), glycine (≥99.7%) and TFA (≥99%) were supplied by Merck (Darmstadt, Germany). CNBr-activated-Sepharose 4B was from GE Healthcare (Waukesha, WI, USA). Sodium Chloride (NaCl, ≥99.5%), DTT (≥99%), iodoacetamide (IAA), ammonium hydrogen carbonate, water (LC-MS grade) and acetonitrile (LC-MS grade) and bovine α_1-acid-glycoprotein (bAGP) were supplied by Sigma-Aldrich (Madrid, Spain). Tris (≥99.5%) was supplied by J.T. Baker (Deventer, Holland). Trypsin (sequencing grade modified) was provided by Promega (Madison, WI, USA). RapiGest® from Waters (Bedford, MA, USA) was used to enhance the enzymatic digestion. ProteaseMAX™ surfactant was supplied by Promega. Goat polyclonal antibody against Tf (immunogen affinity purified) was purchased from Abcam (Cambridge, UK). ESI low concentration tuning mix was supplied by Agilent Technologies (Waldbronn, Germany) for tuning and calibration of the oa-TOF mass spectrometer.

2.2 Tf samples

Tf standard was purchased from Sigma-Aldrich. Tf standards were prepared at 1500 μg/mL (~19 μM) in 50 mM NH_4HCO_3 buffer (pH 7.9).

Commercial available serum samples with a non-pathological profile (control, SERODOS® plus) were purchased from Bio-Rad (Hercules, CA, USA). Serum control samples from healthy volunteers were obtained in the laboratory [33]. Briefly, venous blood was collected in 9 mL Vacuette tubes (Greiner Bio-One, Frickenhausen, Germany) with Z serum separation clot activator, and then allowed to coagulate by leaving it undisturbed at room temperature for 9 h. Afterwards, the clot was kept at 4°C for 12–16 h to improve the clot retraction. The supernatant serum was subsequently separated from the clot with a Pasteur pipette and centrifuged at 1200 × *g* for 20 min at 4°C. Clear serum was then separated and aliquoted to store in a freezer at −20°C when not in use. Serum samples with CDG-type I and CDG-type II pathological profile were kindly provided by the Balagué Center S.A., the Institute of Clinical Biochemistry (Hospital Clínic, University of Barcelona) and the Center for the Diagnosis of Molecular Diseases (Autonomous University of Madrid).

Tf standards and serum samples previously purified by immunoaffinity chromatography (IAC) were reduced, alkylated, and immediately subjected to enzymatic digestion in the presence of RapiGest® as explained in [10]. Briefly,

1.90 μL of 0.5 M DTT in 50 mM NH_4HCO_3 buffer was added to an aliquot of 50 μL of Tf 1500 μg/mL with 0.1% w/v RapiGest®. The mixture was incubated in a TS-100 Thermo-Shaker digester (Bio-San, Riga, Latvia) at 56°C for 30 min and then alkylated with 50 mM IAA for 30 min at room temperature in the dark (5.25 μL of 0.73 M IAA). Excess reagent was removed by ultracentrifugation with Microcon YM-10 (MW cut-off 10 kDa, Millipore, Bedford, MA, USA), washing three times with NH_4HCO_3 buffer. The final residue was recovered from the upper reservoir by being centrifuged upside down in a new vial, and reconstituted to the initial volume (50 μL) with NH_4HCO_3 buffer with 0.1% RapiGest®. Trypsin in an enzyme to sample ratio of 1:40 by mass was added and the mixture was carefully vortexed and incubated overnight at 37°C. In this work, the hydrolyzation of the surfactant was performed after the overnight incubation as follows: formic acid (HFor) was added to the digest to a final concentration of 5% v/v and the mixture was incubated in the digester at 37°C for 30 min. Then, the solution was centrifuged for 10 min at 12 000 rpm to separate the RapiGest® residues. A residual fraction of RapiGest® still remained in the solution but it was completely compatible with LC-MS analysis as indicated by the manufacturer. Finally, the supernatant was carefully collected and stored at −20°C until its analysis. pH measurements were performed with a Crison 2002 potentiometer and a Crison electrode 52-03 (Crison Instruments, Barcelona, Spain). Centrifugation procedures were carried out in a Mikro 20 centrifuge (Hettich, Tuttlingen, Germany) at room temperature.

2.3 Purification of Tf by IAC

In order to isolate Tf from the rest of serum proteins, an immunoaffinity purification was carried out using a cyanogen-bromide sepharose column prepared as explained in a previous work [10]. Briefly, the immunoaffinity procedure consisted in: 50 μL of serum sample were diluted 1/4 in 10 mM Tris-HCl (pH 7.6) in order to improve antigen–antibody interaction, and consequently, to maximize Tf recoveries. Before passing the sample through the column, a conditioning step was carried out consisting in two washes with 10 mM Tris-HCl. Afterwards, the diluted serum was passed through the column ten times. After washing with 10 mM Tris-HCl and 0.5 M NaCl (pH 7.6) to eliminate the non-specifically retained fraction (washing fraction), retained Tf was eluted with 100 mM glycine-HCl (pH 2.5). Eluted Tf was immediately neutralized with 0.5 M Tris. Afterwards, glycine-HCl buffer was exchanged for 50 mM NH_4HCO_3 buffer by ultracentrifugation using Microcon YM-10 (Millipore). Finally, Tf was digested as explained in Section 2.2.

2.4 RapiGest® elimination by desalting with μElution plates

In order to properly eliminate the residual RapiGest®, after the hydrolyzation of the surfactant, an additional desalting step was carried out using Oasis® HLB 96-well μElution plates from Waters S.A. (Bedford, MA, USA) following the manufacturer instructions [34]. Briefly, the μElution cartridges were conditioned with 200 μL of acetonitrile and further equilibrated with 200 μL of 0.1% TFA solution. Afterwards, the digest (typical volumes of 50 μL) was mixed with 0.1% TFA solution to a final volume of 200 μL. The resulting solution was passed through the cartridges at less than 1 mL/min to prevent the breakthrough of peptides and glycopeptides. Two additional washing steps were performed, first with 800 μL of 0.1% TFA solution to remove salts and later with 200 μL of purified water to remove excess buffer and salts. Finally, to elute the retained peptides and glycopeptides 50 μL of 70% ACN solution were added to the cartridges. The solvent was then evaporated in a vacuum concentrator (SpeedVac™, Thermo Fisher Scientific, Waltham, MA, USA) and the sample was reconstituted in the same volume as prior to the desalting procedure with 50 mM NH_4HCO_3 pH 7.9.

2.5 μLC-TOF-MS

The μLC-TOF-MS experiments were performed in the 1200 series capillary LC system coupled to a 6220 oa-TOF LC/MS mass spectrometer with an orthogonal G1385-44300 interface (Agilent Technologies). LC and MS control, separation, data acquisition, and processing were performed with the MassHunter workstation software (Agilent Technologies). The oa-TOF mass spectrometer was tuned and calibrated following the manufacturer's instructions. Once a day, or even twice a day if working days were too long, a 'Quick Tune' of the instrument was carried out in positive mode followed by a mass-axis calibration to ensure accurate mass assignments. MS measurement parameters were as indicated: capillary voltage 4000 V, drying gas (N_2) temperature 200°C, drying gas flow rate 4 L/min, nebulizer gas (N_2) 15 psig, fragmentor voltage 215 V, skimmer voltage 60 V, 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 working in the highest resolution mode (4 GHz).

For the separation of the tryptic digests, a Zorbax 300SB-C18 column (3.5 μm particle diameter, 300Å pore diameter, 150 × 0.3 mm L_T x id, Agilent Technologies) was used. Experiments were performed at room temperature with gradient elution at a flow rate of 4 μL/min. Eluting solvents were A: water with 0.1% v/v of HFor, and B: acetonitrile with 0.1% v/v HFor. Solvents were degassed for 10 min by sonication before use. The optimum elution program was: solvent B from 10 to 60% v/v within 45 min as linear gradient, followed by cleaning and re-equilibration steps of B: 60 to 100% v/v (5 min), 100% v/v (10 min), 100 to 10% v/v (5 min) and 10% v/v (10 min). Before analysis, samples were filtered using a 0.22 μm polyvinyldene difluoride centrifugal filter (Ultrafree-MC, Millipore) at 12 000 rpm for 4 min. Sample injection was performed with an autosampler refrigerated at 4°C and the injection volume was 0.15 μL.

2.6 CE-TOF-MS

The CE-TOF-MS experiments were performed in a 7100 CE system coupled to a 6220 oa-TOF LC/MS mass spectrometer with an orthogonal G1603A sheath-flow interface (Agilent Technologies). The sheath liquid was delivered at a flow rate of 3.3 μL/min by a KD Scientific 100 series infusion pump (Holliston, MA, USA). CE control was performed using ChemStation software running in combination with MassHunter workstation software (both from Agilent Technologies) for control, data acquisition, and processing of the oa-TOF mass spectrometer. The oa-TOF was calibrated as explained in Section 2.4. Instrument parameters were the same as those used for μLC-TOF-MS, except for the nebulizer gas (N_2), which in this case was 7 psig.

A bare fused-silica capillary of 70 cm total length (L_T) × 75 μm id × 375 μm od (Polymicro) supplied by Composite Metals Service (Worcester, England) was used for CE-TOF-MS analyses. New capillaries were activated with 30 min of 1 M NaOH, water and BGE (50 mM HAc and 50 mM HFor, pH 2.3). Capillaries were conditioned every day by rinsing for 5 min with NaOH, 7 min with water and for 10 min with BGE. Activation and conditioning procedures were performed offline in order to avoid NaOH entering the mass spectrometer. Electrophoretic separations were carried out at 25°C under normal polarity (18 kV). Between runs, the capillary was conditioned for 1 min with water, 3 min with 1 M HAc, 1 min with water and 5 min with BGE. A sheath liquid of iPrOH:H_2O (60:40, v/v) with 0.05% v/v HFor was employed, degassing for 10 min by sonication before use. Injection was performed hydrodynamically at 50 mbar for 15 s.

3 Results and discussion

3.1 Hydrolysis of Rapigest® surfactant

In a previous work, a μLC-TOF-MS method was developed for the separation and detection of Tf glycopeptides [10]. The enzymatic digestion was improved by using an acidic-labile anionic surfactant called Rapigest® to increase the digestion yield of Tf. This surfactant is described to be compatible with LC-MS analyses after hydrolysis with 0.5% of TFA and incubation at 37°C for 45 min [32]. After this treatment the non-volatile by-product precipitates and thus, it can be easily removed so that only a sulphonate salt (the other by-product) compatible with MS detection remains in solution. However, under these conditions, we observed that the peak area of the 2Ant/1SiA glycoform was too high for a Tf standard, and that non-expected glycoforms, such as 2Ant/0SiA, were detected, which could be explained by the desialylation of other glycoforms, in this case, the 2Ant/2SiA. In the present work, to minimize desialylation, softer conditions were evaluated with regard to the acid used in the hydrolysis, its concentration and incubation time. The following experimental conditions were tested: 0.5% TFA for 30 min and 45 min at 37°C, 0.1, 0.5 and 5% of HFor for 45 min at 37°C and, 5% of HFor for 30 min at 37°C. Figure 1 shows the extracted ion chromatograms of the 2Ant/1SiA and 2Ant/2SiA glycoforms for the N_{413} glycopeptide under different acidic conditions and incubation times. The peak area of the 2Ant/2SiA glycoform increased while the peak area of the 2Ant/1SiA decreased when using 5% HFor instead of 0.5% TFA, which means that the desialylation of the major glycoform 2Ant/2SiA was reduced when using a weaker acid such as HFor. Moreover, as can be observed in Fig. 1, when the concentration of HFor was reduced from 5 to 0.5 or 0.1%, the peak area of the 2Ant/2SiA and 2Ant/1SiA glycoforms substantially decreased, but also the signal of the rest of peptides and glycopeptides of the tryptic digest. Additionally, some glycoforms were no longer detected when the concentration of HFor was 0.5% or 0.1%. For instance, the glycoform N_{413}-3Ant/3SiA was not observed at either concentrations, and the glycoforms N_{611}-2Ant/1SiA, N_{413}-3Ant/3SiA and N_{611}-3Ant/3SiA also disappeared when a concentration of 0.1% of HFor was used. This could be explained because at low concentration of HFor, the acidic conditions were not strong enough to completely hydrolyze the surfactant and, thus, the remaining surfactant interfered with MS detection. Hence, a concentration of 5% of HFor was selected as it reduced the desialylation of certain glycopeptide glycoforms without decreasing the signal of peptides and glycopeptides. Finally, no significant differences were observed in the area of both glycopeptides using 45 or 30 min of incubation. Therefore, we selected 5% of HFor for 30 min of incubation at 37°C as the optimum conditions to hydrolyze Rapigest® surfactant.

3.2 Elimination of Rapigest® surfactant

The μLC-TOF-MS method permitted to detect the different Tf glycopeptide glycoforms. However, the method still failed when trying to separate the glycoforms of each glycopeptide, due to the use of a C18 column. CE has been described as an excellent analytical technique for the separation of glycopeptide glycoforms, which vary in the composition of the monosaccharides, especially if they differ in the number of sialic acids [20–25]. In the previous work, we evaluated CE-MS for the separation of Tf glycopeptide glycoforms, but the surfactant seemed to interfere with the inner surface of the silica capillary walls, provoking a distortion of the electrophoretic peaks, which were much wider than expected for a glycopeptide or a peptide [10]. Additionally, reproducibility was rather low when using this surfactant. In the present work, with the aim of establishing a CE-MS method able to separate the glycoforms of Tf glycopeptides, an alternative surfactant called ProteaseMAX™ was studied to evaluate whether the problem associated with CE separation was only due to the use of RapiGest® or was also observed when other commercial acid-labile anionic surfactants were used. An aliquot of 1500 μg/mL (~19 μM) Tf standard was digested with trypsin using ProteaseMAX™ following the manufacturer's instructions. Unfortunately, this surfactant also seemed to interfere with the capillary walls distorting the electrophoretic peaks

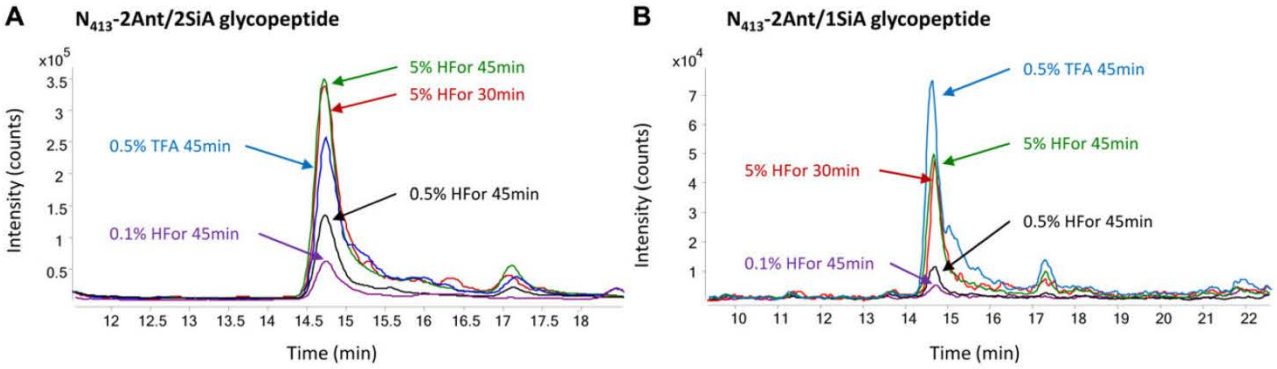


Figure 1. Extracted ion chromatograms (EICs) for the 2Ant/2SiA glycoform (A) and 2Ant/1SiA glycoform (B) for the N_{413} glycopeptide of Tf using different RapiGest® hydrolyzation conditions.

of Tf peptides and glycopeptides, fairly similar to the results observed with RapiGest® (data not shown). Finally, to prove that the alteration of the peaks was a widespread problem observed with any glycoprotein digested with RapiGest®, two additional samples were analyzed by CE-TOF-MS: a 1500 μg/mL (~36 μM) aliquot of bAGP digested with trypsin without using any surfactant and using RapiGest®. Figure 2 shows the extracted ion electropherograms (EIEs) of the N_{118}-2Ant/2SiA glycopeptide and peptide EYQTIEDK when bAGP was tryptically digested with and without RapiGest®. As bAGP is not a proteolytic resistant glycoprotein, we were able to perfectly separate and detect by CE-TOF-MS the peptides and glycopeptides of the digest without using any surfactant. However, when bAGP was digested in the presence of Rapigest® and analyzed by CE-TOF-MS, the peaks of the peptides and glycopeptides were also distorted as happened with Tf, which can be clearly observed in Fig. 2. These results confirmed that the remaining parts after decomposition of this type of surfactants alter the electrophoretic separation of the protein digest by CE-MS and that their removal is mandatory to perform the analysis by CE-MS.

To ensure the complete removal of Rapigest® residuals (i.e. the sulphonate salt), a desalting procedure using μElution plates was evaluated. Initially, a 1500 μg/mL (~19 μM) Tf standard was digested following the optimized method described in this paper, and subsequently desalted following the procedure described in Section 2.4. Concurrently, another aliquot of Tf standard was digested using RapiGest® but not subjected to the desalting procedure. Figure 3 shows the EIEs for 2Ant/2SiA glycoform of N_{611} glycopeptide in both samples, as well as the EIE obtained for the tryptic peptide 197–206 (DGAGDVAFVK) as an example. As can be stated, after desalting and removing the residual RapiGest®, the intensity and width of the peaks improved considerably. As observed in Fig. 3B for peptide 197–206, this improvement was not only seen in the glycopeptides but in the rest of the peptides of the digest as well. These results demonstrated that the initial hypothesis suggested in our previous paper [10] was correct and that the residual surfactant remaining in the digest interferes with the silica capillary walls.

Figure 4 shows the comparison between the Tf standard analysed by CE-TOF-MS and μLC-TOF-MS, bearing in mind that in both cases the optimized hydrolyzation step of the surfactant, explained in Section 3.1, was used. Figure 4A and B illustrate the separation of some N_{611} glycopeptide glycoforms by CE-TOF-MS and μLC-TOF-MS, respectively. Unlike μLC-TOF-MS, the use of CE-TOF-MS permitted to separate the different glycopeptide glycoforms, and analysis time was shorter (runs were of 30 min for CE and 75 min for μLC, and the separation of the glycopeptide glycoforms was obtained in 14 and 23 min, respectively). Moreover, despite injecting less amount of digest (83 nL in CE vs. 150 nL in μLC), the area of the glycoforms by CE-TOF-MS was slightly higher compared to those obtained by μLC-TOF-MS. This improved sensitivity could be explained because the glycoforms separated by CE-TOF-MS did not suffer any ionic suppression between them. Furthermore, the separation of the several glycoforms of a given glycopeptide is also essential to ensure that the detected glycoforms are not artefacts originated from the fragmentation in the ion source of other co-eluting glycoforms. When ion source fragmentation occurs, originated fragments co-migrate or co-elute with the parental ion. Hence, detected glycopeptide glycoforms that do not co-elute with any other glycoform of the same glycopeptide are, very likely, glycoforms from the native glycoprotein. As reversed-phase μLC is not able to separate the glycoforms of the same glycopeptide, such assumption cannot be made.

3.3 Method application to serum samples

The development of a CE-MS method able to separate and detect the glycopeptide glycoforms of Tf could solve the most impending problems of routine methods for the analysis of Tf in serum samples. These routine methods usually consist in the analysis of intact Tf by IEF or anion exchange LC with UV detection [35–37]. However, IEF is an arduous and time-consuming technique which is not suitable for automation and UV detection does not allow the reliable characterization of Tf glycoforms. Consequently, developing a methodology able not only to rapidly separate the different glycoforms but

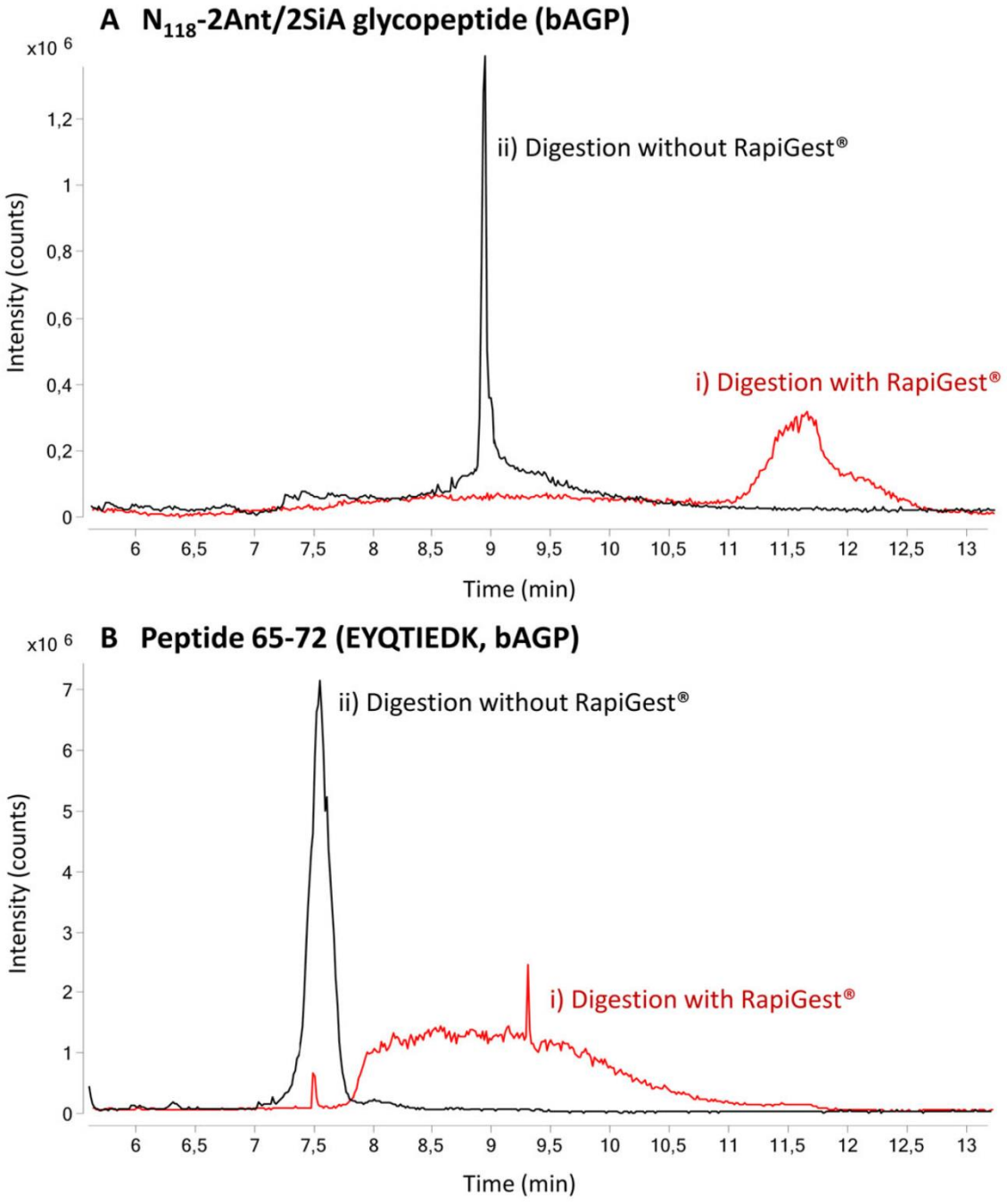


Figure 2. Extracted ion electropherograms (EIEs) for the 2Ant/2SiA glycoform of the N_{118} glycopeptide (A) and the peptide 65–72 (EYQTIEDK) (B) of bAGP digested with (i) and without (ii) RapiGest®.

also unequivocally identify them due to the use of a mass spectrometer is essential. And, certainly, as it has been extensively proved in the last years, MS is the most reliable way to obtain structural information about protein glycosylation as it allows fast and high sensitivity profiling and accurate characterization of heterogeneous glycan structures [2, 38]. Additionally, Tf natural variants may lead to erroneous conclusions when analysing Tf glycoforms by the routine methods described before due to a shift in the bands or peaks, which can be completely avoided using MS [39]. The analysis of Tf glycoforms by MS can be carried out by analysing the intact glycoprotein [38, 40] or by analysing the glycans or glycopeptides obtained by enzymatic digestion [10, 41–44]. Glycopeptide analysis shows better sensitivity compared to

Electrophoresis 2016, *37*, 987–997

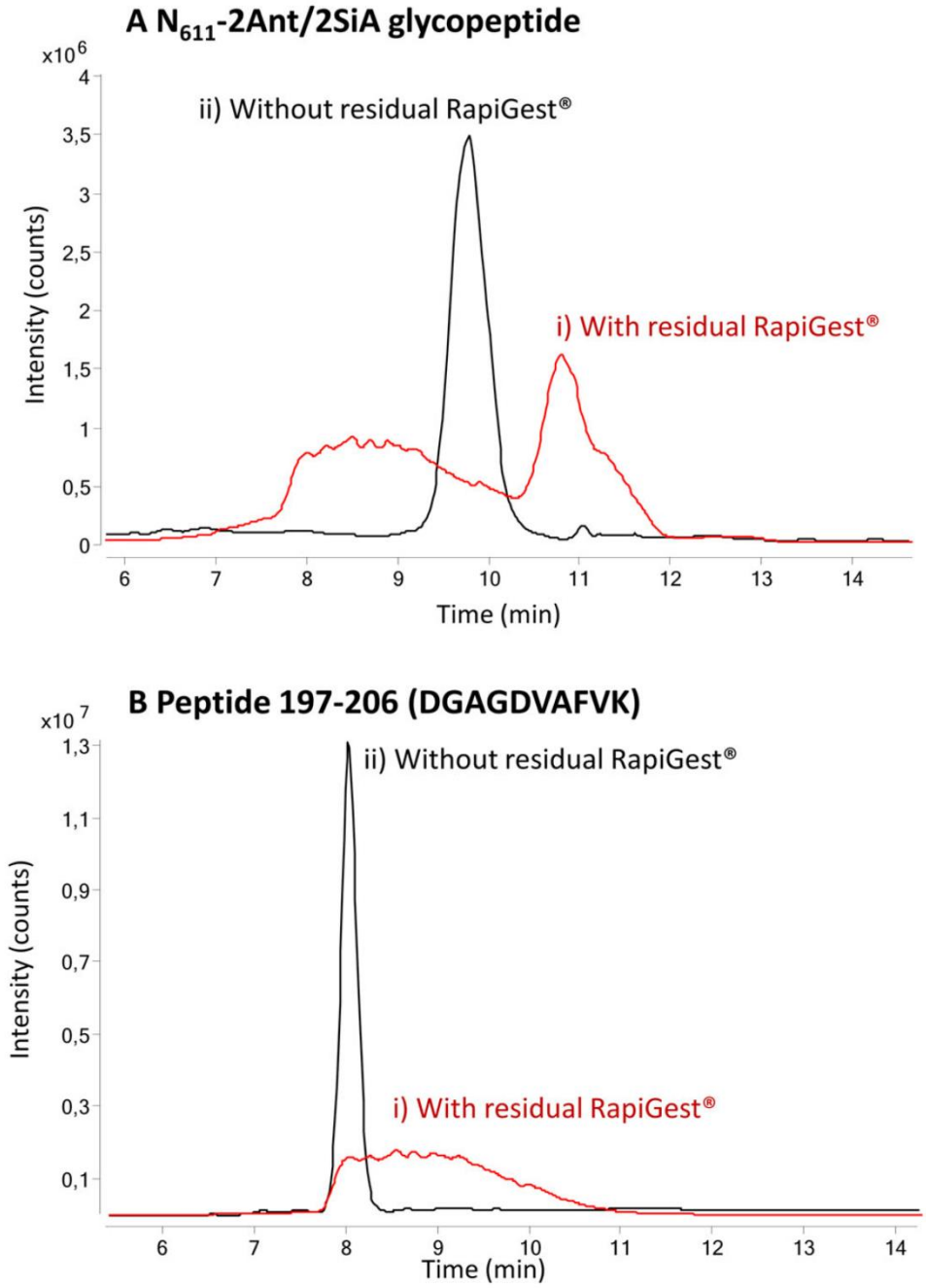


Figure 3. Extracted ion electropherograms (EIEs) for the 2Ant/2SiA glycoform of the N_{611} glycopeptide (A) and the peptide 197–206 (DGAGDVAFVK) of Tf (B) with (i) and without (ii) residual RapiGest®.

the intact Tf analysis, and it provides information about glycosylation sites and their degree of occupation in contrast to the analysis of the glycans.

In this work, after optimization of a sample treatment able to reduce desialylation of the glycopeptide glycoforms and able to remove the residual RapiGest®, which impeded the correct separation of the glycoforms by CE-TOF-MS, the established methodology was applied to the analysis of Tf glycopeptides in biological samples. Serum samples from a healthy volunteer and individuals with CDGs (type I and type II) were subjected to immunopurification by IAC to isolate Tf from the rest of serum components. Then, Tf was digested in the presence of RapiGest®, and the resulting digest was desalted with the μElution plates and subsequently analyzed by CE-TOF-MS. Table 1 shows the main glycoforms of both N_{611} and N_{413} glycopeptides detected in Tf standard and different serum samples (healthy control and pathological samples), as well as their normalized peak area (A_{norm}) and %RSD values. As can be observed, despite serum sample complexity, the healthy control gave similar results to those obtained with Tf standard. A_{norm} values were very similar for all glycopeptide glycoforms, as both samples (Tf standard and healthy control)

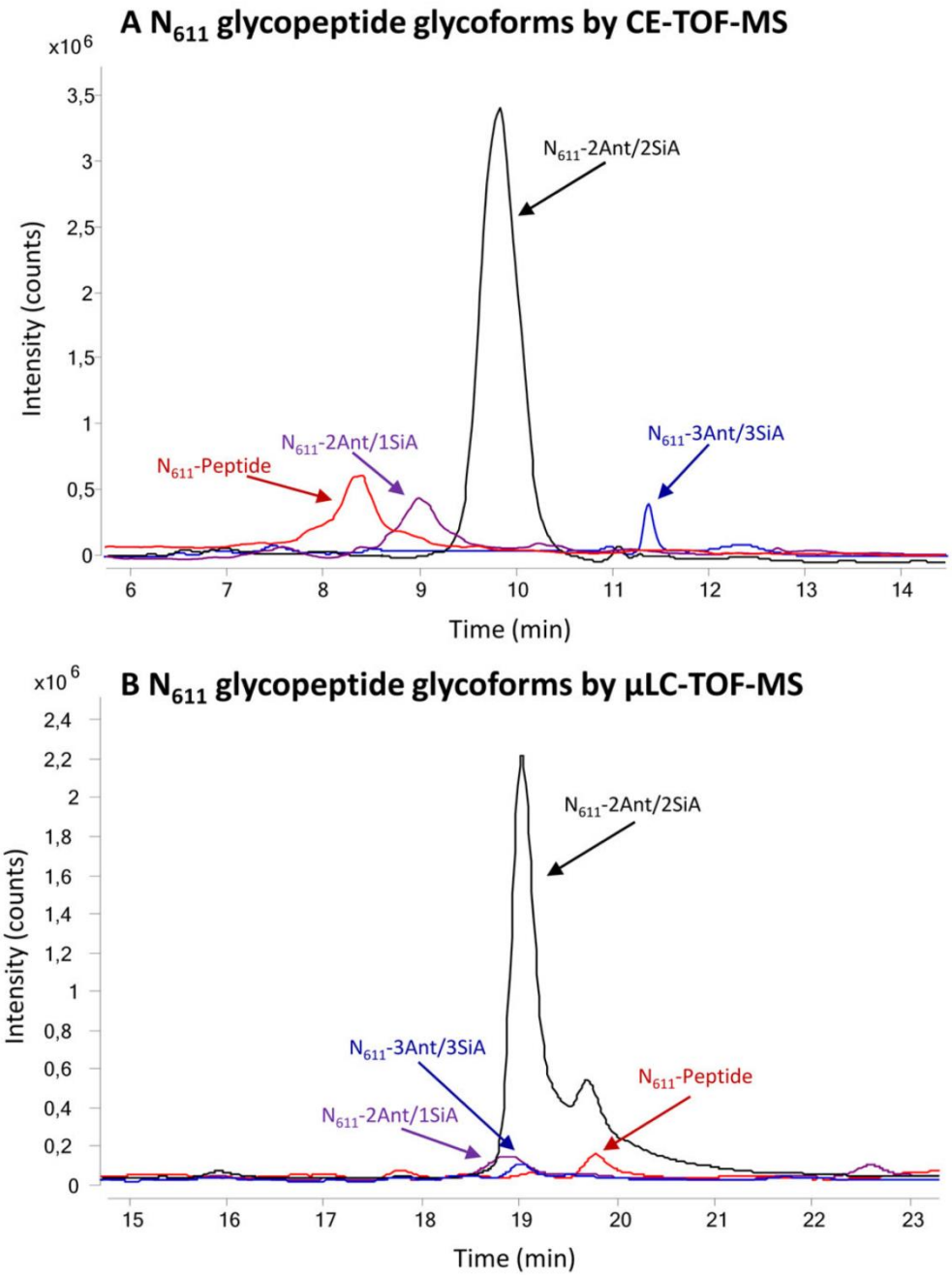


Figure 4. Extracted ion electropherograms (EIEs) and extracted ion chromatograms (EICs) of the main glycoforms for the N_{611} glycopeptide of Tf by (A) CE-TOF-MS and (B) µLC-TOF-MS, respectively.

showed the same glycosylation pattern (similar composition of Tf intact glycoforms). Regarding CDG serum samples, a decrease in 2Ant/2SiA glycoforms for both glycopeptides was observed in both types of CDGs (type I and type II). Moreover, in CDG type I samples (PMM2 and MPI), an increase in the peptide glycoform for both glycopeptides was clearly observed while the other glycoforms remained almost unaltered, besides the 2Ant/2SiA. On the other hand, 2Ant/1SiA glycoform increased for both glycopeptides in CDG type II sample, while, again, the other glycoforms were barely modified. These glycosylation patterns were in concordance with the expected alteration of the intact glycoforms of Tf in CDG samples, as lack of glycosylation is expected in CDG type I (i.e. increased peptide glycoforms), and carbohydrate chains are truncated in CDG type II (i.e. increased 2Ant/1SiA glycoforms).

To further demonstrate the ability of the developed methodology to separate and detect by CE-MS Tf glycopeptide glycoforms, three of the main N_{611} glycopeptide glycoforms for the serum control and the three CDG samples are shown in Fig. 5. As can be observed, in all four samples the three glycoforms were separated as they were in the

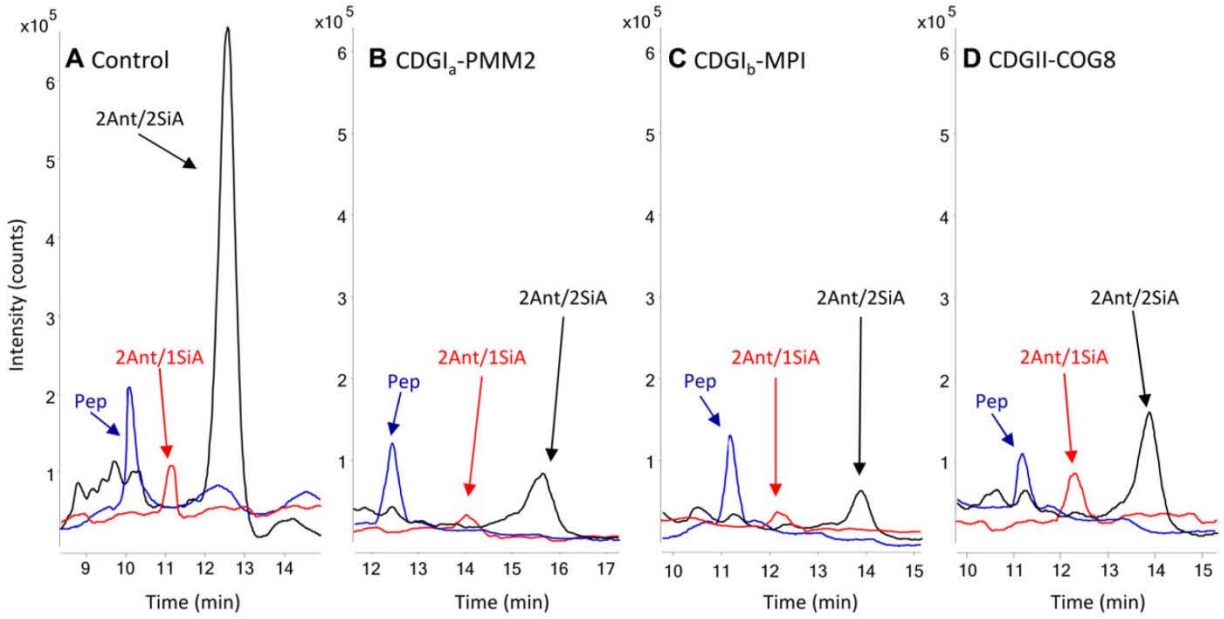


Figure 5. Extracted ion electropherograms (EIEs) for the three main glycoforms of the N_{611} glycopeptide in serum samples: (A) control serum, (B) CDGI$_a$-PMM2, (C) CDGI$_b$-MPI and (D) CDGII-COG8.

Table 1. Normalized peak area and RSD of the main glycoforms for both Tf N-glycopeptides detected in standard, serum control and three CDG serum samples

Glycoforms		Samples									
		Standard		*Serum control*		*CDGI$_a$-PMM2*[a)]		*CDGI$_b$-MPI*[b)]		*CDGII-COG8*[c)]	
		A_{norm} (%)	RSD (%)	A_{norm} (%)	RSD (%)	A_{norm} (%)	RSD (%)	A_{norm} (%)	RSD (%)	A_{norm} (%)	RSD (%)
N_{413}	2Ant/2SiA	53.0	3.2	60.2	1.9	25.8	4.6	25.1	3.2	28.4	0.5
	3Ant/3SiA	7.3	4.3	7.1	7.5	4.0	6.1	3.5	6.5	5.4	8.1
	2Ant/1SiA	5.0	9.4	3.5	8.0	2.5	7.8	1.7	1.5	13.4	6.3
	2Ant/2SiA1Fuc	1.7	2.4	2.2	2.0	3.3	7.5	2.5	5.0	2.8	2.0
	Peptide	0.0	0.0	0.0	0.0	77.3	1.2	76.5	4.5	0.0	0.0
N_{611}	2Ant/2SiA	75.9	1.0	83.3	3.1	31.9	9.1	23.2	1.5	36.0	7.4
	3Ant/3SiA	2.3	2.7	3.0	1.7	2.3	4.3	1.8	6.6	1.8	7.9
	2Ant/1SiA	5.4	9.1	6.3	9.0	4.4	6.1	4.1	7.4	15.6	10.2
	2Ant/2SiA1Fuc	7.6	3.7	6.9	2.9	6.4	4.9	6.4	2.0	6.7	3.4
	Peptide	13.0	7.0	13.0	3.7	31.6	6.1	46.3	1.1	9.9	5.3

Normalized peak areas were calculated as: (Glycoform peak area/peptide 125–143 peak area) × 100.
a) Phosphomannomutase 2 official gene name.
b) Mannose phosphate isomerase official gene name.
c) Component of oligomeric golgi complex 8 official gene name.

standard. Additionally, the electropherogram obtained for the serum control is fairly similar to the electropherogram of the Tf standard, bearing in mind that less sensitivity is expected when analysing serum samples, as the recoveries after the immunopurification process were around 40% [10]. Regarding the CDG samples, it can be clearly observed that the relative abundance of the 2Ant/2SiA glycoform decreases substantially in all of them. However, a direct comparison of the peak areas would lead to misleading conclusions as the recoveries after the IAC purification extremely depends on the amount of Tf present in the serum sample, and this amount varies considerably between samples (e.g. children vs. adults). In order to avoid these differences in Tf concentration, the areas of each glycoform are calculated as normalized areas with respect to the area of a peptide (125–143) of the digest, as shown in Table 1, thus, allowing the comparison of the different samples. These results confirmed that the established methodology can be applied to real samples, and thus, the analysis of Tf by CE-TOF-MS could be regarded as a potential alternative to LC analysis techniques since similar results are obtained, but reduced amounts of reagents and sample are required, analysis times are considerably shorter and, above

all, glycopeptides with larger peptide fraction (such is the case of Tf glycopeptides) can be easily detected and separated.

4 Concluding remarks

In this work, a novel methodology has been developed to successfully digest and analyse by CE-MS the glycopeptide glycoforms of proteolytic resistant glycoproteins, which usually require a surfactant to obtain proper digestion yields. In particular, Tf, as an example of a compacted glycoprotein, was digested with trypsin in the presence of an acid-labile anionic surfactant called RapiGest®, and the resulting glycopeptides were analysed by CE-TOF-MS. Desialylation of the glycopeptide glycoforms was reduced using 5% of HFor and no peak distortion was observed as residual RapiGest® was removed with a desalting procedure using μElution plates. The established methodology was further validated with the analysis of Tf glycopeptides in serum samples from a healthy control and patients with CDGs. This methodology could be regarded as an excellent alternative to other separation techniques such as LC with polar stationary phases, e.g. HILIC or PGC, even when glycoforms of larger glycopeptides are analysed.

Part of this study was supported by the Spanish Ministry of Economy and Competitiveness (CTQ2011-27130 and CTQ2014-56777-R). Albert Barroso thanks the Ministry of Education, Culture and Sport for a FPU fellowship. The authors acknowledge Balagué Center S.A., Institute of Clinical Biochemistry (Hospital Clínic, University of Barcelona) and Centre for the Diagnosis of Molecular Diseases (Autonomous University of Madrid) for kindly providing the serum samples from healthy individuals and CDG patients. The authors also thank Marta Burrull from Waters for her help regarding the elimination of RapiGest® surfactant.

The authors have declared no conflict of interest.

5 References

[1] Alley, W. Jr., Novotny, M., *Annu. Rev. Anal. Chem.* 2013, *6*, 237–265.

[2] Wuhrer, M., *Glycoconj. J.* 2013, *30*, 11–22.

[3] Levery, S. B., Steentoft, C., Halim, A., Narimatsu, Y., Clausen, H., Vakhrushev, S. Y., *Biochim. Biophys. Acta - Gen. Subj.* 2015, *1850*, 33–42.

[4] Kailemia, M. J., Ruhaak, L. R., Lebrilla, C. B., Amster, I. J., *Anal. Chem.* 2014, *86*, 196–212.

[5] Sturiale, L., Barone, R., Garozzo, D., *J. Inherit. Metab. Dis.* 2011, *34*, 891–899.

[6] Zhang, Y., Jiao, J., Yang, P., Lu, H., *Clin. Proteomics* 2014, *11*, 18.

[7] Alley, W. R., Jr., Mechref, Y., Novotny, M. V., *Rapid Commun. Mass Spectrom.* 2009, *23*, 495–505.

[8] Novotny, M., Alley, W. R., Jr., Mann, B. F., *Glycoconj. J.* 2013, *30*, 89–117.

[9] Mariño, K., Bones, J., Kattla, J. J., Rudd, P. M., *Nat. Chem. Biol.* 2010, *6*, 713–723.

[10] Barroso, A., Giménez, E., Benavente, F., Barbosa, J., Sanz-Nebot, V., *Anal. Chim. Acta* 2013, *804*, 167–175.

[11] Ruhaak, L. R., Deelder, A. M., Wuhrer, M., *Anal. Bioanal. Chem.* 2009, *394*, 163–174.

[12] Zauner, G., Deelder, A., Wuhrer, M., *Electrophoresis* 2011, *32*, 3456–3466.

[13] Stavenhagen, K., Kolarich, D., Wuhrer, M., *Chromatographia* 2014, *78*, 307–320.

[14] Giménez, E., Balmaña, M., Figueras, J., Fort, E., Bolós, C., Sanz-Nebot, V., Peracaula, R., Rizzi, A., *Anal. Chim. Acta* 2015, *866*, 59–68.

[15] Kašička, V., *Electrophoresis* 2012, *33*, 48–73.

[16] Righetti, P. G., Sebastiano, R., Citterio, A., *Proteomics* 2013, *13*, 325–340.

[17] Wang, Y., Fonslow, B. R., Wong, C. C. L., Nakorchevsky, A., Yates, J. R., *Anal. Chem.* 2012, *84*, 8505–8513.

[18] Sun, L., Zhu, G., Yan, X., Champion, M. M., Dovichi, N. J., *Proteomics* 2014, *14*, 622–628.

[19] Haselberg, R., de Jong, G. J., Somsen, G. W., *Electrophoresis* 2013, *34*, 99–112.

[20] Zamfir, A. D., Flangea, C., Serb, A., Zagrean, A.-M., Rizzi, A. M., Sisu, E., *Methods Mol. Biol.* 2013, *951*, 145–169.

[21] Amon, S., Zamfir, A. D., Rizzi, A., *Electrophoresis* 2008, *29*, 2485–2507.

[22] Giménez, E., Ramos-Hernan, R., Benavente, F., Barbosa, J., Sanz-Nebot, V., *Rapid Commun. Mass Spectrom.* 2011, *25*, 2307–2316.

[23] Giménez, E., Ramos-Hernan, R., Benavente, F., Barbosa, J., Sanz-Nebot, V., *Anal. Chim. Acta* 2012, *709*, 81–90.

[24] Zamfir, A. D., Peter-Katalinić, J., *Electrophoresis* 2004, *25*, 1949–1963.

[25] Zaia, J., *Methods Mol. Biol.* 2013, *984*, 13–25.

[26] Janecki, D. J., Reilly, J. P., *Rapid Commun. Mass Spectrom.* 2005, *19*, 1268–1272.

[27] Liu, X., Chan, K., Chu, I. K., Li, J., *Carbohydr. Res.* 2008, *343*, 2870–2877.

[28] Segu, Z. M., Hammad, L. A., Mechref, Y., *Rapid Commun. Mass Spectrom.* 2010, *24*, 3461–3468.

[29] Zhang, Y., Fonslow, B. R., Shan, B., Baek, M.-C., Yates, J. R. I., *Chem. Rev.* 2013, *113*, 2343–2394.

[30] Huang, H. Z., Nichols, A., Liu, D., *Anal. Chem.* 2009, *81*, 1686–1692.

[31] Min, L., Choe, L. H., Lee, K. H., *Electrophoresis* 2015, *36*, 1690–1698.

[32] WATERS S.A. Protocol: Desalting of Peptide Mixture Prior To ESI-MS Analysis. S.A.

[33] Pont, L., Benavente, F., Barbosa, J., Sanz-Nebot, V., *Electrophoresis* 2015, *36*, 1265–1273.

[34] WATERS S.A. RapiGest Sf Surfactant Care and Use Manual.

[35] Mohamed, M., Guillard, M., Wortmann, S. B., Cirak, S., Marklova, E., Michelakakis, H., Korsch, E., Adamowicz, M., Koletzko, B., van Spronsen, F. J., Niezen-Koning, K. E., Matthijs, G., Gardeitchik, T., Kouwenberg, D., Lim, B. C., Zeevaert, R., Wevers, R. A., Lefeber, D. J., Morava, E., *Biochim. Biophys. Acta* 2011, *1812*, 691–698.

[36] Jaeken, J., *Ann. N. Y. Acad. Sci.* 2010, *1214*, 190–198.

[37] Marklová, E., Albahri, Z., *J. Clin. Lab. Anal.* 2009, *23*, 77–81.

[38] Sanz-Nebot, V., Balaguer, E., Benavente, F., Neusüss, C., Barbosa, J., *Electrophoresis* 2007, *28*, 1949–1957.

[39] Zühlsdorf, A., Park, J. H., Wada, Y., Rust, S., Reunert, J., DuChesne, I., Grüneberg, M., Marquardt, T., *Clin. Biochem.* 2015, *48*, 11–13.

[40] Del Castillo Busto, M. E., Montes-Bayón, M., Blanco-González, E., Meija, J., Sanz-Medel, A., *Anal. Chem.* 2005, *77*, 5615–5621.

[41] Xia, B., Zhang, W., Li, X., Jiang, R., Harper, T., Liu, R., Cummings, R., He, M., *Anal. Biochem.* 2013, *442*, 178–185.

[42] Pérez, B., Medrano, C., Ecay, M. J., Ruiz-Sala, P., Martínez-Pardo, M., Ugarte, M., Pérez-Cerdá, C., *J. Inherit. Metab. Dis.* 2013, *36*, 535–542.

[43] Landberg, E., Åström, E., Kågedal, B., Påhlsson, P., *Clin. Chim. Acta* 2012, *414*, 58–64.

[44] Heywood, W., Mills, P., Grunewald, S., Worthington, V., Jaeken, J., Carreno, G., Lemonde, H., Clayton, P., Mills, K., *J. Proteome Res.* 2013, *12*, 3471–3479.

Analytica Chimica Acta 854 (2015) 169–177

Contents lists available at ScienceDirect

Analytica Chimica Acta

journal homepage: www.elsevier.com/locate/aca

Modelling the electrophoretic migration behaviour of peptides and glycopeptides from glycoprotein digests in capillary electrophoresis-mass spectrometry

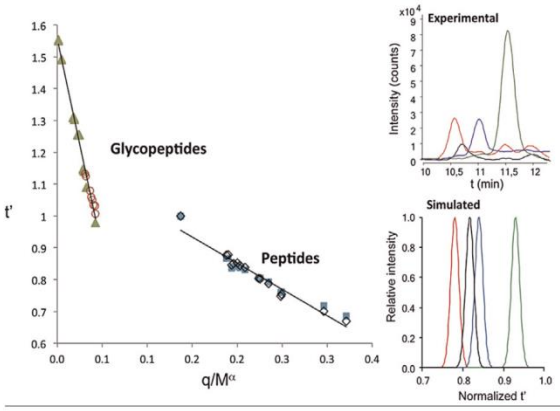
Albert Barroso, Estela Giménez, Fernando Benavente, José Barbosa, Victoria Sanz-Nebot *

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

HIGHLIGHTS

- Semiempirical models are used to predict rhEPO peptide–glycopeptide migration times.
- The Stoke's law ($q/M^{1/3}$) result in better linear correlations for rhEPO peptides.
- rhEPO glycopeptides fit better with the classical polymer model ($q/M^{1/2}$).
- These models permit to simulate rhEPO peptide and glycopeptide CE separations.
- Potential applicability to estimate electrophoretic maps of other digested glycoproteins.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 7 July 2014
Received in revised form 8 October 2014
Accepted 26 October 2014
Available online 29 October 2014

Keywords:
Capillary electrophoresis
Mass spectrometry
Migration prediction
Glycopeptides
rhEPO

ABSTRACT

In this study, the classical semiempirical relationships between the electrophoretic mobility and the charge-to-mass ratio (m_e vs. q/M^{α}) were used to model the migration behaviour of peptides and glycopeptides originated from the digestion of recombinant human erythropoietin (rhEPO), a biologically and therapeutically relevant glycoprotein. The Stoke's law ($\alpha = 1/3$), the classical polymer model ($\alpha = 1/2$) and the Offord's surface law ($\alpha = 2/3$) were evaluated to predict migration of peptides and glycopeptides, with and without sialic acids (SiA), in rhEPO digested with trypsin and trypsin–neuraminidase. The Stoke's law resulted in better correlations for the set of peptides used to evaluate the models, while glycopeptides fitted better with the classical polymer model. Once predicted migration times with both models, it was easy to simulate their separation electropherogram. Results were later validated predicting migration and simulating separation of a different set of rhEPO glycopeptides and also human transferrin (Tf) peptides and glycopeptides. The excellent agreement between the experimental and the simulated electropherograms with rhEPO and Tf digests confirmed the potential applicability of this simple strategy to predict, in general, the peptide–glycopeptide electrophoretic map of any digested glycoprotein.

* Corresponding author. Tel.: +34 934039083; fax: +34 934021233.
E-mail address: vsanz@ub.edu (V. Sanz-Nebot).

http://dx.doi.org/10.1016/j.aca.2014.10.038
0003-2670/

1. Introduction

Capillary electrophoresis mass spectrometry (CE–MS) is nowadays a mature technique for the analysis of peptides and proteins [1–8], that has been widely explored as an alternative to liquid chromatography mass spectrometry (LC–MS) for separation and characterization of protein digests in the typical bottom-up strategies applied in proteomic analysis [5–10]. The recent development of novel nanoESI sheathless interfaces is revitalising the interest in CE–MS as a high-separation efficiency and sensitive tool to achieve high sequence coverage of minute diluted protein digests [3,7,8,11]. Furthermore, several authors have demonstrated the excellent performance of CE–MS for the analysis of glycopeptides from glycoprotein digests [11–15].

Several authors have regarded prediction of electrophoretic migration behaviour of the compounds of interest, including peptides, as an excellent tool to speed-up method development in CE–MS, as well as to refine the structural assignments made on the basis of the measured molecular mass (M) [16–24]. Migration of ionisable compounds in capillary zone electrophoresis (CZE) can be easily predicted using the classical semiempirical relationships between the electrophoretic mobility and charge-to-mass ratio (m_e vs. q/M^{α}) or other models, such as those based on the Hubbard–Onsager dielectric friction theory of ion mobility [4-6,16–34]. In our previous works, the Stoke's law ($\alpha = 1/3$, peptides are modelled as spherical particles), the classical polymer model ($\alpha = 1/2$, peptides are polymers with lower q densities) and the Offord's surface law ($\alpha = 2/3$, for larger and more rigid structures, which experience frictional forces that are proportional to their surface area) yielded excellent correlations when they were employed to study the m_e of several peptide hormones [18], neuropeptides [31], quinolones [32] and metallothioneins [35], when good estimates of acidity constant values were available for charge calculations. Migration prediction of peptides from protein hydrolysates has been also explored by several authors [4-6,20–24], but those works were mainly focused in the analysis of proteins without post-translational modifications (PTMs) [20–23]. To the best of our knowledge, only J. Kim et al. evaluated before several semiempirical models with post-translationally modified peptides resulting from tryptic digestion of human myelin basic protein, taking into account citrullination, deamidation, oxidation, phosphorylation and methylation [24]. Modelling migration behaviour of glycoprotein digests, which are complex mixtures of peptides and glycopeptides [11–15], represents a further challenging task. On the one hand, we have the evident structural dissimilarities between peptides and glycopeptides. On the other hand, glycopeptides show up as a mixture of glycoforms due to the different composition of the carbohydrate chains attached to the peptide core [14,15,36–40]. Thus, while the number and type of glycans of the carbohydrate chains affects glycoform size, and hence M, the specific presence of sialic acid (SiA = *N*-acetylneuraminic acid) strongly contributes to their charge and it is critical for the electrophoretic separation resolution [14,15].

In this paper, we investigated the classical semiempirical relationships between m_e and q/M^{α} to model the migration behaviour in CE–MS of peptides and glycopeptides originated from the digestion of recombinant human erythropoietin (rhEPO), a biologically and therapeutically relevant glycoprotein [11–15]. The aim is to describe a simple strategy to easily generate 'dry-lab' peptide–glycopeptide electrophoretic maps of glycoproteins, which can be later useful to assist separation optimization and identification.

2. Materials and methods

2.1. Chemicals

All chemicals used in the preparation of buffers and solutions were of analytical reagent grade. Isopropanol (iPrOH, ≥99.9%), acetic acid (HAc, glacial), formic acid (HFor 98–100%), ammonia (25%), ammonium acetate (NH_4Ac, ≥99.99%) and sodium hydroxide (≥99%) were supplied by Merck (Darmstadt, Germany). DL-Dithiothreitol (DTT, ≥99%), iodoacetamide (IAA, ≥98%), ammonium hydrogen carbonate (≥99.9%) and human Tf standard (≥98%) were supplied by Sigma–Aldrich (Madrid, Spain). Trypsin (sequencing grade modified, 16,000 U mg^{-1}) from Promega (Madison, WI, USA) and neuraminidase (sialidase, 100 U mg^{-1}) were obtained from Roche (Mannheim, Germany). Water with a conductivity lower than 0.05 mS cm^{-1} was obtained using a Milli-Q water purification system from Millipore (Molsheim, France). ESI Low Concentration (ESI-L) tuning mix was supplied by Agilent Technologies (Waldbronn, Germany) for tuning and calibration of the TOF mass spectrometer.

2.2. Protein samples

rhEPO produced in a Chinese hamster ovary (CHO) cell line was provided by the European Pharmacopoeia as a Biological Reference Product (BRP-lot3). Each sample vial contained 250 μg of EPO (a mixture of epoetin alpha and beta), 0.1 mg of Tween 20, 30 mg of trehalose, 3 mg of arginine, 4.5 mg NaCl, and 3.5 mg of Na_2HPO_4. The content of each vial was dissolved in water to obtain a 1000 mg L^{-1} solution of rhEPO. Excipients of low molecular mass were removed from the rhEPO sample by passage through a Millipore Microcon YM-10 centrifugal filter (molecular weight cut-off, MWCO, 10 kDa) [14,15]. All the following centrifugation steps were performed for 10 min at 13,000 × g unless otherwise indicated. First, the filter was washed with water before loading the sample. After sample filtration, the sample residue was washed three times with an appropriate volume of water. The final residue was recovered from the upper reservoir by upside-down centrifugation in a new vial (3 min at 1000 × g). Finally, sufficient water was added to adjust the rhEPO concentration to 1000 mg L^{-1}.

rhEPO was reduced, alkylated and immediately subjected to enzymatic digestion [14,15]. Briefly, 2.5 μL of 0.5 M DTT in 50 mM NH_4HCO_3 (pH 7.9) were added to an aliquot of 100 μL of the filtered 1000 mg L^{-1} rhEPO solution. The mixture was incubated in a water bath at 56 °C for 30 min and then alkylated in 50 mM IAA for 30 min at room temperature in the dark (7 μL of 0.73 M IAA were added). Excess of low molecular mass reagents was removed with Microcon YM-10 centrifugal filters as explained before. Tryptic digestion (rhEPO-T digest) [14,15]: the final residue was reconstituted in 100 μL of 50 mM NH_4HCO_3 (pH 7.9). Trypsin was added in an enzyme to sample ratio of 1:40 w/w, the mixture was carefully vortexed and later incubated at 37 °C in a water bath for 18 h. Digestion was stopped by heating for 5 min in boiling water and stored at −20 °C until its use [14,15]. Neuraminidase digestion (rhEPO-TN digests) [14,15]: SiA residues were released from rhEPO tryptic glycopeptides by enzymatic digestion with neuraminidase [15]. Once rhEPO sample was subjected to tryptic digestion, sample was evaporated to dryness with air and reconstituted with 100 μL of 50 mM NH_4Ac (pH 5.0). Subsequently, 1 μL of neuraminidase (50 mU) was added, and solution was incubated at 37 °C for 18 h. Digestion was stopped by heating for 5 min in boiling water and stored at −20 °C until its use [15]. T- and TN-digests were obtained in different days.

Tryptic digests of human Tf (Tf-T) were obtained following the procedure described above for the tryptic digestion of rhEPO.

2.3. Apparatus and procedures

pH measurements were performed with a Crison 2002 potentiometer and a Crison electrode 52-03 (Crison Instruments, Barcelona, Spain). Centrifugation procedures were carried out in a Mikro 20 centrifuge (Hettich, Tuttlingen, Germany) at room temperature.

The CE–TOF-MS experiments were performed in a HP3DCE system coupled to a 6220 oaTOF LC–MS mass spectrometer with an orthogonal G1603A sheath-flow interface (Agilent Technologies)

Table 1
Theoretical *M*,*t*′, *q*, carbohydrate content and carbohydrate composition of the peptides and glycopeptides of rhEPO-T and rhEPO-TN digests used to evaluate the classical semiempirical relationships.

N	Peptides	M_{theo} (Da)	*q*	*t*′ Mean	s (*n* = 3)	Carbohydrate fraction (%)[a]	Carbohydrate composition
1	APPR	439.2543	1.8882	0.759	0.005	–	–
2	LICDSR	762.3694	1.8288	0.845	0.003	–	–
3	VLER	515.3067	1.8819	0.792	0.004	–	–
4	YLLEAK	735.4167	1.8819	0.832	0.003	–	–
5	VNFYAWK	926.4650	1.8882	0.836	0.003	–	–
6	MEVGQQAVEVWQGLALLSEAVLR[c]	2525.3311	1.8694	1.00	–	–	–
7	AVSGLR	601.3547	1.8882	0.806	0.005	–	–
8	SLTTLLR	802.4912	1.8882	0.837	0.004	–	–
9	ALGAQK	586.3438	1.8882	0.805	0.004	–	–
10	TITADTFR	923.4712	1.8288	0.867	0.003	–	–
11	LFR	434.2641	1.8882	0.761	0.005	–	–
12	VYSNFLR	897.4708	1.8882	0.841	0.004	–	–
13	GK	203.1270	1.8882	0.686	0.003	–	–
14	LK	259.1896	1.8882	0.718	0.005	–	–
15	LYTGEACR	968.4385	1.8819	0.867	0.004	–	–
	N-glycopeptides in rhEPO-T						
N	N_{83} (GQALLVNSSQPWEPLQLHVDK)[b]						
16	3Ant/2SiA	5074.1958	2.1547	1.14	0.003	53.5	6Hex5HexNAc2SiA1Fuc
17	3Ant/3SiA	5365.2911	1.8209	1.26	0.009	56.0	6Hex5HexNAc3SiA1Fuc
18	4Ant/2SiA	5439.3280	2.1547	1.14	0.001	56.6	7Hex6HexNAc2SiA1Fuc
19	4Ant/3SiA	5730.4234	1.8209	1.26	0.009	58.8	7Hex6HexNAc3SiA1Fuc
20	4Ant/4SiA	6021.5188	1.4870	1.30	0.01	60.8	7Hex6HexNAc4SiA1Fuc
21	4Ant1LacNAc/2SiA	5804.4602	2.1547	1.15	0.003	59.4	8Hex7HexNAc2SiA1Fuc
22	4Ant1LacNAc/3SiA	6095.5556	1.8209	1.26	0.009	61.3	8Hex7HexNAc3SiA1Fuc
23	4Ant1LacNAc/4SiA	6386.6510	1.4870	1.31	0.01	63.1	8Hex7HexNAc4SiA1Fuc
24	4Ant2LacNAc/3SiA	6460.6878	1.8209	1.26	0.009	63.5	9Hex8HexNAc3SiA1Fuc
25	4Ant2LacNAc/4SiA	6751.7832	1.4870	1.31	0.01	65.1	9Hex8HexNAc4SiA1Fuc
26	4Ant3LacNAc/4SiA	7116.9154	1.4870	1.31	0.06	66.9	10Hex9HexNAc4SiA1Fuc
N	N_{24}–N_{38} (EAENITTGCAEHCSLNENITVPDTK)[b]						
27	G3–7SiA	9837.7194	0.4666	1.49	0.02	71.5	14Hex12HexNAc7SiA2Fuc
28	G3–8SiA	10128.8149	0.1327	1.56	0.02	72.3	14Hex12HexNAc8SiA2Fuc
29	G4–7SiA	10202.8516	0.4666	1.49	0.02	72.5	15Hex13HexNAc7SiA2Fuc
30	G4–8SiA	10493.9471	0.1327	1.55	0.02	73.3	15Hex13HexNAc8SiA2Fuc
	N-glycopeptides in rhEPO-TN						
N	N_{83} (GQALLVNSSQPWEPLQLHVDK)[b]						
31	3Ant/0SiA	4492.0048	2.8224	1.03	0.0006	47.5	6Hex5HexNAc1Fuc
32	4Ant/0SiA	4857.1370	2.8224	1.04	0.001	51.4	7Hex6HexNAc1Fuc
33	4Ant1LacNAc/0SiA	5222.2692	2.8224	1.05	0.0008	54.8	8Hex7HexNAc1Fuc
34	4Ant2LacNAc/0SiA	5587.4014	2.8224	1.06	0.002	57.8	9Hex8HexNAc1Fuc
35	4Ant3LacNAc/0SiA	5952.5336	2.8224	1.08	0.0005	60.4	10Hex9HexNAc1Fuc
N	N_{24}–N_{38} (EAENITTGCAEHCSLNENITVPDTK)[b]						
36	G3–0SiA	7800.0515	2.8036	1.12	0.001	64.1	14Hex12HexNAc2Fuc
37	G4–0SiA	8165.1837	2.8036	1.13	0.001	65.7	15Hex13HexNAc2Fuc
	O-glycopeptides in rhEPO-T						
N	O_{126} (EAISPPDAASAAPLR)[b]						
38	O126–0SiA	1829.8894	1.8226	0.980	0.003	20.0	1Hex1HexNAc
39	O126–1SiA	2120.9851	1.4887	1.09	0.002	30.9	1Hex1HexNAc1SiA
40	O126–2SiA	2412.0808	1.1548	1.25	0.009	39.3	1Hex1HexNAc2SiA
	O-glycopeptide in rhEPO-TN						
N	O_{126} (EAISPPDAASAAPLR)[b]						
41	O126–0SiA	1829.8894	1.8226	1.01	0.0006	20.0	1Hex1HexNAc

M_{theo}: theoretical mass, *q*:charge, *t*′: relative migration time, *s*: standard deviation.
[a] Carbohydrated fraction: (glycan mass/glycopeptide mass).
[b] Peptide sequence of N_{83}, N_{24}–N_{38}, O_{126}.
[c] Peptide number 6 was used as a reference (average *t* = 7.68 min).

[14,15]. The sheath liquid (50:50 (v/v) iPrOH/H_2O with 0.05% (v/v) of HFor) was delivered at a flow rate of 3.3 μL min^{-1} by a KD Scientific 100 series infusion pump (Holliston, MA, USA) and was 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 with the TOF mass spectrometer. The TOF mass spectrometer was tuned and calibrated following the manufacturer's instructions. At the beginning of a working day, a 'Quick Tune' of the instrument was carried out in positive mode followed by a mass-axis calibration to ensure accurate mass assignments. Measurement parameters were tuned for the analysis of O_{126} and N_{83}-4Ant glycopeptides (Table 1) paying special attention to the fragmentor voltage value [14,15]. The optimum parameters were as follows: capillary voltage 4000 V, drying gas (N_2) temperature 200 °C, drying gas flow rate 4 L min^{-1}, nebulizer gas (N_2) 7 psig, fragmentor voltage 190 V, skimmer voltage 60 V, OCT 1 RF Vpp voltage 300 V. Data were collected in profile (continuum) at 1 spectrum s^{-1} (approx. 10,000 transients spectrum^{-1}) 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) × 50 μm internal diameter (i.d.) × 360 μm outer diameter (o.d.) (Polymicro Technologies, Phoenix, AZ, USA) was used for CE–TOF-MS separations. Activation and conditioning procedures were performed off-line in order to avoid NaOH entering to the mass spectrometer. New capillaries were activated by flushing at 930 mbar for 30 min with 1 M NaOH, water and background electrolyte (BGE) (50 mM HAc and 50 mM HFor (pH 2.2)). Each day, capillaries were conditioned by rinsing for 5 min with 0.1 NaOH, 7 min with water and 10 min with BGE. rhEPO and Tf digests were injected for 15 s at 50 mbar (in triplicate, $n = 3$). Electrophoretic separations were carried out at 25 °C and 18 kV under normal polarity (cathode in the outlet). Between runs, capillaries were rinsed for 1 min with water, 3 min with 1 M HAc, 1 min with water and 5 min with BGE. Capillaries were stored overnight filled with water. All solutions were passed through a 0.45-mm nylon filter (MSI, Westboro, MA USA) before use.

2.4. Calculations of classical semiempirical relationships

In general, the m_e of a peptide is proportional to its q and inversely proportional to its Stoke's radius (r). The r is generally expressed in terms of M, because the volume of a molecule is proportional to its mass if the density is constant [25–27]. The classical equations describing semiempirical models are deduced from assumptions concerning the peptide shapes and the forces that they undergo during electrophoretic motion [25–27]. The general form of the equation relating m_e, M and q is as follows:

$$m_e = A\frac{q}{M^{\alpha}} \quad (1)$$

where A is a constant and, for the various semiempirical models tested in this work: $\alpha = 1/3$ for the Stoke's law, $\alpha = 1/2$ for the classical polymer model and $\alpha = 2/3$ for Offord's surface area law. In general, for peptides, α approaches 1/3 when peptides are modelled as spherical particles, that have high charge densities; α approaches 1/2 when the peptide is considered as a classical polymer with a lower charge density; and α approaches 2/3 for larger and more rigid structures, which experience frictional forces that are proportional to the surface area of the molecule during electrophoretic motion [27].

These models were investigated by CE–TOF-MS at the pH of the electrophoretic separation (pH 2.2) using peptides and glycopeptides of rhEPO-T and rhEPO-TN digests (Table 1). M was calculated from the amino acid sequence of the peptides and glycopeptides, taking also into account for the glycopeptides the M of the carbohydrate moieties with or without SiA residues (and any other post-translational modification (PTM) if it is indicated later in the text). Peptide and glycopeptide q was calculated at the pH of the electrophoretic separation using the average pK values for amino acids given by Rickard et al. [25], a pK value of 2.6 for the carboxylic acid of SiA [41], a pK value of 1.8 for the sulfonic acid of SiA (predicted with ACD/Labs software, Advanced Chemistry Development, Toronto, Canada) and the Sillero and Ribeiro expression [42], which is based on the Henderson–Hasselbalch equation.

$$q = \sum_{n=1-4}\frac{P_n}{1+10^{(\mathrm{pH}-\mathrm{pK}(P_n))}} - \sum_{n=1-6}\frac{N_n}{1+10^{\mathrm{pK}(N_n)-\mathrm{pH})}} \quad (2)$$

where P_n and N_n are the cationic (i.e. P_1 = tNH$_2$, P_2 = His P_3 = Arg and P_4 = Lys) and anionic (i.e. N_1 = tCOOH, N_2 = Asp, N_3 = Glu, N_4 = Cys, N_5 = Tyr and N_6 = SiA) ionisable groups found in the amino acids and the SiA residues, and pK (P_n) and pK (N_n) are the pK values of these groups. Unless otherwise indicated later in the text, no other ionisable groups were considered for q calculations. The accuracy of the q calculated in this way depends on the proposed sequence and the reliability of the pK values considered for the ionisable groups [18,31,32]. The estimated pK values for the individual amino acids, SiA and sulfonated SiA were used, because the real pK values in all the peptide and glycopeptides sequences were not available.

Experimental relative migration times ($t_i' = t_i/t_{\mathrm{reference}}$) were used for the correlations instead of m_e, because no electroosmotic flow marker was added to the samples [17,19,23,43]. The largest rhEPO peptide (peptide N6, Table 1), which had the most similar intensity and migration time to the glycopeptides was used as a reference in rhEPO-T and rhEPO-TN digests.

Once linearity between t' and q/M^{α} was demonstrated for the peptides and glycopeptides at pH 2.2 in rhEPO-T and rhEPO-TN digests, predicted t_i' at this pH value were used to simulate an electropherogram for the separation of the mixture, showing a pure Gaussian peak for each peptide and glycopeptide [31,44]:

$$h(t') = h_0 e^{-\frac{1}{2}\left[\frac{(t'-t_i')}{\sigma}\right]^2} \quad (3)$$

where t' is the relative time values that define the time-scale, h_0 and t_i' are the height and predicted t_i' at the peak maximum for each

Table 2
Summary of linear least squares parameters obtained from application of the classical semiempirical models to the studied peptides and glycopeptides of rhEPO-T and rhEPO-TN digests at pH 2.2 ($y = b + mx$).

	$q/M^{1/2}$			$q/M^{1/3}$			$q/M^{2/3}$		
	b (±s)	m (±s)	R^2	b (±s)	m (±s)	R^2	b (±s)	m (±s)	R^2
N-glycopeptides and *O*-glycopeptide (with and without SiA)	1.557 (±0.008)	−13.3 (±0.3)	0.9891	1.57 (±0.02)	−3.33 (±0.14)	0.9569	1.527 (±0.015)	−49 (±2)	0.9598
Peptides	1.054 (±0.015)	−3.1 (±0.2)	0.9120	1.18 (±0.02)	−1.65 (±0.08)	0.9443	0.989 (±0.015)	−6.4 (±0.5)	0.8756

s: standard deviation.

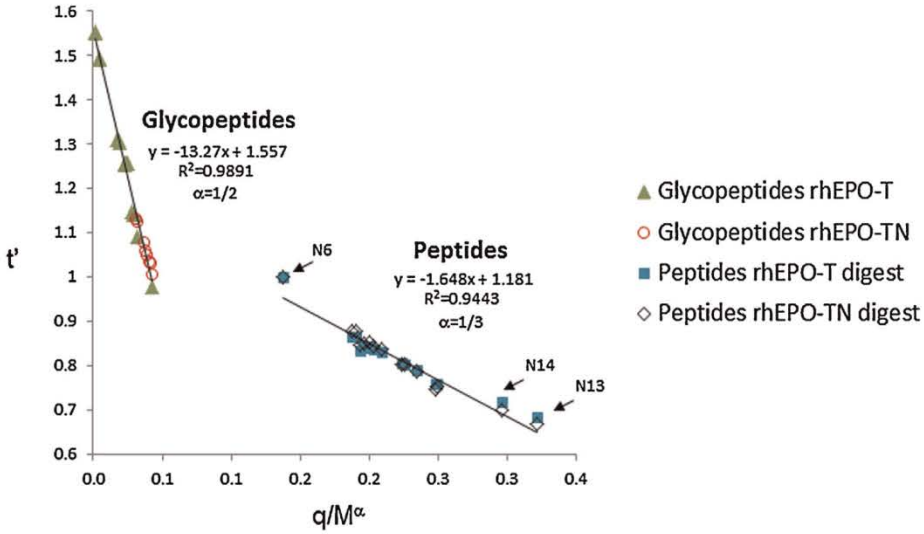


Fig. 1. Experimental (symbols) and predicted (lines) t' vs. $M/q^{1/3}$ and $M/q^{1/2}$ for peptides and glycopeptides of rhEPO-T and rhEPO-TN digests using the best linear correlations of Table 2. (BGE pH 2.2).

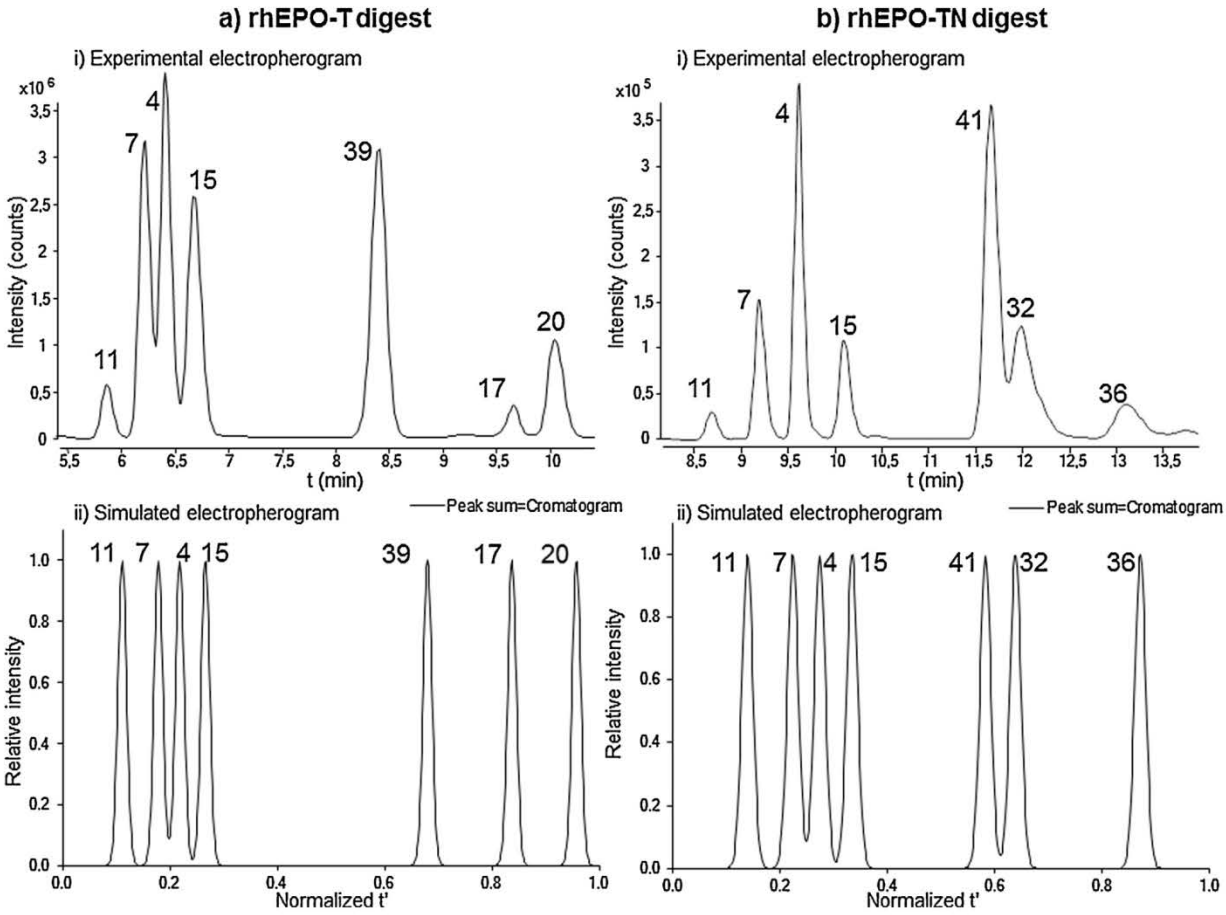


Fig. 2. (i) Experimental and (ii) simulated electropherograms of some peptides and glycopeptides of the modelling set in (a) rhEPO-T and (b) rhEPO-TN digests. (BGE pH 2.2).

peptide or glycopeptide, respectively, and σ is the standard deviation of the Gaussian peak which is related to the value of the width at half height ($w_{1/2}$) of the simulated electrophoretic peak by the following expression: $w_{1/2} = 2.354\sigma$. A value of 1 was arbitrarily selected as h_0 for all electrophoretic peaks. An average value of $w_{1/2}$ was fixed in 0.005 taking into account direct measurements of $w_{1/2}$ in the experimental electropherograms. In order to simplify the simulation procedure, no other assumptions were made regarding peak shape, BGE composition or the influence of the EOF on migration time and resolution. Microsoft Office Excel 2010 for

Windows was used for all the calculations and simulations. Other computer simulation tools (e.g. Simul 5.0) can be alternatively used for the simulations [45].

3. Results and discussion

Several authors have explored the use of different semiempirical approaches which relate m_e, M and q in order to predict electrophoretic migration behaviour, study structural modifications, charge characteristics and conformations [4-6,16–32]. However, there is no general rule to select the optimum semiempirical relationship for each experimental data set, and several modifications have been suggested for Eq. (1) (see Section 2.2), mainly aimed to obtain the best linear correlation between m_e and M/q^{α} values [4–6]. In this paper we tested the Stoke's law ($\alpha = 1/3$, Eq. (1)), the classical polymer model ($\alpha = 1/2$, Eq. (1)) and the Offord's surface area law ($\alpha = 2/3$, Eq. (1)). Table 1 shows the theoretical M, the q, the carbohydrate content and the carbohydrate composition of the peptides and glycopeptides of rhEPO-T and rhEPO-TN digests used to evaluate the semiempirical models. The whole modelling set was very heterogeneous, comprising peptides and *N*- and *O*-glycopeptides with a wide variety of M, q, carbohydrate contents and structures (with or without SiA) (Table 1). All of them were previously identified in rhEPO digests by CE–TOF-MS (with an accuracy lower than 10 ppm) [14,15]. It is important to clarify that all the peptides identified by CE–TOF-MS in those works were used in this modelling set, while only the glycopeptides with the best signal-to-noise ratios [14,15] were included. Some of the remaining identified glycopeptides were used later to validate the models. The use of experimental t' allowed us to study peptides and glycopeptides from rhEPO-T and rhEPO-TN at the same time, despite they were necessarily analysed in separate runs (see the materials and methodssection for details). The reproducibility of relative magnitudes related to migration time in CE, such as m_e or t', is widely accepted to be higher than for raw migration times [43]. This allowed a reliable run-to-run comparison.

Table 2 shows the parameters resulting from the linear correlation of the experimental t' and M/q^{α} ($\alpha = 1/3$, for the Stoke's law, $\alpha = 1/2$ for the classical polymer model, and $\alpha = 2/3$ for Offord's law) for the studied peptides and glycopeptides of rhEPO-T and rhEPO-TN digests. Linear correlations were good with all models for glycopeptides ($R^2 > 0.96$), while for peptides the Stoke's law and the classical polymer law produced significantly better results ($R^2 > 0.91$). We tried to improve these results considering the electrostatic q suppression effect in both models [26,27], as in our previous work with metallothioneins [35]. As the total q on the molecule increases, the effect of other q on its m_e decreases, resulting in an ineffectiveness of a part of the q, which can be simply addressed by using $\ln(1+q)$ instead of q in the classical semiempirical models. However, in this case, no improvement on linearity was observed (data not shown). Results were not markedly better with a model based on the Hubbard–Onsager dielectric friction theory of ion mobility (Eq. (4)), probably because with this complex group of peptides and glycopeptides hydrodynamic friction contribution predominates over dielectric friction

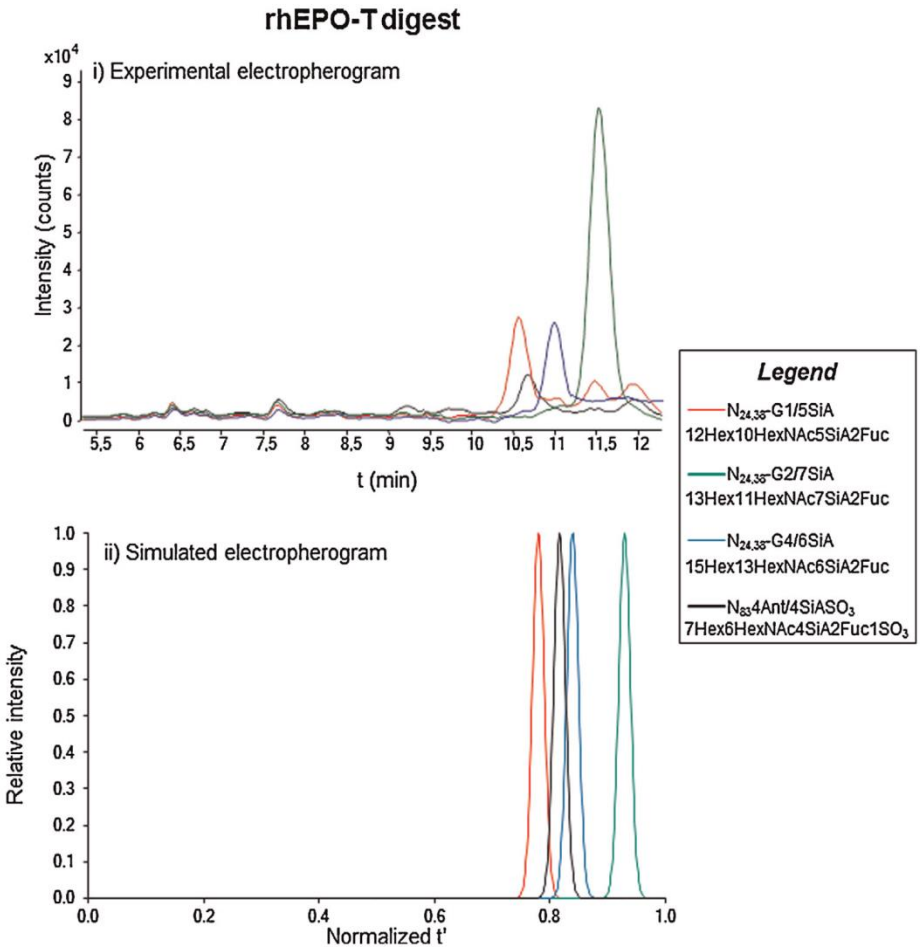


Fig. 3. (i) Experimental and (ii) simulated electropherograms of four N_{83} and N_{24-38} glycopeptides from the rhEPO-T digest that were not in the modelling set. (BGE pH 2.2).

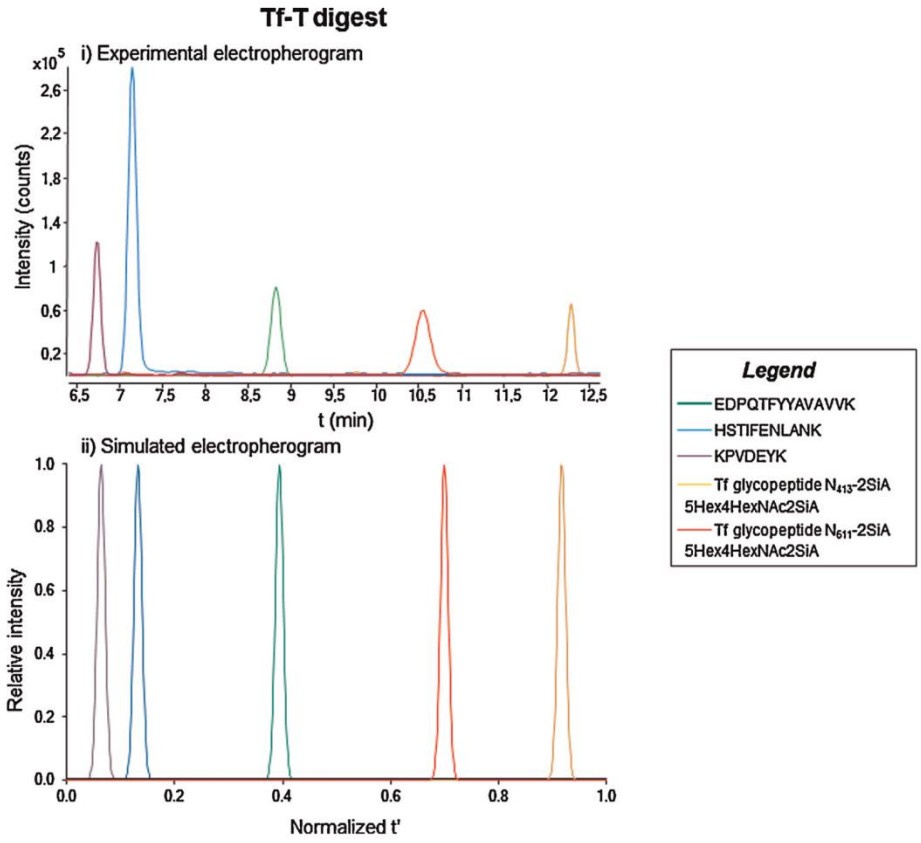


Fig. 4. (i) Experimental and (ii) simulated electropherograms of three peptides and two *N*-glycopeptides from a tryptic Tf digest (Tf-T). (BGE pH 2.2).

contribution [33,34].

$$t' = \frac{aq}{M^C + (bq^2/M)} \quad (4)$$

The linear correlations of Table 2 confirmed the validity of the classical semiempirical relationships at the separation pH when good estimations of the p*K* values were available for charge calculations [18,31,32]. However, it was difficult to draw conclusions about molecular shape based on the slight differences found between the results for the majority of the models. As can be observed from the slopes and *y*-intercepts obtained with each model, it was clear that a different linear equation was necessary for peptides and glycopeptides, probably because the larger *M* of glycopeptides promoted a significant increase of t' (i.e. a reduction of m_e). It is worth mentioning, that *N*- and *O*-glycopeptides with and without SiA, which presented a significantly different fraction of carbohydrates and *q* (Table 1), all fitted well to the same linear equation. This would reinforce the idea that differences in linear equations between peptides and glycopeptides could be a matter of the increase of *M* and the presence of carbohydrate chains (with or without SiA) rather than a matter of *q* (SiA was negatively charged to some extent at pH 2.2) [41]. The Stoke's law was selected for further investigations for peptides because regression coefficients were slightly better ($R^2 > 0.94$), while glycopeptides fitted better with the classical polymer model ($R^2 > 0.99$). In both cases, the residuals were small and randomly distributed with an average value of approximately 0, indicating a good fit for a linear model. The residuals ranged from −0.02 to 0.05 for the peptides and from −0.04 to 0.03 for the glycopeptides. Therefore, the bias in predicting relative migration times was within the range of the standard deviation values of the experimental migration times ($0.0005 \leq s \leq 0.06$, Table 1).

In our previous works with low-molecular mass peptide hormones [18], neuropeptides [31], quinolones [32] and metallothioneins [35], excellent correlations were also observed with the Stoke's and the classical polymer laws. Fig. 1 shows the plot of t' vs. $M/q^{1/3}$ and $M/q^{1/2}$ for peptides and glycopeptides, respectively, in rhEPO digests. The symbols stand for the experimental t' and the solid straight lines indicate the predicted t' values using the appropriate equations of Table 2 (see also the insets in Fig. 1). The small differences in R^2 between the linear models for peptides and glycopeptides could be explained by the higher structural heterogeneity of the set of peptides (Table 1). As can be also observed from Fig. 1, peptides were migrating at lower t' covering a wider range of q/M^{α} values but in a shorter t' window. As an example, in one of the extremes of the studied range we find the symbols corresponding to the reference peptide (Table 1, $N = 6$, $t'_6 = 1$, nominal $M = 2525$ Da, $q = 1.87$), while in the other extreme we have the dipeptides (Table 1, $N = 13$ and 14, nominal $M = 203$ and 259, $q = 1.89$ and 1.89).

The t' values of several peptides and glycopeptides from the modelling set were predicted with the linear equations corresponding to the Stoke's law and the classical polymer model, respectively, to simulate a sole electropherogram that will facilitate visualization of the separation of the tryptic mixtures. As a different model was selected for peptides and glycopeptides, it was necessary to use the predicted t' values to simulate the electropherograms instead of the predicted q/M^{α} values that were used in our previous works [31,32,35]. The x-scale of the simulated electropherograms was normalized to unit time in the investigated t' window for ease understanding ($(t'_i - t'_{min})/(t'_{max} - t'_{min})$). Fig. 2a

and b shows the experimental (i) and simulated (ii) electropherograms of these peptides and glycopeptides from rhEPO-T and rhEPO-TN digests, respectively. A separate electropherogram was simulated for rhEPO-T and rhEPO-TN digests, because the experimental electropherograms were obtained in separate experiments, as indicated before (Section 2.3). As can be observed, despite the simplicity of the approach, where each electrophoretic peak is simulated with a fixed width at half height of 0.005 and a fixed relative intensity of 1, the concordance between the experimental and the predicted electropherograms is good, with peptides and glycopeptides migrating in separate time windows of different width at the beginning and at the final part of the electropherograms. The small differences in the experimental and predicted electrophoretic profiles could be due to a combination of different factors: inaccuracy of SiA p*K* value, longitudinal diffusion and mismatched conductivities sample-BGE.

These results were further validated comparing the experimental and predicted electropherograms for a set of N_{83} and N_{24-38} glycopeptides from the rhEPO-T digest that were not in the modelling set, because they were detected by CE–TOF-MS with low signal-to-noise ratios (Fig. 3) [14,15]. As can be observed in the insets of Fig. 3, the selected glycopeptides differed in the number of sialic acids and one of them showed a sulfonic group, which affected charge calculations at the separation pH. The agreement was again good, even for the N_{83} glycopeptide with a sulfonic group (N_{83}-4Ant/4SiA-SO_3), taking into account that no glycopeptide with such functional group was used in the modelling. The sulfonic group p*K* value in sulfonated SiA was estimated using ACD/Labs software (1.3 ± 0.5), because a reliable p*K* value for sulfonic groups in free SiA or in glycopeptide carbohydrate chains was not available. In our case, the best concordance between the experimental and predicted electropherograms was obtained when a p*K* value of 1.8 was considered for the sulfonic group (Fig. 3). The agreement between both electropherograms showed that the sulfonic group of the substituted SiA in the glycopeptide was only slightly more acidic than the carboxylic acid of SiA (p*K* 2.6 [41]).

The same models could be also applied to predict electrophoretic migration behaviour of peptides and glycopeptides of other glycoprotein digests at acidic pH or at a different pH value. As an example at acidic pH value, Fig. 4 shows the experimental and predicted electropherograms for a set of peptides and *N*-glycopeptides of the tryptic digest of human transferrin (Tf) using the polymer's law and the Stoke's law for the glycopeptides and the peptides, respectively. As can be observed, the concordance is again good, with minimal differences for the glycopeptides at the final part of the electropherogram. Therefore, these results prove that with this simple approach, valuable information about migration order and separation can be obtained before doing any experiment. In addition to *M* and the presence of carbohydrates, the accuracy of migration time prediction especially depend on the goodness of charge estimation, as we proved in a wide pH range with peptide hormones, neuropeptides and quinolones [18,31,32]. Accurate p*K* values are always recommended, but good p*K* estimates may be enough if separation pH differs in at least more than one pH unit from any of the true peptide or glycopeptide p*K* values.

4. Concluding remarks

Several classical semiempirical relationships have been tested for prediction of migration behaviour of a mixture of peptides and glycopeptides, originated from the digestion of rhEPO with trypsin and with trypsin–neuraminidase. Using good estimates of the p*K* values of the ionisable groups at the separation pH value (2.2), the Stoke's law ($q/M^{1/3}$) resulted in better linear correlations for the peptides, while glycopeptides fitted better with the classical polymer model ($q/M^{1/2}$). The obtained models were useful to predict migration times of rhEPO peptides and glycopeptides and to simulate their separation electropherogram at the acidic separation conditions typically used to analyse protein and glycoprotein digests by CE–MS with optimum sensitivity in positive ESI mode. The established models could be used to obtain, in a simple and straightforward way, 'dry-lab' peptide–glycopeptide electrophoretic maps of other biologically relevant glycoproteins, such as alpha-1-acid glycoprotein (AGP), apolipoprotein-CIII (APO-CIII), rhEPO analogues, etc.

Acknowledgements

Part of this work was supported by the Spanish Ministry of Science and Innovation (CTQ2011-27130). Albert Barroso thanks the *Generalitat de Catalunya* for a FI fellowship.

References

[1] Z. El Rassi, Electrophoretic and electrochromatographic separation of proteins in capillaries: an update covering 2007–2009, Electrophoresis 31 (2010) 174–191.
[2] A. Taichrib, M. Pioch, C. Neusüss, Toward a screening method for the analysis of small intact proteins by CE-ESI-TOF MS, Electrophoresis 33 (2012) 1356–1366.
[3] R. Haselberg, G.J. de Jong, G.W. Somsen, CE–MS for the analysis of intact proteins 2010–2012, Electrophoresis 34 (2013) 99–112.
[4] M. Herrero, E. Ibañez, A. Cifuentes, Capillary electrophoresis-electrospray–mass spectrometry in peptide analysis and peptidomics, Electrophoresis 29 (2008) 2148–2160.
[5] V. Kašička, Recent developments in CE and CEC of peptides (2009–2011), Electrophoresis 33 (2012) 48–73.
[6] P.G. Righetti, R. Sebastiano, A. Citterio, Capillary electrophoresis and isoelectric focusing in peptide and protein analysis, Proteomics 13 (2013) 325–340.
[7] Y. Wang, B.R. Fonslow, C.C.L. Wong, A. Nakorchevsky, J.R. Yates, Improving the comprehensiveness and sensitivity of sheathless capillary electrophoresis-tandem mass spectrometry for proteomic analysis, Anal. Chem. 84 (2012) 8505–8513.
[8] L. Sun, G. Zhu, X. Yan, M.M. Champion, N.J. Dovichi, Capillary zone electrophoresis for analysis of complex proteomes using an electrokinetically pumped sheath flow nanospray interface, Proteomics 0 (2013) 1–7.
[9] F. Foret, T.J. Thompson, P. Vouros, B.L. Karger, P. Gebauer, P. Bocek, Liquid sheath effects on the separation of proteins in capillary electrophoresis/electrospray mass spectrometry, Anal. Chem. 66 (1994) 4450–4458.
[10] J.H. Wahl, D.C. Gale, R.D. Smith, Sheathless capillary electrophoresis-electrospray ionization mass spectrometry using 10 mu m I.D. capillaries: analyses of tryptic digests of cytochrome c, J. Chromatogr. A 659 (1994) 217–222.
[11] 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, Methods Mol. Biol. 951 (2013) 145–169.
[12] S. Amon, A.D. Zamfir, A. Rizzi, Glycosylation analysis of glycoproteins and proteoglycans using capillary electrophoresis–mass spectrometry strategies, Electrophoresis 29 (2008) 2485–2507.
[13] A.D. Zamfir, J. Peter-Katalinić, Capillary electrophoresis-mass spectrometry for glycoscreening in biomedical research, Electrophoresis 25 (2004) 1949–1963.
[14] 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.
[15] 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.
[16] V. Sanz-Nebot, F. Benavente, J. Barbosa, Liquid chromatography–mass spectrometry and capillary electrophoresis combined approach for separation and characterization of multicomponent peptide mixtures. Application to crude products of leuprolide synthesis, J. Chromatogr. A 950 (2002) 99–111.
[17] M. Sugimoto, S. Kikuchi, M. Arita, T. Soga, T. Nishioka, M. Tomita, Large-scale prediction of cationic metabolite identity and migration time in capillary electrophoresis mass spectrometry using artificial neural networks, Anal. Chem. 77 (2005) 78–84.
[18] F. Benavente, E. Balaguer, J. Barbosa, V. Sanz-Nebot, Modelling migration behavior of peptide hormones in capillary electrophoresis-electrospray mass spectrometry, J. Chromatogr. A 1117 (2006) 94–102.
[19] M. Sugimoto, A. Hirayama, M. Robert, S. Abe, T. Soga, M. Tomita, Prediction of metabolite identity from accurate mass, migration time prediction and

isotopic pattern information in CE-TOFMS data, Electrophoresis 31 (2010) 2311–2318.
[20] C. Simó, R. González, C. Barbas, A. Cifuentes, Combining peptide modeling and capillary electrophoresis–mass spectrometry for characterization of enzymes cleavage patterns: recombinant versus natural bovine pepsin A, Anal. Chem. 77 (2005) 7709–7716.
[21] B.J. Williams, W.K. Russell, D.H. Russell, Utility of CE–MS data in protein, Anal. Chem. 79 (2007) 3850–3855.
[22] S. Mittermayr, M. Olajos, T. Chovan, G.K. Bonn, A. Guttman, Mobility modeling of peptides in capillary electrophoresis, Trends Anal. Chem. 27 (2008) 407–417.
[23] S. Català-Clariana, F. Benavente, E. Giménez, J. Barbosa, V. Sanz-Nebot, Identification of bioactive peptides in hypoallergenic infant milk formulas by CE–TOF-MS assisted by semiempirical model of electromigration behavior, Electrophoresis 34 (2013) 1886–1894.
[24] J. Kim, R. Zand, D.M. Lubman, Electrophoretic mobility for peptides with post-translational modifications in capillary electrophoresis, Electrophoresis 24 (2003) 782–793.
[25] E.C. Rickard, M.M. Strohl, R.G. Nielsen, Correlation of electrophoretic mobilities from capillary electrophoresis with physicochemical properties of proteins and peptides, Anal. Biochem. 197 (1991) 197–207.
[26] N.J. Adamson, E.C. Reynolds, Rules relating electrophoretic mobility, charge and molecular size of peptides and proteins, J. Chromatogr. B 699 (1997) 133–147.
[27] A. Cifuentes, H. Poppe, Behavior of peptides in capillary electrophoresis: effect of peptide charge, mass and structure, Electrophoresis 18 (1997) 2362–2376.
[28] M. Castagnola, D.V. Rossetti, M. Corda, M. Pellegrini, F. Misiti, A. Olianas, B. Giardina, I. Messana, The pH dependence of predictive models relating electrophoretic mobility to peptide chemico-physical properties in capillary zone electrophoresis, Electrophoresis 19 (1998) 2273–2277.
[29] V. Solínová, V. Kasicka, D. Koval, J. Hlavácek, Separation and investigation of structure-mobility relationships of insect oostatic peptides by capillary zone electrophoresis, Electrophoresis 25 (2004) 2299–2308.
[30] V. Solínová, V. Kasicka, P. Sázelová, T. Barth, I. Miksík, Separation investigation of structure-mobility relationship of gonadotropin-releasing hormones by capillary zone electrophoresis in conventional isoelectric acidic background electrolytes, J. Chromatogr. A 1155 (2007) 146–153.
[31] V. Sanz-Nebot, F. Benavente, E. Hernández, J. Barbosa, Evaluation of the electrophoretic behaviour of opioid peptides separation by capillary electrophoresis-electrospray ionization mass spectrometry, Anal. Chim. Acta 577 (2006) 68–76.
[32] F. Benavente, E. Giménez, D. Barrón, J. Barbosa, V. Sanz-Nebot, Modeling the electrophoretic behavior of quinolones in aqueous and hydroorganic media, Electrophoresis 31 (2010) 965–972.
[33] K.I. Roy, C.A. Lucy, Influence of methanol as a buffer additive on the mobilities of organic cations in capillary electrophoresis, Electrophoresis 24 (2003) 370–379.
[34] R. Lee, A.S. Ptolemy, L. Niewczas, P. Britz-McKibbin, Integrative metabolomics for characterizing unknown low-abundance metabolites by capillary electrophoresis–mass spectrometry with computer simulations, Anal. Chem. 79 (2007) 403–415.
[35] F. Benavente, B. Andón, E. Giménez, J. Barbosa, V. Sanz-Nebot, Modeling the migration behavior of rabbit liver apothioneins in capillary electrophoresis, Electrophoresis 29 (2008) 2790–2800.
[36] G. Zauner, R.P. Kozak, R.A. Gardner, D.L. Fernandes, A.M. Deelder, M. Wuhrer, Protein *O*-glycosylation analysis, Biol. Chem. 393 (2012) 687–708.
[37] M. Wuhrer, Glycomics using mass spectrometry, Glycoconjugate J. 30 (2013) 11–22.
[38] J. Zaia, Capillary electrophoresis-mass spectrometry of carbohydrates, Methods Mol. Biol. 984 (2013) 13–25.
[39] D.S. Dalpathado, H. Desaire, Glycopeptide analysis by mass spectrometry, Analyst 133 (2008) 731–738.
[40] H. Desaire, Glycopeptide analysis, recent developments and applications, Mol. Cell. Proteomics 12 (2013) 893–901.
[41] R.M.C. Dawson, D.C. Elliott, W.H. Elliott, K.M. Jones, Data for Biochemical Research, third ed., Clarendon Press, 1989.
[42] Sillero, J.M. Ribeiro, Isoelectric points of proteins: theoretical determination, Anal. Biochem. 179 (1989) 319–325.
[43] T.T. Lee, E.S. Yeung, Facilitating data transfer and improving precision in capillary zone electrophoresis with migration indexes, Anal. Chem. 63 (1991) 2842–2848.
[44] J.R. Torres-Lapasió, J.J. Baeza-Baeza, M.C. Garcı'a-Alvarez-Coque, Description of the retention behaviour in micellar liquid chromatography as a function of pH, surfactant and modifier concentration, J. Chromatogr. A 769 (1997) 155–168.
[45] V. Hruška, M. Beneš, J. Svobodová, I. Zusková, B. Gaš, Simulation of the effects of complex-formation equilibria in electrophoresis: I. mathematical model, Electrophoresis 33 (2012) 938–947.

Chapter 4. Glycosylation study of glycoproteins by ion mobility mass spectrometry

The analysis of the abnormal glycosylation pattern of certain glycoproteins might not be enough to fully understand the extent of certain diseases. The scientific community has reported quite recently that the alteration of certain **isomeric glycans** might be the key to comprehend the behavior of these pathologies and propose proper treatments. In complex N-glycans, isomers are mainly originated due to the different orientation or location of the **sialic acid** and **fucose** monosaccharide residues.

Numerous works have been reported which focus their attention on derivatization protocols or more recent stationary phases in LC, with the aim of achieving the so-desired isomeric separation. MS/MS methodologies have also been proposed as interesting, albeit complex, attempts to distinguish between isomeric glycoconjugates. However, all these methods are still in their early stages of development, and, certain limitations are expected. Consequently, other analytical techniques able to separate and distinguish isomeric glycoconjugates in a straightforward manner are much needed.

In this regard, **IM-MS** has aroused great interest in the last few years for its power to separate isomeric compounds in the gas-phase. And, even more recently, IM-MS is becoming an alluring analytical technique for the separation of isomeric glycoconjugates. For this reason, a study aimed at assessing the proficiency of IM-MS for the separation of isomeric glycoconjugates due to the type of sialic acid linkage, at the three possible levels (intact glycoprotein, glycopeptides and glycans) was the last goal of this thesis.

The work obtained in this study has resulted in the publication indicated below:

- **Publication 4.1:** Ion mobility separation of glycoconjugate isomers due to different types of sialic acid linkage at intact glycoprotein, glycopeptide and glycan level. Albert Barroso, Estela Giménez, Albert Konijnenberg, Jaime Sancho, Victoria Sanz-Nebot, Frank Sobott. *Analytica Chimica Acta*, submitted for publication.

ION MOBILITY SEPARATION OF GLYCOCONJUGATE ISOMERS DUE TO DIFFERENT TYPES OF SIALIC ACID LINKAGE, AT THE INTACT GLYCOPROTEIN, GLYCOPEPTIDE AND GLYCAN LEVEL

Albert Barroso[1], Estela Giménez[1], Albert Konijnenberg[2], Jaime Sancho[3], Victoria Sanz-Nebot[1], Frank Sobott[2,4,5]

[1] Department of Chemical Engineering and Analytical Chemistry, University of Barcelona, Diagonal 645, 08028 Barcelona, Spain

[2] Biomolecular & Analytical Mass Spectrometry group, Department of Chemistry, University of Antwerp, Antwerp, Belgium

[3] Instituto de Parasitología y Biomedicina "López-Neyra" (IPBLN), CSIC, Armilla, Granada, Spain

[4] Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds LS2 9JT, United Kingdom

[5] School of Molecular and Cellular Biology, University of Leeds, LS2 9JT, United Kingdom

Keywords: glycosylation, ion mobility, isomers, mouse transferrin, sialic acid.

Abbreviations: ATD: arrival time distribution; CCS: collision cross section; CIA: collagen induced arthritis; hAGP: human α-1-acid glycoprotein; hTf: human transferrin; IM-MS: ion mobility – mass spectrometry; mTf: mouse transferrin; nano-ESI: nanoelectrospray ionization; nano-UPLC: nano ultra performance liquid chromatography; TOF: time-of-flight; TWIMS: travelling wave ion mobility spectrometry.

Abstract

The study of protein glycosylation can be regarded as an intricate but very important task, making glycomics one of the most challenging and interesting, albeit under-researched, type of "omics" science. Complexity escalates remarkably when considering that carbohydrates can form severely branched structures with many different constituents, which often leads to the formation of multiple isomers. In this regard, ion mobility (IM) spectrometry has recently demonstrated its power for the separation of isomeric compounds. In the present work, the potential of traveling wave IM (TWIMS) for the separation of isomeric glycoconjugates was evaluated, using mouse transferrin (mTf) as model glycoprotein. Particularly, we aim to assess the performance of this platform for the separation of isomeric glycoconjugates due to the type of sialic acid linkage, at the intact glycoprotein, glycopeptide and glycan level. Straightforward separation of isomers was achieved with the analysis of released glycans, as opposed to the glycopeptides which showed a more complex pattern. Finally, the developed methodology was applied to serum samples of mice, to investigate its robustness when analysing real complex samples.

1. Introduction

Glycosylation is by far one of the most common and complex posttranslational modifications, with more than half of all secretory and cellular proteins being glycosylated [1–3]. Carbohydrates enhance the functional diversity of proteins, but they can also define their destination or elicit an immune response. The presence of glycans in the surface of eukaryotic cells is vital, as they take part in important cellular events, such as cell–cell interactions and receptor recognition [4]. Notwithstanding its importance and the major role of glycosylation in a multitude of biological processes [5–7], the analysis and characterization of carbohydrates is usually difficult due to their inherent complexity - the main reason why advances in glycomics

have been scarcer compared to other "omics" sciences [8,9]. Very often, in contrast to more linearly assembled biological molecules such as proteins or oligonucleotides, carbohydrates can form complex structures, severely branched, with many monosaccharide constituents, which usually results in a multitude of isomers [10].

Mass spectrometry (MS)-based techniques are the prime option for the characterization of glycoproteins, as reliable structural information can be obtained [7,11]. MS is frequently used in conjunction with chromatographic or electrophoretic separation techniques, as this allows high sensitivity profiling and accurate characterization of heterogeneous glycan structures [12–14]. However, when analyzing isomeric glycan structures, MS often fails to separate them [8,15–17], as they have identical mass and atomic composition - thus they remain undistinguishable unless derivatization or less common stationary phases in liquid chromatography (LC) are used [18–22]. But even then, derivatization protocols can be time-consuming, expensive or hinder the ionization of some glycans, or, in the case of LC, the unambiguous identification is still impossible when different isomers coelute. Moreover, in the last few years, several tandem mass spectrometry methods have been reported that allow the identification of glycan isomers and the characterization of their structure [23–25]. However, few authors have studied the fragmentation of glycans with different sialic acid linkages. Even then, distinguishing by MS/MS between isomeric glycans due to sialic acid linkage is not trivial and, quite often, is based on differences in the relative abundance of certain fragment ions [26,27]. Therefore, a straightforward technique that helps to separate and differentiate those isomeric compounds is much needed.

In this regard, ion-mobility (IM) spectrometry coupled with MS has aroused some interest in the last years, not only in the glycomics field but also in other omics sciences, as a proficient analytical technique for the separation of isomeric compounds [3,8,10,15–17,28–30]. Ion mobility provides an additional dimension for 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 - thereby resolving ions that would be otherwise indistinguishable solely by MS, such as, for instance, isomers [31–35]. Particularly, IM measures the time (drift time) that a particular ion takes to cross a cell filled with an inert, neutral background gas (N_2 and He are most commonly used) at a controlled pressure under the influence of a weak electric field. The drift time of a specific ion is mainly due to ion-gas collisions; therefore, ions are separated due to their ion-neutral collision cross-section (Ω), related to the overall shape and topology of the ion [31–35]. Small compact ions have the shortest drift times, i.e. they arrive first, as a result of their smaller Ω. Moreover, the higher the charge of the ion, the greater the accelerating electric force, and therefore the more quickly the ion will cross the chamber. Consequently, the drift time of an ion is often described as being determined by the collision cross-section-to-charge ratio (Ω/z) [34]. When coupled on-line with MS (IM-MS), ion mobility provides three-dimensional analytical information for each detected species, i.e. shape-to-charge, mass-to-charge and abundance, thus allowing reliable analyte identification.

Nowadays, there are several IM methods next to the classical drift-time ion mobility spectrometry (DTIMS), such as field asymmetric waveform ion mobility spectrometry (FAIMS), but among them, traveling wave ion mobility spectrometry (TWIMS) is the one that has seen a major

growth in the last years [36,37]. In TWIMS, ions are propelled thanks to a sequence of symmetric potential waves continually propagating through a cell, each ion with its own velocity, thus different species transit the cell in different times. 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 analyzers in MS, such as time-of-flight (TOF), has made this platform an alluring option for structural analysis and isomer separation [37–39]. This platform, along with other IM methods, have been recently explored for the analysis of glycans or glycoconjugates by several authors [8,15,16,40–46].

In this work, TWIMS combined with TOF-MS was used for the study of glycoconjugate isomers which differ in the type of sialic acid linkage, with mouse transferrin (mTf) as a model glycoprotein. Sialic acid, an important monosaccharide residue of complex type N-glycans, may form primarily two types of linkages: α-2→3 or α-2→6. We aim to assess the capacity of TWIMS-TOF-MS (from now on referred to as IM-MS) as an analytical platform to separate α-2→3 and α-2→6 isomeric glycoconjugates at the intact glycoprotein, glycopeptide and glycan level. The developed methodology was also applied to serum samples of mice, to confirm its robustness when analyzing real complex samples.

2. Materials and methods

2.1. Chemicals

All chemicals used in the preparation of buffers and solutions were of analytical reagent grade. Isopropanol (iPrOH), hydrochloric acid (HCl, 37% (w/v)), formic acid (FA, 98–100%), ammonium acetate (NH_4Ac, ≥98.0%) and glycine (≥99.7%) were supplied by Merck (Darmstadt, Germany). CNBr-activated-Sepharose 4B was provided by GE Healthcare (Waukesha, WI, USA) and "NP-40 alternative" by Calbiochem (Darmstadt, Germany). Sodium chloride (NaCl, ≥99.5%), DL-Dithiothreitol (DTT, ≥99%), sodium cyanoborohydride ($NaBH_3CN$), 2-mercaptoethanol (β-ME), sodium dodecyl sulfate (SDS), iodoacetamide (IAA), ammonium hydrogencarbonate, sodium azide (NaN_3, ≥99.5%) water (LC-MS grade), acetonitrile (LC–MS grade) and mouse apotransferrin (mTf, reference: T0523) were supplied by Sigma–Aldrich (St. Louis, MO, USA) and Tris(hydroxymethyl) aminomethane (TRIS, ≥99.5%) by J.T. Baker (Deventer, Holland). Trypsin (Sequencing grade modified) was provided by Promega (Madison, WI, USA). RapiGest® from Waters (Bedford, MA, USA) was used to facilitate the enzymatic digestion. Goat polyclonal antibody against human transferrin (hTf) (immunogen affinity purified) was purchased from Abcam (Cambridge, UK). Human transferrin (hTf) and human α-1-acid glycoprotein (hAGP) were used as additional examples of other glycosylated glycoproteins and were also supplied by Sigma–Aldrich.

2.2. Mice and induction of arthritis

WT mice were from Harlan Ibérica (Barcelona, Spain). All studies with live animals were authorized by the Institute of Parasitology and Biomedicine "López-Neyra" (IPBLN) and Universidad de Cantabria Institutional Laboratory Animal Care and Use Committees. For the induction of collagen-induced arthritis (CIA), 8-12 weeks-old male mice were immunized as described elsewhere [47,48].

2.3 Purification of serotransferrin from mouse serum samples by immunoaffinity chromatography (IAC)

In order to isolate mTf from the rest of serum proteins, an immunoaffinity (IA) purification was carried out using a cyanogen-bromide sepharose column

where a polyclonal antibody against human transferrin (hTf) was bound, as detailed previously [49]. The IA procedure consisted of: first, a conditioning step with two washes of 10 mM Tris-HCl; second, approximately 25 µL of serum were diluted 1:8 in 10 mM Tris-HCl (pH 7.6-7.7) in order to improve antigen-antibody interaction, and passed through the column ten times. After washing with 10 mM Tris-HCl and 0.5 M NaCl (pH 7.6-7.7), retained mTf was eluted with 100 mM glycine-HCl (pH 2.5). Eluted mTf was immediately neutralized with 0.5 M Tris. Afterwards, glycine-HCl buffer was exchanged for water by ultracentrifugation, using Microcon YM-10 (MW cut-off 10 kDa, Millipore, Bedford, MA, USA). Then, samples were evaporated to dryness using a SpeedVac™ concentrator (Thermo Fisher Scientific, Waltham, MA, USA) and stored at -20˚C until use. Finally, the IA column was washed and stored in 10 mM Tris-HCl and 0.01% (w/v) NaN_3 (pH 7.6-7.7).

2.4. Analytical approaches for mTf glycosylation study

2.4.1. Intact glycoprotein analysis

mTf standard (25 µg) was desalted using three different procedures: dialysis, size exclusion and ultracentrifugation. Briefly, in the first method, D-Tube™ dialyzers from Merck-Millipore were left in contact with 100 mM NH_4Ac for 15 min. Afterwards, the sample was placed into the dialyzer and left in contact with 500 mL of 100 mM NH_4Ac for 2 h at 4˚C. Later, the buffer was renewed with 500 mL and the dialysis was allowed to continue for 2 more hours, repeating this process twice. Finally, the sample was carefully recovered and stored at -20˚C until analysis. Regarding the size exclusion procedure, the sample was desalted and the buffer exchanged using Micro Bio-Spin™ columns from BioRad (Hercules, California, USA) following the manufacturer instructions. Columns were centrifuged to remove the excess of packing buffer and washed three times with 500 µL of 100 mM NH_4Ac. Finally, the sample was added and collected in a new tube after centrifugation. Lastly, the ultracentrifugation procedure was carried out using Microcon YM-10 (MW cut-off 10 kDa) to desalt and exchange the buffer of the sample. Filters were washed with 100 mM NH_4Ac and centrifuged for 10 min at 10000 g. Afterwards, the sample was added and washed with 50 µL of 100 mM NH_4Ac a total of 4 times, centrifuging each time for 10 min at 10000 g. Finally, the final volume was recovered in a new vial after centrifugation for 2 min at 1000 g, and reconstituted to the initial volume (25 µL). Centrifugation procedures were carried out in a MiniSpin® centrifuge (Eppendorf, Hamburg, Germany) at room temperature. In all cases, intact mTf in 100 mM NH_4Ac was injected directly into the mass spectrometer under non-denaturing conditions and detected in positive ion mode.

2.4.2. Glycopeptide analysis

mTf standard (25 µg) was reduced, alkylated and immediately subjected to trypsin digestion in the presence of RapiGest® as explained in a previous work [50]. mTf tryptic digests were stored at -20˚C until analysis.

For the analysis of glycopeptides, a Waters Nano ACQUITY UPLC® was used with a double binary gradient pump, using a peptide BEH C18 column (1.7 µm particle diameter, 130 Å pore, 100 x 0.1 mm length x ID; Waters). Experiments were performed at room temperature with gradient elution at a flow rate of 400 nL min^{-1}. Eluting solvents were A: water with 0.1% (v/v) of formic acid (FA), and B: acetonitrile with 0.1% (v/v) FA. Solvents were degassed by sonication (10 min) before use. The optimum elution program was: solvent B from 10 to 60% (v/v) within 20 min as

linear gradient, followed by cleaning and re-equilibration steps of B: 60 to 100% (v/v) (5 min), 100% (v/v) (5 min), 100 to 10% (v/v) (5 min) and 10% (v/v) (5 min). Before analysis, samples were filtered using a 0.22 µm polyvinylidene difluoride centrifugal filter (Ultrafree-MC, Millipore, Bedford, MA, USA) centrifuging at 10,000 g for 4 min. Sample injection (300 nL) was performed with an autosampler refrigerated at 4˚C. Control of the instrument was performed using MassLynx 4.1 (Waters).

2.4.3. Glycan analysis

IAC purified mTf or mTf standard (25 µg) was reduced with 0.5% 2-mercaptoethanol (β-ME) and 0.5% SDS and subjected to enzymatic digestion with PNGase F as described in [19]. Afterwards, released glycans were isolated by solid phase extraction (SPE) using Hypercarb cartridges (25 mg, 1 mL volume, Thermo Fisher Scientific) and, subsequently, purified by ice-cold acetone precipitation following the procedure reported in [19] in both cases. Reduced glycans were diluted with 50:50 H_2O/ACN with 0.1% FA and directly analyzed by IM-MS in negative ion mode.

2.5. Ion mobility-mass spectrometry

For IM-MS analysis a Synapt G2 HDMS instrument from Waters was used. Samples were directly introduced into the vacuum of the mass spectrometer using home-made nano-ESI gold-coated borosilicate capillaries, unless when analyzed by nano-UPLC-IM-MS, in which case an in-line nano-ESI interface with commercially available coated needles was used.

Spectra were acquired in positive mode for the analysis of intact glycoproteins and glycopeptides, and in negative mode for glycans, and conditions were optimized in each case. The voltages for spray capillary, sampling cone, trap CE, trap DC bias and transfer CE were, respectively: intact glycoproteins, 1.4-1.6 kV, 30 V, 4 V, 40 V and 0 V; glycopeptides, 1.5-1.7 kV, 50 V, 4 V, 20 V and 0 V; and glycans, 1.5 kV, 50 V, 4 V, 45 V and 0 V. The “trap CE” voltage was only increased to 60 V when fragmentation of the glycopeptides was the goal. Mass spectrometer control and spectra processing were carried out using MassLynx 4.1 (Waters).

The software IMoS [51,52], available for free at *imospedia.com*, was used for the theoretical calculations of the collision cross sections (CCS) of glycans, using their minimum energy structures. The online tool *carbohydrate builder*, available at *glycam.org* [53], was used to generate the required input for theoretical calculations. The tool allows building different glycan isomers based on monosaccharide unit and linkage type, and generates minimum energy structures.

3. Results and discussion

3.1. Intact glycoprotein analysis

It is well established that intact proteins can be analyzed mainly in two different ways when using electrospray ionization: under denaturing or non-denaturing (i.e., “native”) conditions. In the first case, the protein appears highly charged with a broad charge state distribution; thus, the separation of the different glycoforms is difficult. On the other hand, an advantage of native MS is that it strongly reduces charging, hence the different glycoforms can be more clearly resolved [54].

Using native MS in this work, the concentration of ammonium acetate and the nano-ESI source parameters were optimized to improve and obtain the best possible sensitivity and separation between mTf glycoforms in positive ion mode. Moreover, in order to obtain an acceptable separation between the several glycoforms, mTf was washed repeatedly with 100 mM NH_4Ac to eliminate excipients and salts, and injected directly into the mass spectrometer without further separation. Without washing, no glycoform separation was observed whatsoever. Three methods were evaluated to desalt the glycoprotein: dialysis, size exclusion and ultracentrifugation, with the latter one giving the best results (see supplementary Figure 1). Afterwards, the drift times of intact glycoforms were determined by ion mobility. In Figure 1-a, the ion with charge +19 of mTf is shown, as well as the arrival time distributions (ATD) of different sections of this peak (Figure 1-a (i-v)), each corresponding to one intact glycoform of mTf). As can be observed in the ATD profiles, only one, relatively wide peak is obtained for each glycoform. Hence, these results imply that, at least for large glycoproteins with a low degree of glycosylation, the glycan part has barely any influence on the overall size of the whole molecule as the CCS appears

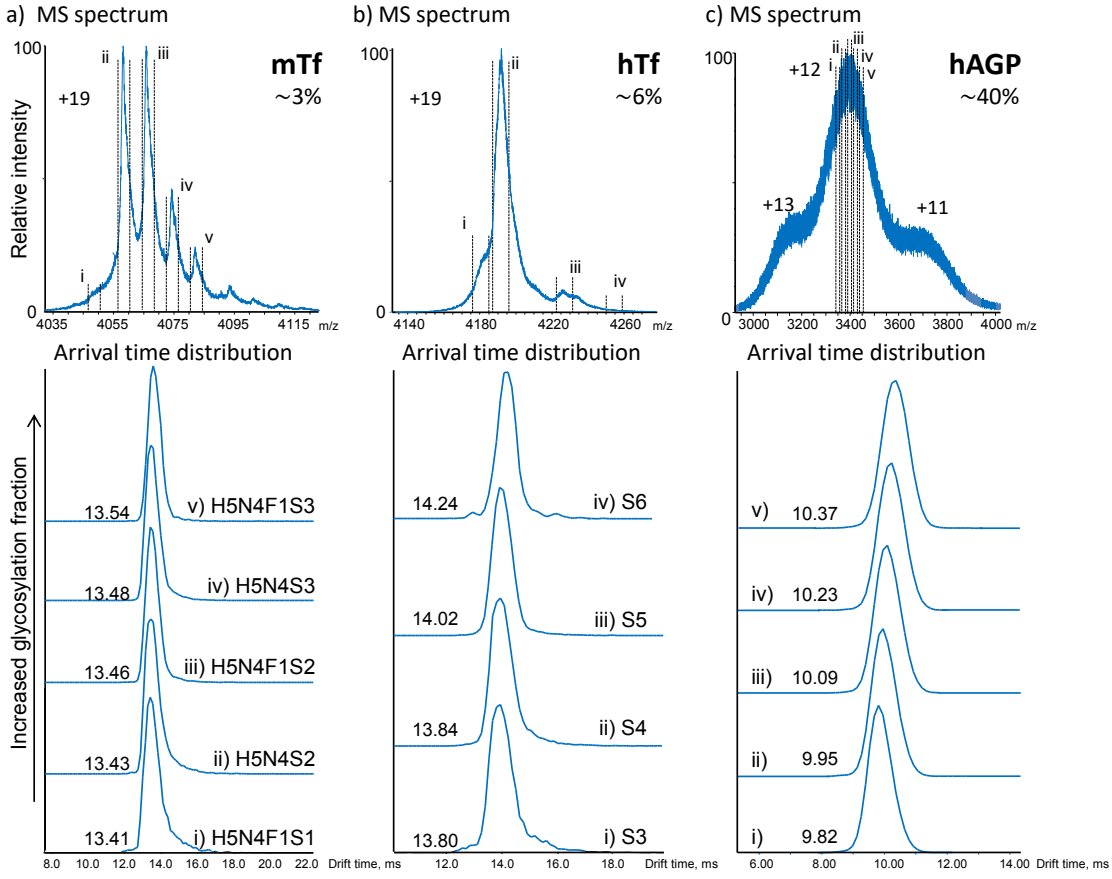


Figure 1: Mass spectra showing the ion with charge +19 and the corresponding drift time (arrival time) distributions, or ATD, of a) intact mTf and b) intact hTf; and also c) ions with charge +13, +12 and +11 of intact hAGP and the corresponding ATD. The value indicated corresponds to the approximate glycosylation percentage (w/w) of each protein, calculated as the mass of the most abundant glycan per glycosylation site divided by the mass of the glycoprotein. i)-v): indicates the glycoform or the region of the mass spectrometric peak. H: hexose; N: N-acetylhexosamine; F: fucose; S: sialic acid (in this case, all S are N-glycolylneuraminic acid)

Table 1: Relative increase in m/z and drift time between glycoforms i) and ii) (Figure 1) of mTf, hTf and hAGP

		mTf	hTf	hAGP
Relative m/z difference (%)		0.247	0.239	0.746
Relative drift time difference (%)		0.149	0.290	1.324
Normalization*	**Relative m/z (%)**	0.247	0.247	0.247
	Relative drift time (%)	0.149	0.299	0.438

* The relative m/z for hTf and hAGP was changed to have the same value as mTf (0.247%) and, therefore, the relative drift time was proportionally modified

virtually unaffected by glycosylation. hTf and hAGP were also studied to see if the same behaviour was observed with other glycoproteins. Figure 1-b and 1-c show the separation between intact glycoforms in the MS spectra of hTf (for the ion with charge +19) and hAGP, respectively. As can be observed, the peak was better resolved for hTf, but no separation was observed for hAGP, as the number of glycoforms was considerably higher and the overlapping of several ion distributions was unavoidable. Moreover, one wide peak was observed again when the drift time of each section of the mass spectrum was measured for both glycoproteins. Regarding hTf, each section corresponded to one glycoform (Figure 1-b (i-iv)), whereas, in the case of hAGP, the resolution was not high enough to ensure that a selected region of the peak corresponded to one defined glycoform (Figure 1-c (i-v)), because of the high glycosylation degree of hAGP. The higher the m/z region of this peak however, the bigger the carbohydrate fraction is (usually meaning more complex and branched glycans). This is due to the fact that an increase of m/z within the same charge state can only be due to an increase in the number of glycan subunits of the carbohydrate fraction, as the polypeptide backbone remains unaltered. For the sake of consistency, we will refer to each section of the peak as glycoform, even if, as in the case of hAGP, they are not resolved. Interestingly, the drift time increased with increasing carbohydrate fraction, i.e. complexity and branching of the glycoforms, in all three cases (see Figure 1-a (i-v), 1-b (i-iv) and 1-c (i-v), type of glycoform indicated where possible). The relative drift time differences between glycoforms were higher in proteins with higher glycosylation content, as can be observed for hAGP (see Figure 1-c). Moreover, if we compare the drift time between different glycoforms of mTf and hTf, the increase in drift time was more noticeable in hTf, as the percentage of glycosylation is slightly higher here in relation to the total protein mass. As Table 1 shows, for the same relative increase in m/z, the relative drift time increase is higher for hAGP, which demonstrates that for glycoproteins with a high percentage of glycosylation, differences in the glycosylation have a greater effect on the drift time of the whole molecule (i.e., they significantly alter the CCS of the whole glycoprotein). It could then also be conceivable to separate isomers due to the different sialic acid linkage (of the same glycoform) at the intact protein level if the glycan:peptide ratio was high enough. However, this option would not have been viable for all types of glycoproteins, thus, other alternatives were studied.

Table 2: Theoretical and experimental molecular mass (Mr), mass error and detected charge states of the 10 glycoforms for the N_{494} glycopeptide of mTf detected by nano-UPLC-IM-MS

Glycopeptide	Glycoform*	Theoretical M_r	Experimental M_r	Mass error (ppm)	Observed charge state
N_{494}	H3N3S1	2892.2138	2892.2840	24	+2
	H3N3F1S1	3038.2717	3038.2920	7	+2
	H5N4S1	3419.3988	3419.3908	2	+2, +3
	H5N4F1S1	3565.4568	3565.4656	2	+3
	H5N4S2	3726.4892	3726.4993	3	+2, +3
	H5N4F1S2	3872.5471	3872.5600	3	+2, +3
	H5N4S3	4033.579	4033.6210	10	+2, +3
	H5N4F1S3	4179.6374	4179.6958	14	+2, +3
	H6N5S3	4398.7117	4398.7813	16	+3
	H6N5F1S3	4544.7696	4544.8321	14	+3

* H: hexose; N: N-acetylhexosamine; F: fucose; S: sialic acid (in this case, all S are N-glycolylneuraminic acid)

3.2. Glycopeptide analysis

As mTf only shows one glycosite at asparagine 494, only one glycopeptide is expected after tryptic digestion (N_{494} glycopeptide, peptide: NSTLCDLCIGPLK). However, with the direct injection of the digest, the N_{494} glycopeptide was not detected, hence, an additional separation before the MS was mandatory. In this regard, nano-UPLC was used as a separation technique prior to IM-MS detection, in order to separate the glycopeptide from the other peptides and simultaneously determine the drift time of the different glycoforms. Injection volume, flow rate and gradient were optimized to obtain the best sensitivity with a stable spray (see section 2.4, Materials and Methods). Table 2 shows the detected glycoforms at the peptide level, their theoretical and experimental masses, mass error and observed charge states.

Regarding the determination of the drift time of the different glycopeptide glycoforms, even though a range of different values for the wave velocity (WV) and wave height (WH) of the TWIMS device were tested, only one drift time value was observed. This suggests that no isomer separation was possible at this mobility resolution, which is probably due to the fact that the different isomers had similar CCS despite the distinct orientation of the sialic acid. Recently, Hinneburg et al. [15] and Guttman et al. [44] also described this observation analyzing glycopeptides directly by IM-MS. They observed no isomeric separation, unless fragmentation of the glycopeptide was carried out and one of the observed smaller fragments still contained the sialic acid. This fragment (obtained before the IM cell) had different drift times depending on the sialic acid linkage, because a change in the orientation of the sialic acid was more noticeable (i.e. the CCS was more affected) in a smaller analyte. We also tested this approach, by fragmenting the most intense glycopeptide glycoform (N(H5N4S2)STLCDLCIGPLK) before the IM cell. Figure 2-a shows the collision induced dissociation (CID) MS/MS spectrum for this glycoform. Several fragments were obtained, e.g., the glycolylneuraminic acid

(S1) and H1N1 and H2N1 fragments. Among all the fragments however, H1N1S1 was the one that still retained the sialic acid linkage and had enough intensity to yield a good and accurate drift time measurement. Isomer separation was observed in the aforementioned glycan fragment. Specifically, as it is shown in Figure 2-c, two drift time peaks were clearly observed at this m/z, which are postulated to be due to the sialic acid being α-2→6 and α-2→3 linked.

In order to confirm that the distinct drift times observed for the glycan fragment could be due to different sialic acid linkages, the theoretical CCS of the H1N1S1 isomers were calculated. The obtained CCS values were: 233.4 Å^2 for the α-2→6 linkage and 243.7 Å^2 for the α-2→3 glycan, different enough to be separated by IM. This suggests that the more prominent, lower drift time peak (i.e.,

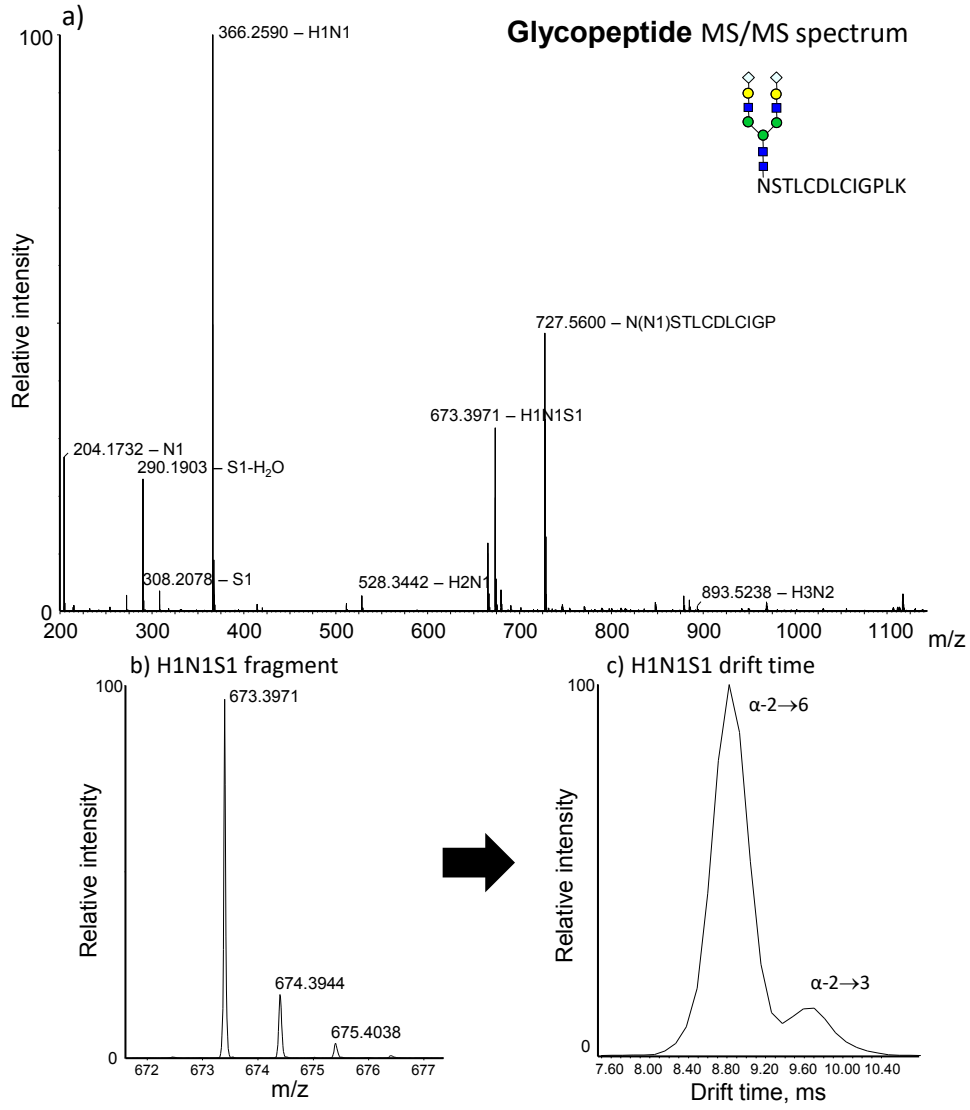


Figure 2: a) MS/MS spectrum for the mTf glycopeptide glycoform N(H5N4S2)STLCDLCIGPLK; b) mass spectra of a fragment that still keeps the sialic acid (H1N1S1) and c) arrival time distribution of this fragment (m/z range: 673.3-673.5). The symbols used for the representation of the glycoform H5N4S2 follow the Consortium for Functional Glycomics (CFG) rules [56]. H: hexose; N: N-acetylhexosamine; F: fucose; S: sialic acid (in this case, all S are N-glycolylneuraminic acid)

lower CCS) corresponds to the α-2→6 glycan, whereas the second peak represents the α-2→3 glycan.

Our results are in accordance with those obtained by Hinneburg et al. [15], who, using modelling, drew the same conclusions. This is also in agreement with Guttman et al. [44], who found that biantennary glycoforms showed lower content of α-2→3 sialylation. In our case, the percentage of α-2→3 glycan was approximately 14% (taking into account the measured area of each drift time peak, see Figure 2-c). However, this ratio can vary depending on the glycoprotein studied [44]. With this method, using nano-UPLC-IM-MS and drift time measurement of the MS/MS fragments, separation of glycopeptide isomers was achieved.

3.3. Glycan analysis

Finally, we proceeded with the study of the enzymatically released glycans using PNGase F. Different solvents were tested to obtain the best spray and ionization yield, and a slightly superior glycan signal was obtained at 50:50 H_2O:ACN with 0.1% FA. Table 3 shows the detected mTf glycans along with their theoretical and experimental masses, mass error and observed charge states. In Figure 3 the ATD profile of H5N4S2 glycans at different values of WV and WH is shown. Two peaks are observed which are believed to correspond to two different isomers of the glycan H5N4S2, in analogy to the previously obtained results with the glycopeptide N(H5N4S2)STLCDLCIGPLK. Fine tuning of the IM parameters was mandatory in order to resolve both drift time peaks. A WV value of 450 m s^{-1} and WH of 25 V was selected as optimal for the analysis of all mTf glycans. It is worth mentioning that in this case, glycans were analyzed in negative mode and, as can be observed in Figure 3, the first isomer was clearly the less abundant one. This is in contrast to the glycopeptide analysis which was done in positive mode, where the less abundant isomer was the second one. This could be due to the fact that both molecules were, actually, quite different - specifically, the glycan had eleven monosaccharide units, as opposed to the fragment observed in the glycopeptide

Table 3: Theoretical and experimental molecular mass (Mr), mass error and detected charge states of the 10 mTf glycans detected by IM-MS

Glycan*	Theoretical M_r	Experimental M_r	Mass error (ppm)	Observed charge state
H3N3S1	1420.4975	1420.4659	22	-1
H3N3F1S1	1566.5554	1566.5692	9	-1
H5N4S1	1947.6825	1947.7031	11	-1
H5N4F1S1	2093.7404	2093.7891	23	-1
H5N4S2	2254.7729	2254.7650	4	-2
H5N4F1S2	2400.8308	2400.8240	3	-2
H5N4S3	2561.8632	2561.9150	20	-2
H5N4F1S3	2707.9211	2707.9470	10	-2
H6N5S3	2926.9954	2926.9334	21	-2, -3
H6N5F1S3	3073.0533	3073.0036	16	-3

* H: hexose; N: N-acetylhexosamine; F: fucose; S: sialic acid (in this case, all S are N-glycolylneuraminic acid)

analysis which had only three. With regard to other glycans, three different peaks, i.e., three drift times, were observed for H6N5S3 (see supplementary Figure 2-c), albeit separation was slightly worse than observed for H5N4S2. H5N4S1 on the other hand only showed one wide peak (see supplementary Figure 2-d), most probably implying that only one glycan isomer exists. However, the absence of other drift time peaks does not preclude the presence of other isomers, as they can have similar CCS. It is also worth pointing out that when the glycan contained a fucose unit, the isomeric separation was somewhat hampered, as observed in supplementary Figure 2-b. We reckon that the addition of one extra monosaccharide unit might affect the global CCS of the glycan, and as the whole glycan is bigger, a small variation in the orientation of the sialic acid is less noticeable.

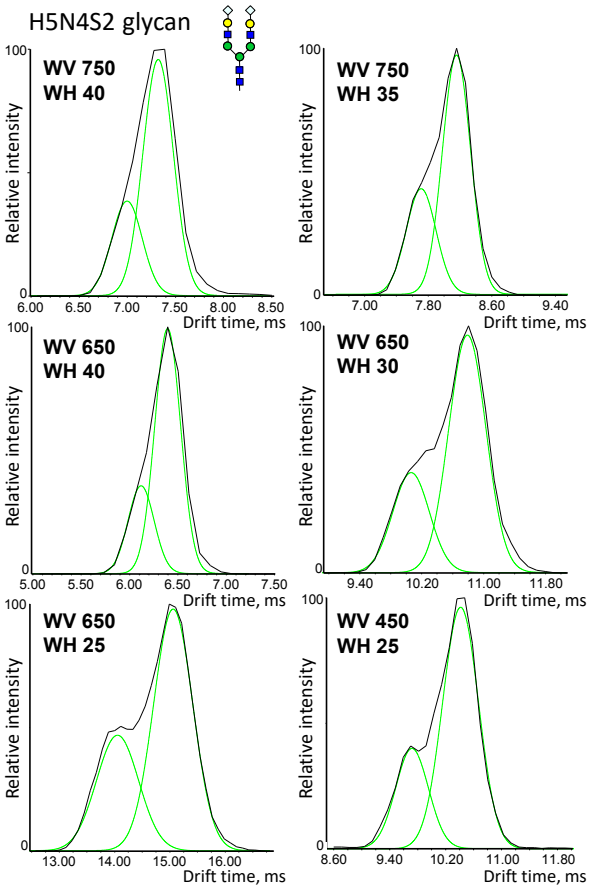


Figure 3: Arrival time distributions for the H5N4S2 glycan released from mTf at different wave height (WH, in V) and wave velocity (WV, in m s-1) combinations. The symbols used for the representation of the H5N4S2 glycan follow the Consortium for Functional Glycomics (CFG) rules [56]. H: hexose; N: N-acetylhexosamine; F: fucose; S: sialic acid (in this case, all S are N-glycolylneuraminic acid)

To confirm that separation of isomeric glycans due to the sialic acid linkage was also possible with other glycoproteins, hAGP was also digested with PNGase F and the released glycans analyzed by IM-MS. Separation of isomeric glycans was achieved, seemingly obtaining the same results as with mTf. As an example, the arrival time distributions of the H5N4S2 and H6N5S2 glycans are shown in supplementary Figure 3. As can be seen, two drift time peaks were obtained for the H5N4S2 glycan and three peaks for H6N5S3, albeit separation was poorer in both cases when compared to mTf.

Moreover, to confirm that ion mobility separation of these possible isomers is due to different sialic acid linkages, the theoretical CCS values of H5N4S2 with $(\alpha\text{-}2\rightarrow3)_2$, $(\alpha\text{-}2\rightarrow6)(\alpha\text{-}2\rightarrow3)$ and $(\alpha\text{-}2\rightarrow6)_2$ sialic acids were calculated. The obtained CCS were: 652.3 Å^2 for the $(\alpha\text{-}2\rightarrow3)_2$ glycan, 635.4 Å^2 for $(\alpha\text{-}2\rightarrow6)(\alpha\text{-}2\rightarrow3)$, and 623.9 Å^2 for the $(\alpha\text{-}2\rightarrow6)_2$ form. The differences between these calculated values suggest that the observed ion mobility peaks could be due to isomeric glycans with different types of sialic acid linkage. With the knowledge of theoretical CCS, and the abundance of the $\alpha\text{-}2\rightarrow3$ H1N1S1 fragment being lower than the $\alpha\text{-}2\rightarrow6$ (based on fragmentation of the H5N4S2 glycopeptide, see above), the two peaks observed for the H5N4S2 glycan were tentatively assigned. We suggest that the peak with the highest drift time, and

highest abundance, could be the glycan (α-2→6)(α-2→3) linkage. Whereas, the one with the lowest drift time (and lowest abundance) could be the (α-2→6)$_2$-linked glycan. However, alternative approaches for the study of isomeric glycoconjugates, for instance using specific sialidases [18], could be useful to obtain complementary information and reliably assign the different isomers.

3.4. Mice serum sample

To further asses the ability of the established method to separate isomeric glycoconjugates in biological samples, we measured the drift time of mTf glycans purified from serum samples. Only the analysis of the released glycans by IM-MS was included in this study, as it was found to be the most sensitive and straightforward approach to obtain information about isomeric forms. Two serum samples were analyzed: one healthy control and one sample with collagen-induced arthritis (CIA), an autoimmune disease known to alter the glycosylation pattern of mTf [50,55]. As can be seen in Figure 4, two peaks were observed for H5N4S2 glycan and three peaks for H6N5S3 in both samples, with the same drift time and similar relative intensities as in the mTf standard (compare Figure 4 with

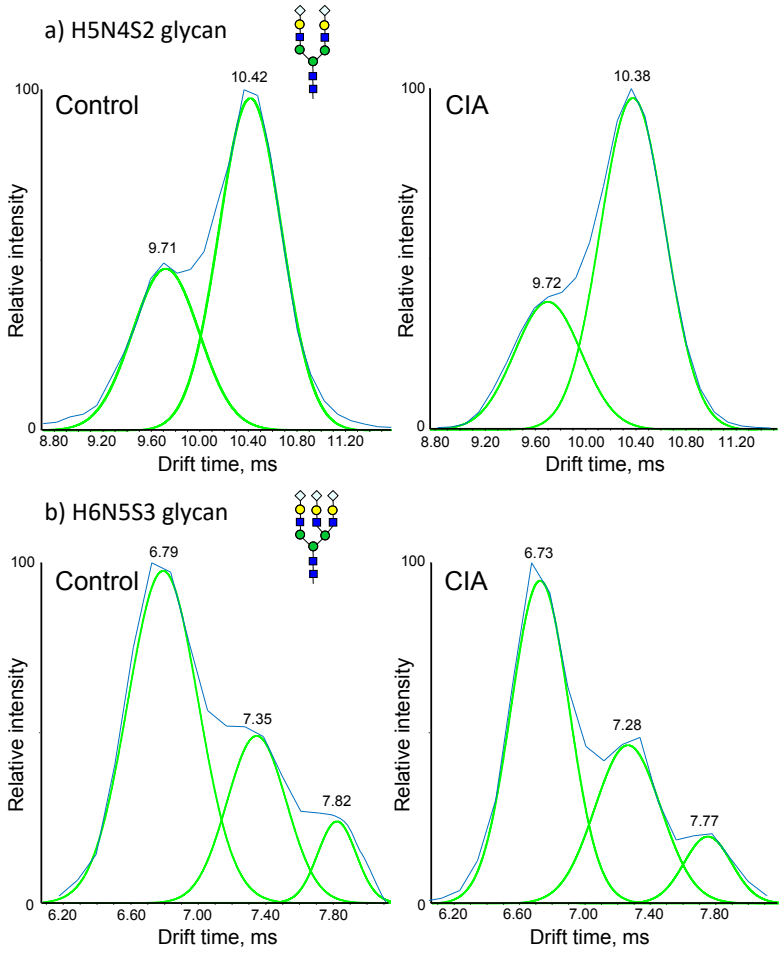


Figure 4: Arrival time distributions for the glycans a) H5N4S2 and b) H6N5S3 released from mTf in a healthy mouse serum and a serum from a mouse with collagen-induced arthritis (CIA). The symbols used for the glycan representation follow the Consortium for Functional Glycomics (CFG) rules [56]. H: hexose; N: N-acetylhexosamine; F: fucose; S: sialic acid (in this case, all S are N-glycolylneuraminic acid).

supplementary Figure 2). Although additional samples are required to observe possible differences in the relative abundance of glycan isomers between control and pathological samples, the presented methodology shows great potential for the separation of isomeric glycans and the discovery of novel biomarkers in glycomic studies.

4. Concluding remarks

In the present paper, the capability of IM-MS to separate isomeric glycoconjugates which are due to different types of sialic acid linkage (i.e. α-2→3 and α-2→6) has been evaluated at three different levels: intact glycoprotein, glycopeptides, and the released glycans. Separation of isomeric glycoconjugates is an important task in the glycomics field, because it has been reported that differences in the abundance of some glycan isomers might be of great importance for the early diagnosis or control of, for instance, inflammatory diseases and certain types of cancer.

With the current capabilities of Synapt IM-MS instrumentation, isomeric separation cannot be obtained at the intact glycoprotein and glycopeptide level. Released glycans however can be separated after optimization of the IM parameters. As stated before by others [15,44], and also demonstrated in this work, there is a workaround to distinguish different types of sialic acid linkage in glycopeptides that takes into account the mobility of an MS/MS fragment which still retains the sialic acid. This method can however be time-consuming and rather difficult, as glycopeptides must be separated from the rest of the digest and, besides, the obtained glycopeptide fragment is, actually, a glycan.

The interest in using ion mobility for glycoconjugate separation and identification has seen a major growth in the last years. Thus, it is likely that new technologies and improvements will become available soon, including the advent of ion mobility instrumentation with up to ten times higher resolution. Therefore, the separation of sialic acid linkage isomers may also become possible at the level of glycopeptides and intact proteins in the future.

Acknowledgements

Part of this study was supported by the Spanish Ministry of Economy and Competitiveness (CTQ2011-27130 and CTQ2014-56777-R), and the Synapt instrument at the University of Antwerp was funded by the Hercules Foundation-Flanders. Albert Barroso thanks the Ministry of Education, Culture and Sport for the FPU (FPU13/01596) fellowship and the grant for short stays (EST15/00398).

5. References

[1] A.P. Corfield, M. Berry, Glycan variation and evolution in the eukaryotes, Trends Biochem. Sci. 40 (2015) 351–359.

[2] H.A. Abbass, H.A. El Hassan, H. Bahmad, A. Zebian, R. Youssef, J. Ismail, R. Zhu, S. Zhou, X. Dong, M. Nasser, M. Bahmad, H. Darwish, Y. Mechref, F. Kobeissy, Glycosylation and other post translational modifications alterations in neurodegenerative diseases: current status and future role in neurotrauma, Electrophoresis. 37 (2016) 1549–1561.

[3] F. Zhu, J.C. Trinidad, D.E. Clemmer, Glycopeptide Site Heterogeneity and Structural Diversity Determined by Combined Lectin Affinity Chromatography/IMS/CID/MS Techniques, J. Am. Soc. Mass Spectrom. 26 (2015) 1092–1102.

[4] A. Grigorian, S. Torossian, M. Demetriou, T-cell growth, cell surface organization, and the galectin glycoprotein lattice, Immunol. Rev. 230 (2009) 232–246.

[5] R.A. Dwek, Glycobiology: towards understanding the function of sugars, Chem. Rev. 96 (1996) 683–720.

[6] P.M. Rudd, T. Elliott, P. Cresswell, I.A. Wilson, R.A. Dwek, Glycosylation and the immune system, Science. 291 (2001) 2370–2376.

[7] H.J. Allen, E.C. Kisailus, eds., Glycoconjugates: Composition: Structure, and Function, CRC Press, 1992.

[8] J. Hofmann, H.S. Hahm, P.H. Seeberger, K. Pagel, Identification of carbohydrate anomers using ion mobility-mass spectrometry, Nature. 526 (2015) 241–244.

[9] 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.

[10] 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.

[11] N. Leymarie, J. Zaia, Effective use of mass spectrometry for glycan and glycopeptide structural analysis, Anal. Chem. 84 (2012) 3040–3048.

[12] M.L.M. Hennrich, A.A. Gavin, Quantitative mass spectrometry of posttranslational modifications: Keys to confidence, Sci. Signal. 8 (2015) 1–5.

[13] M. Wuhrer, Glycomics using mass spectrometry, Glycoconj. J. 30 (2013) 11–22.

[14] Y. Zhang, J. Jiao, P. Yang, H. Lu, Mass spectrometry-based N-glycoproteomics for cancer biomarker discovery, Clin. Proteomics. 11 (2014) 18.

[15] 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.

[16] D.J. Harvey, C.A. Scarff, M. Edgeworth, M. Crispin, C.N. Scanlan, F. Sobott, S. Allman, K. Baruah, L. Pritchard, J.H. Scrivens, Travelling wave ion mobility and negative ion fragmentation for the structural determination of N-linked glycans, Electrophoresis. 34 (2013) 2368–2378.

[17] P. Both, A.P. Green, C.J. Gray, R. Šardzík, J. Voglmeir, C. Fontana, M. Austeri, M. Rejzek, D. Richardson, R.A. Field, G. Widmalm, S.L. Flitsch, C.E. Eyers, Discrimination of epimeric glycans and glycopeptides using IM-MS and its potential for carbohydrate sequencing, Nat Chem. 6 (2014) 65–74.

[18] M. Mancera-Arteu, E. Giménez, J. Barbosa, V. Sanz-Nebot, 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, Anal. Chim. Acta. 940 (2016) 92–103.

[19] E. Giménez, M. Balmaña, J. Figueras, E. Fort, C. 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.

[20] 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.

[21] P. Jackson, M.I. Attalla, N-Nitrosopiperazines form at high pH in post-combustion capture solutions containing piperazine: a low-energy collisional behaviour study, Rapid Commun. Mass Spectrom. 24 (2010) 3567–3577.

[22] F. Tousi, 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.

[23] 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.

[24] E. V. Da Costa, A.S.P. Moreira, F.M. Nunes, M.A. Coimbra, D. V. Evtuguin, M.R.M. Domingues, Differentiation of isomeric pentose disaccharides by electrospray ionization tandem mass spectrometry and discriminant analysis, Rapid Commun. Mass Spectrom. 26 (2012) 2897–2904.

[25] 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.

[26] 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.

[27] 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.

[28] D. Isailovic, R.T. Kurulugama, M.D. Plasencia, S.T. Stokes, Z. Kyselova, R. Goldman, Y. Mechref, M. V. Novotny, D.E. Clemmer, Profiling of Human Serum Glycans Associated with Liver Cancer and Cirrhosis by IMS–MS, J. Proteome Res. 7 (2008) 1109–1117.

[29] M. Sarbu, F. Zhu, J. Peter-Katalinic, D.E. Clemmer, A.D. Zamfir, Application of ion mobility tandem mass spectrometry to compositional and structural analysis of glycopeptides extracted from the urine of a patient diagnosed with Schindler disease, Rapid Commun. Mass Spectrom. 29 (2015) 1929–1937.

[30] M.M. Maurer, G.C. Donohoe, S.J. Valentine, Advances in ion mobility-mass spectrometry instrumentation and techniques for characterizing structural heterogeneity, Analyst. 140 (2015) 6782–6798.

[31] F. Lanucara, S.W. Holman, C.J. Gray, C.E. Eyers, The power of ion mobility-mass spectrometry for structural characterization and the study of conformational dynamics, Nat. Chem. 6 (2014) 281–294.

[32] D.M. Williams, T.L. Pukala, Novel insights into protein misfolding diseases revealed by ion mobility-mass spectrometry, Mass Spectrom. Rev. 32 (2013) 169–187.

[33] B.T. Ruotolo, J.L.P. Benesch, A.M. Sandercock, S.-J. Hyung, C. V Robinson, Ion mobility-mass spectrometry analysis of large protein complexes, Nat. Protoc. 3 (2008) 1139–1152.

[34] M.F. Bush, Z. Hall, K. Giles, J. Hoyes, C. V. Robinson, B.T. Ruotolo, Collision cross sections of proteins and their complexes: A calibration framework and database for gas-phase structural biology, Anal. Chem. 82 (2010) 9557–9565.

[35] A. Konijnenberg, J.F. van Dyck, L.L. Kailing, F. Sobott, Extending native mass spectrometry approaches to integral membrane proteins, Biol. Chem. 396 (2015) 991–1002.

[36] Y. Sun, S. Vahidi, M.A. Sowole, L. Konermann, Protein Structural Studies by

Traveling Wave Ion Mobility Spectrometry: A Critical Look at Electrospray Sources and Calibration Issues, J. Am. Soc. Mass Spectrom. 27 (2016) 31–40.

[37] A.B. Kanu, P. Dwivedi, M. Tam, L. Matz, H.H. Jr. Hill, Ion mobility–mass spectrometry, J. Mass Spectrom. 43 (2008) 1–22.

[38] J.P. Williams, M. Kipping, J.P.C. Vissers, Traveling Wave Ion Mobility Mass Spectrometry. Proteomic and Biopharmaceutical Applications, Waters. Application Notes.

[39] A.A. Shvartsburg, R.D. Smith, Fundamentals of traveling wave ion mobility spectrometry, Anal. Chem. 80 (2008) 9689–9699.

[40] D.J. Harvey, C.A. Scarff, M. Edgeworth, K. Pagel, K. Thalassinos, 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.

[41] 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.

[42] D.J. Harvey, C.A. Scarff, M. Edgeworth, W.B. Struwe, K. Pagel, K. Thalassinos, M. Crispin, J. Scrivens, Travelling-wave ion mobility and negative ion fragmentation of high mannose N-glycans, J. Mass Spectrom. 51 (2016) 219–235.

[43] D. Bitto, D.J. Harvey, S. Halldorsson, K.J. Doores, L.K. Pritchard, J.T. Huiskonen, T.A. Bowden, M. Crispin, Determination of N-linked Glycosylation in Viral Glycoproteins by Negative Ion Mass Spectrometry and Ion Mobility, in: B. Lepenies (Ed.), Carbohydrate-Based Vaccines Methods Protoc., Springer New York, 2015: pp. 93–121.

[44] M. Guttman, K.K. Lee, Site-Specific Mapping of Sialic Acid Linkage Isomers by Ion Mobility Spectrometry, Anal. Chem. 88 (2016) 5212–5217.

[45] Y. Pu, M.E. Ridgeway, R.S. Glaskin, M.A. Park, C.E. Costello, C. Lin, Separation and Identification of Isomeric Glycans by Selected Accumulation-Trapped Ion Mobility Spectrometry-Electron Activated Dissociation Tandem Mass Spectrometry, Anal. Chem. 88 (2016) 3440–3443.

[46] 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.

[47] J. Postigo, M. Iglesias, D. Cerezo-Wallis, A. Rosal-Vela, S. García-Rodríguez, M. Zubiaur, J. Sancho, R. Merino, J. Merino, Mice deficient in CD38 develop an attenuated form of collagen type II-induced arthritis, PLoS One. 7 (2012) e33534.

[48] 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) R113.

[49] 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.

[50] 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.

[51] C. Larriba, C.J. Hogan Jr., Free molecular collision cross section calculation methods for nanoparticles and complex ions with energy accommodation, J. Comput. Phys. 251 (2013) 344–363.

[52] C. Larriba, C.J. Hogan Jr., Ion mobilities in diatomic gases: Measurement versus prediction with non-specular scattering models, J. Phys. Chem. A. 117 (2013) 3887–3901.

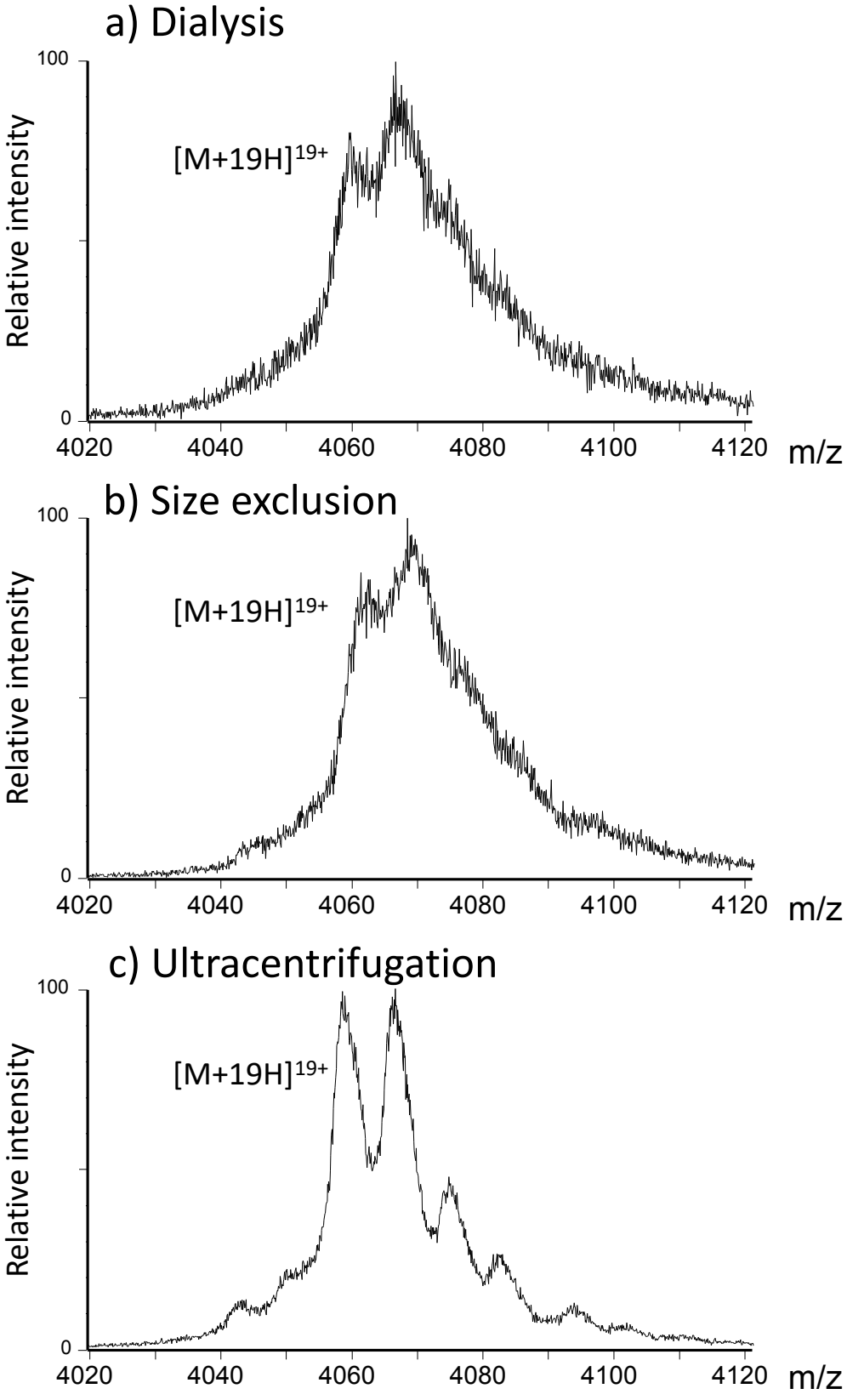
[53] www.glycam.org. The University of Georgia. Complex Carbohydrate Research Center (CCRC).

[54] J. Marcoux, T. Champion, O. Colas, E. Wagner-Rousset, N. Corvaïa, A. Van Dorsselaer, A. Beck, S. Cianférani, Native mass spectrometry and ion mobility characterization of trastuzumab emtansine, a lysine-linked antibody drug conjugate, Protein Sci. 24 (2015) 1210–1223.

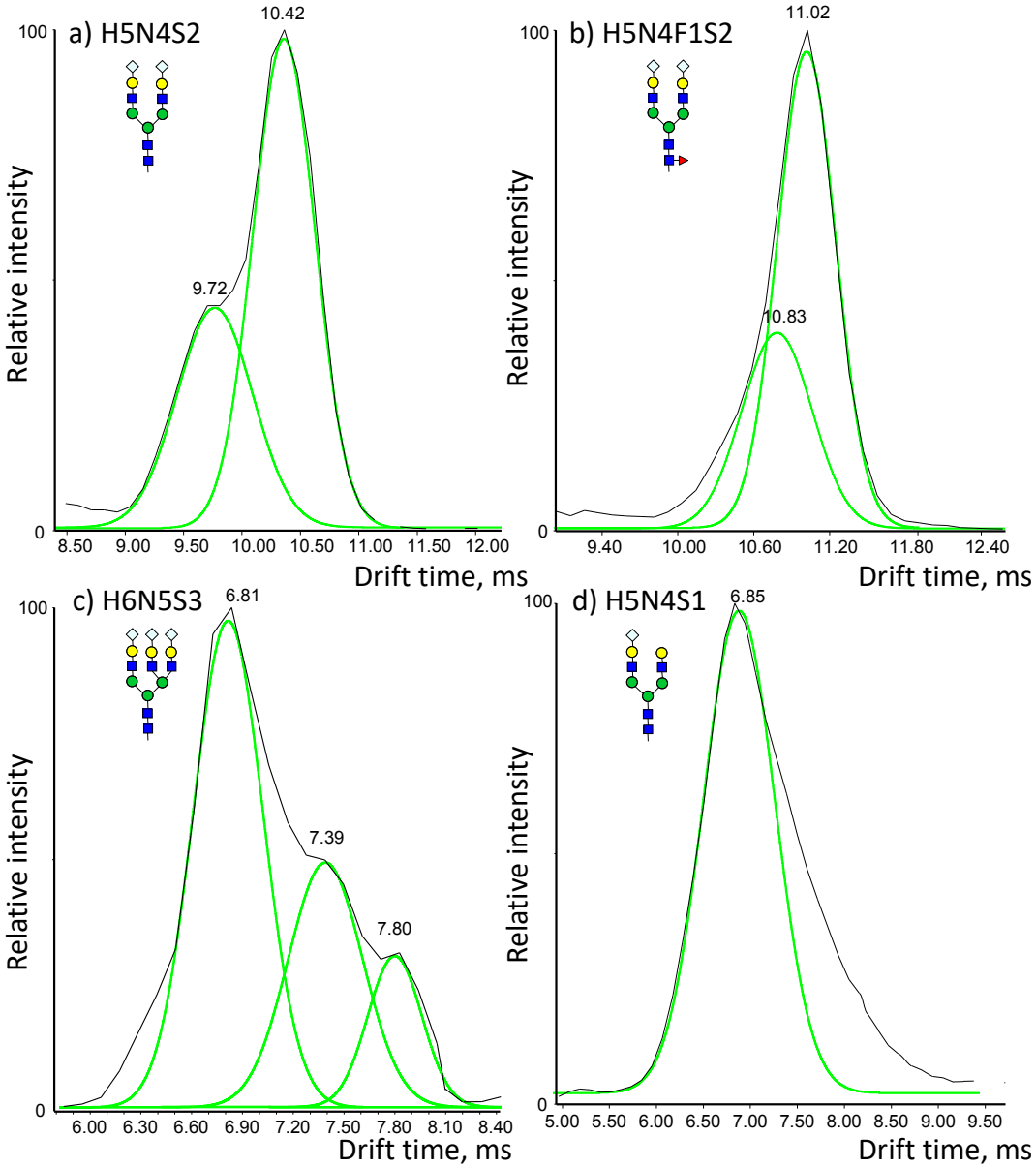
[55] R.A. Feelders, G. Vreugdenhil, G. de Jong, A.J.G. Swaak, H.G. van Eijk, Transferrin microheterogeneity in rheumatoid arthritis. Relation with disease activity and anemia of chronic disease, Rheumatol. Int. 12 (1992) 195–199.

[56] www.functionalglycomics.org. Consortium for Functional Glycomics. Functional Glycomics Gateway.

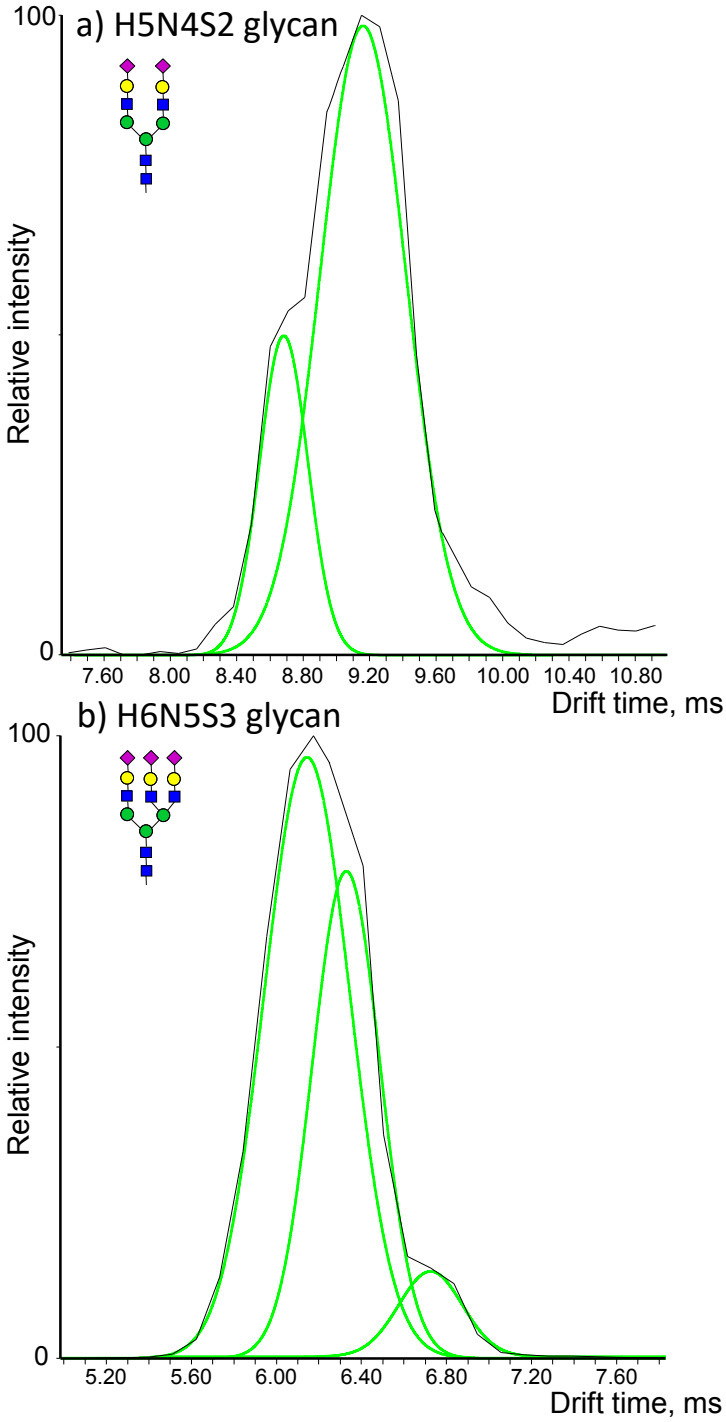
Supplementary Figures



Supplementary Figure 1: Mass spectra of the ion with +19 of intact mTf after desalting with three different methods: a) dialysis, b) size exclusion columns, and c) ultracentrifugation



Supplementary Figure 2: Arrival time distributions for several glycans of mTf standard: a) H5N4S2, b) H5N4F1S2, c) H6N5S3 and d) H5N4S1. The symbols used for the glycan representation follow the Consortium for Functional Glycomics (CFG) rules [56]. H: hexose; N: N-acetylhexosamine; F: fucose; S: sialic acid (in this case, all S are N-glycolylneuraminic acid).



Supplementary Figure 3: Arrival time distributions for a) H5N4S2 glycan and b) H6N5S3 glycan of hAGP. The symbols used for the glycan representation follow the Consortium for Functional Glycomics (CFG) rules [56]. H: hexose; N: N-acetylhexosamine; F: fucose; S: sialic acid (in this case, all S are N-acetylneuraminic acid).

Capítulo 5. Resultados y discusión

La **glicómica** se ha convertido en una de las ciencias ómicas de mayor interés, en gran parte debido a la importancia de los **carbohidratos** en multitud de procesos biológicos y a su implicación en un gran número de enfermedades. En la última década, los trabajos científicos relacionados con la glicómica han aumentado considerablemente, poniendo de manifiesto la gran novedad e impacto que posee esta ciencia ómica en campos como el biomédico, biotecnológico o farmacológico.

En la actualidad, la **espectrometría de masas (MS)** es la técnica analítica más utilizada en el estudio de la glicosilación de proteínas (**glicoproteómica**). Esto es debido a su elevada capacidad para obtener información estructural de biomoléculas, y por consiguiente, cualquier trabajo desarrollado hoy en día que tenga como objetivo estudiar la glicosilación de una proteína, depende, en gran medida, de la MS para alcanzar su objetivo satisfactoriamente.

El estudio de la glicosilación de proteínas por MS se puede llevar a cabo a tres niveles diferentes: analizando la glicoproteína intacta, los glicopéptidos y/o los glicanos, tal y como se ha mencionado en el apartado 1.3 de esta tesis doctoral. De todos ellos, el análisis de los **glicopéptidos** proporciona varias ventajas, ya que se obtiene información de la composición de los glicanos de cada punto de glicosilación, así como del grado de ocupación. No obstante, el análisis de los glicopéptidos obtenidos tras la digestión enzimática de una glicoproteína no se puede llevar a cabo con éxito mediante infusión directa del digesto en el espectrómetro de masas, ya que los péptidos presentes en el digesto se ionizan con mayor facilidad, suprimiendo la ionización de los glicopéptidos. Por tanto, es necesaria una técnica de separación previa a la detección por MS.

Por estos motivos, esta tesis doctoral se ha centrado principalmente en el estudio de la glicosilación de diferentes proteínas a nivel glicopeptídico, utilizando la cromatografía de líquidos capilar y la electroforesis capilar acopladas a la espectrometría de masas (**CapLC-MS** y **CE-MS**, respectivamente), utilizando un analizador de tiempo de vuelo de aceleración

ortogonal (**oa-TOF**). Ambas técnicas se engloban dentro de las llamadas técnicas de separación **miniaturizadas** y, como ya se ha explicado en la introducción, proporcionan diferentes ventajas que las hacen excelentes para el análisis de la glicosilación en muestras biológicas.

No obstante, la determinación de la composición de glicanos presentes en una proteína puede no ser suficiente para detectar la presencia y/o evolución de un proceso patológico. De hecho, se ha sugerido que en ciertas enfermedades, además de una modificación en la composición total de glicanos de algunas proteínas, puede existir una alteración de la abundancia relativa de ciertos glicanos isómeros [251–254]. En este sentido, la **movilidad iónica acoplada a la espectrometría de masas (IM-MS)**, ha suscitado gran interés como una técnica con potencial para la separación de isómeros, tal y como han demostrado varios autores [140–143, 152–155]. Por este motivo, en esta tesis doctoral, también se ha evaluado la eficacia del IM-MS para la separación de isómeros glicoconjugados debidos al tipo de enlace del ácido siálico, tanto a nivel de glicoproteína intacta como a nivel de glicopéptidos y glicanos.

5.1. Estudio de la glicosilación de glicoproteínas por cromatografía de líquidos capilar acoplada a la espectrometría de masas

5.1.1. Desarrollo y optimización de la metodología analítica

Para poder establecer y optimizar la metodología analítica basada en CapLC-MS, se escogió la **transferrina humana (hTf)**, una glicoproteína de interés biológico implicada en varios procesos patológicos, cuya glicosilación, a priori, es relativamente sencilla. La hTf muestra tres puntos de glicosilación, dos N y uno O, siendo los glicanos en los puntos de N-glicosilación de tipo complejo los que pueden verse alterados en presencia de ciertas enfermedades.

Optimización de la digestión enzimática de la hTf

Tal y como se detalla en el artículo 2.1, la digestión enzimática de la hTf con tripsina no resultó ser eficaz empleando condiciones de digestión en solución convencionales y ampliamente descritas en la bibliografía [82, 277–279] (0.5 M DTT, 0.73 M IAA, 50 mM NH_4HCO_3 pH 7.9, 37˚C de incubación durante 18h). Con estas condiciones convencionales se detectaron los péptidos y glicopéptidos de la hTf a muy baja concentración, lo que dificultaba la detección de las glicoformas minoritarias. Este bajo rendimiento de la digestión seguramente fue ocasionado por la estructura compacta de la hTf y la presencia de un elevado número de puentes disulfuro, lo que podía dificultar la hidrólisis enzimática de la glicoproteína [280, 281]. Con el objetivo de aumentar el rendimiento, se probaron diferentes reactivos y condiciones de digestión descritos en la bibliografía encaminados a facilitar la acción de la enzima [282–285]. No obstante, no fue hasta la utilización de un surfactante aniónico como coadyuvante, llamado RapiGest® (WATERS® S.A.), cuando se consiguió desnaturalizar la glicoproteína y obtener digestiones eficaces [286–290]. La principal ventaja de este surfactante es su compatibilidad con la cromatografía de líquidos y la espectrometría de masas tras acidificar la disolución. Esta

acidificación provoca la hidrólisis del surfactante, generando un producto no volátil que se elimina por precipitación, y otro que permanece en el digesto pero que no interfiere en el análisis por LC-MS (ver Figura 5.1).

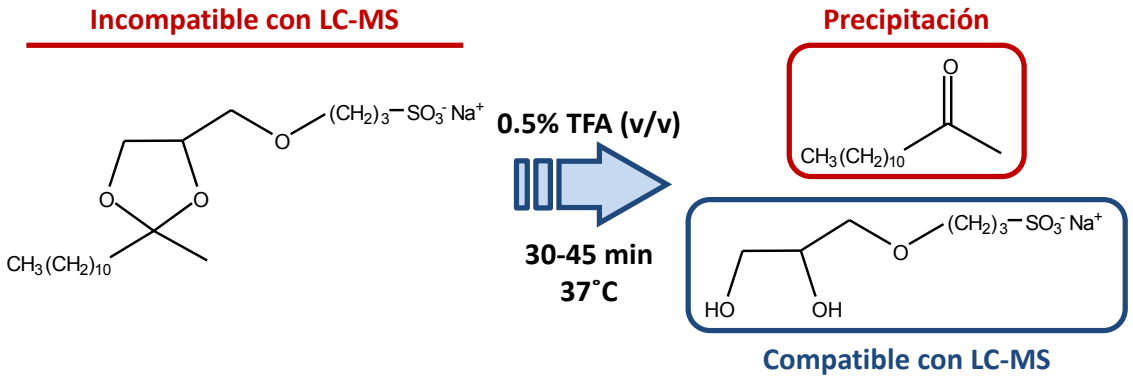


Figura 5.1: Hidrólisis del surfactante RapiGest® tras la acidificación de la disolución.

La eficacia de este surfactante a la hora de mejorar el rendimiento de la digestión de proteínas ya había sido descrita previamente por algunos autores [286–290]. No obstante, las condiciones para la hidrólisis del surfactante recomendadas por el fabricante (ver Figura 5.1) resultaron no ser adecuadas para el análisis de glicoproteínas, dado que producían una desialilación parcial de los glicopéptidos de la hTf (artículo 3.1). Para evitar esta pérdida de ácidos siálicos en la hTf, en esta tesis doctoral se evaluaron diferentes condiciones de hidrólisis. Se compararon dos ácidos, el ácido trifluoroacético (TFA) y otro ácido más débil, como el ácido fórmico (HFor), a diferentes concentraciones, además de modificar el tiempo de reacción. La Figura 5.2 muestra los cromatogramas de iones extraídos (EICs) de las glicoformas H5N4S2 y H5N4S1 del glicopéptido N_{413} en las diferentes condiciones de acidificación y reacción para hidrolizar el surfactante.

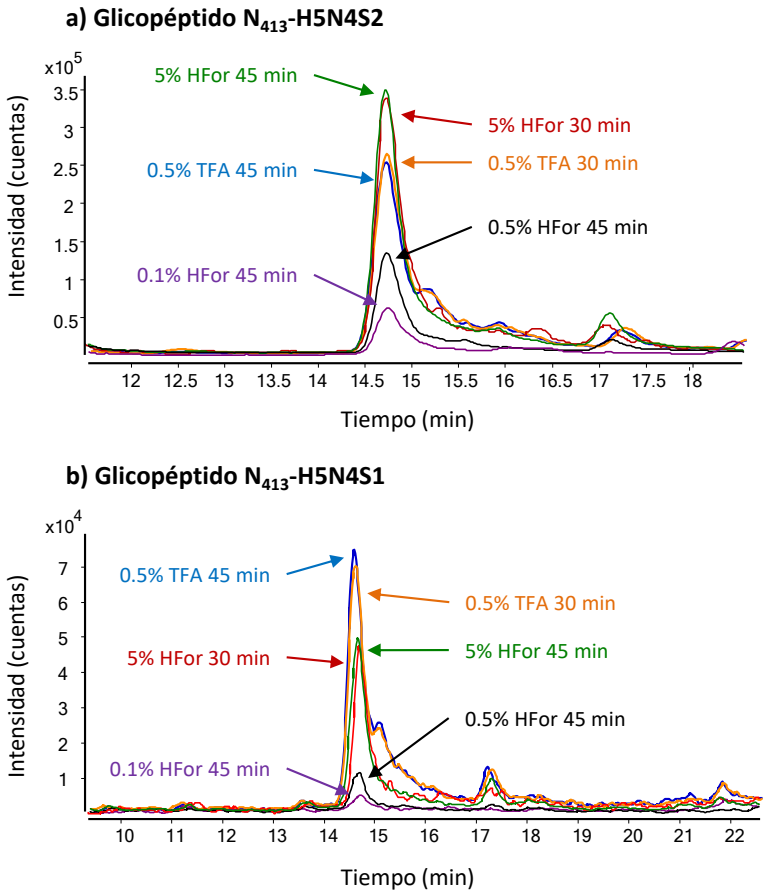


Figura 5.2: EICs de las glicoformas a) H5N4S2 y b) H5N4S1 del glicopéptido N_{413} de la hTf con diferentes condiciones de hidrólisis del surfactante.

Como se puede observar, al sustituir el TFA por el HFor a una concentración del 5% (v/v), la glicoforma H5N4S2 aumenta mientras la glicoforma H5N4S1 disminuye, lo que pone de manifiesto que al disminuir la acidez del medio se reduce la pérdida de ácidos siálicos. Sin embargo, cuando se disminuye la concentración final de HFor por debajo del 5% (v/v), la intensidad tanto de la glicoforma H5N4S2 como de la H5N4S1 disminuye considerablemente. Esta disminución puede ser debida a que el HFor es más débil que el TFA y a concentraciones menores del 5% (v/v) no es capaz de asegurar la hidrólisis completa del surfactante y, por tanto, la compatibilidad por MS. Se observó además que una concentración de HFor inferior al 5% (v/v) provocaba la disminución de la intensidad tanto de las diferentes glicoformas como de los péptidos del digesto, lo que impedía detectar las glicoformas minoritarias. Por ello, se

escogió una concentración de HFor del 5% (v/v) como la más adecuada para la hidrólisis del surfactante.

A continuación, se decidió comprobar la influencia del tiempo de reacción (30 o 45 min) en la desialilación. Tal y como se muestra en la Figura 5.2, no se observa un cambio significativo en el área de ninguna de las dos glicoformas y, en general, la intensidad de ambos picos se mantiene igual, independientemente del tiempo de reacción. Consecuentemente, en esta tesis doctoral (artículo 3.1) se seleccionaron como condiciones óptimas para la hidrólisis del surfactante, minimizando la desialilación de los glicopéptidos y sin provocar una disminución general de la sensibilidad, una concentración de HFor del 5% (v/v) y un tiempo de reacción de 30 min. Estas condiciones han sido las empleadas en todos los estudios posteriores al artículo 2.1.

Optimización de las condiciones del sistema CapLC-MS

El análisis por CapLC-MS de los péptidos y glicopéptidos de la hTf se llevó a cabo con una columna cromatográfica C18 (Zorbax 300SB-C18, 3.5 µm diámetro de partícula, 300 Å poro, 150 × 0.3 mm L_T × i.d., Agilent Technologies) y una fase móvil H_2O:ACN con 0.1% HFor (v/v). No obstante, a fin de mejorar la separación y detección, se optimizaron diferentes parámetros cromatográficos como el gradiente de fase móvil y el volumen de inyección. En cuanto al espectrómetro de masas, se utilizó un analizador de tiempo de vuelo de aceleración ortogonal (oa-TOF), del cual se optimizó principalmente el voltaje del *fragmentor* para mejorar la sensibilidad sin producir una pérdida de los ácidos siálicos. Esta optimización se detalla en el artículo 2.1, aun así, las condiciones óptimas se indican a continuación: volumen inyección: 150 nL; gradiente: de 10% a 60% (v/v) de ACN 0.1% HFor (v/v) en 45 min, seguido de un 60% a 100% (v/v) (5 min), 100% (v/v) (10 min), de 100% a 10% (v/v) (5 min) y 10% (v/v) (10 min); voltaje del *fragmentor*: 215 V. En la Figura 5.3 se pueden observar los EICs de las glicoformas mayoritarias y de algunas glicoformas minoritarias detectadas para ambos N-glicopéptidos (a),

así como los EICs del O-glicopéptido y de algunos de los péptidos observados (b). Utilizando RapiGest® como surfactante y CapLC-MS como técnica de análisis, de los dos N-glicopéptidos originados tras la digestión enzimática, se consiguieron detectar tanto las glicoformas mayoritarias como las minoritarias, así como cerca del 100% de los péptidos trípticos del digesto. Sin embargo, como se aprecia en la Figura 5.3-a, no existe separación entre las diferentes glicoformas de un mismo glicopéptido. En la Tabla 5.1, se recogen las áreas normalizadas (A_{norm}) de todas las glicoformas detectadas para ambos N-glicopéptidos en un patrón de hTf con la metodología establecida. Las glicoformas observadas para los glicopéptidos provienen de diferentes glicoformas intactas de la hTf (S2-S6) tal y como se detalla en la Figura 1.8 del apartado 1.2.1.

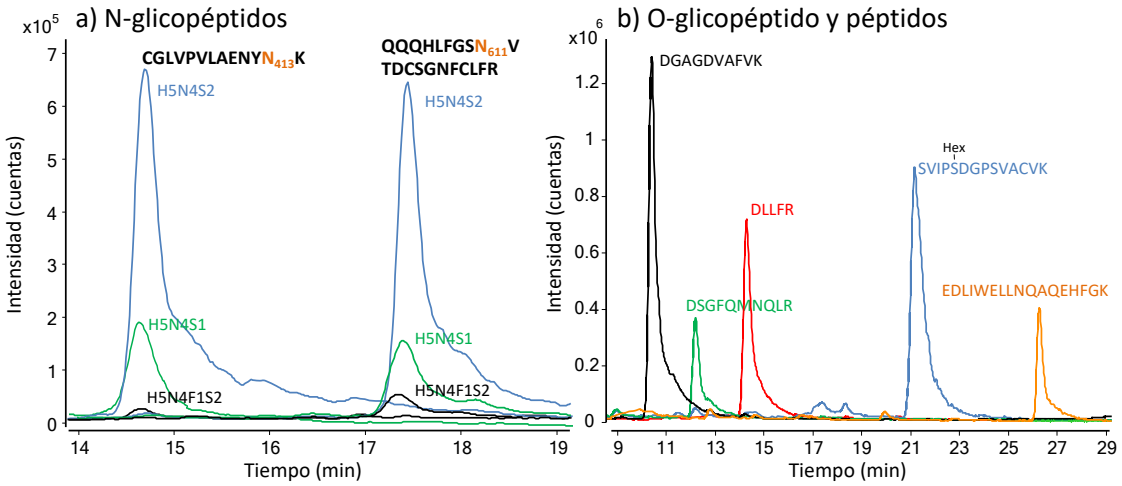
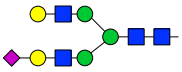
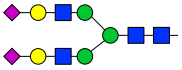
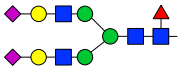
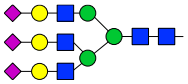
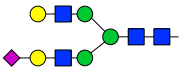
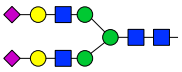
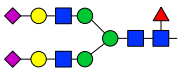
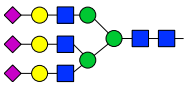


Figura 5.3: EICs a) de las glicoformas mayoritarias y algunas minoritarias detectadas de ambos N-glicopéptidos de la hTf y b) del O-glicopéptido y algunos péptidos del digesto de un patrón de hTf.

Tabla 5.1: Masa teórica (M_{teo}) y experimental (M_{exp}), error de masa, cargas observadas, tiempo de retención (t_R) y área normalizada (A_{norm}) para las diferentes glicoformas detectadas de ambos N-glicopéptidos en un patrón de hTf por CapLC-MS.

Glico-péptido	Glicoforma	M_{teo} (Da)	M_{exp} (Da) (n=6)	Error (ppm)	Cargas observadas	t_R (min) (n=6)	RSD (%) (t_R)	A_{norm}[a] (%) (n=6)	RSD (%) (A_{norm})
N_{413}	Péptido	-	-	-	-	-	-	-	-
	H5N4S1 	3389.4213	3389.4130	2.4	+2, +3, +4	14.3	1.1	**17.0**	3.3
	H5N4S2 	3680.5167	3680.5081	2.3	+2, +3, +4	14.4	1.1	**70.6**	3.2
	H5N4F1S2 	3826.5746	3826.5796	1.3	+3, +4	14.3	1.9	**1.2**	13.9
	H6N5S3 	4336.7443	4336.7527	1.9	+4	14.3	1.3	**1.0**	7.2
N_{611}	Péptido	2514.1169	2514.1022	5.8	+2, +3	17.8	0.9	**0.6**	13.2
	H5N4S1 	4427.7939	4427.7875	1.4	+3, +4	16.9	1.0	**14.5**	5.8
	H5N4S2 	4718.8893	4718.8838	1.2	+3, +4, +5	17.1	0.9	**74.5**	1.6
	H5N4F1S2 	4864.9472	4864.9336	2.8	+3, +4	16.9	0.9	**4.8**	2.2
	H6N5S3 	5375.1169	5375.1201	0.6	+4, +5	17.1	0.8	**1.9**	5.0

[a]Las áreas normalizadas se calcularon como: (Área glicoforma / Área péptido DGAGDVAFVK) x 100.

Los símbolos utilizados para la representación de los glicanos siguen las reglas sugeridas por el *Consortium for Functional Glycomics* (CFG) [13].

Para poder comparar los cambios de glicosilación entre diferentes muestras se debe tener en cuenta que ciertos factores como la cantidad de proteína presente en una muestra biológica, el rendimiento de la digestión o las condiciones del espectrómetro de masas pueden afectar al área absoluta de las diferentes glicoformas. Por este motivo, es necesario normalizar las áreas

de las diferentes glicoformas respecto al área de un péptido del propio digesto (A_{norm}). De este modo, gran parte de las variaciones entre muestras ocasionadas por los factores mencionados anteriormente, se corrigen y, consecuentemente, se pueden asignar las diferencias entre muestras a cambios de glicosilación de la glicoproteína estudiada. En concreto, se escogió como péptido de referencia para el cálculo de las áreas normalizadas aquel que tenía un tiempo de migración o retención y un área absoluta parecidos al de las glicoformas mayoritarias en cada caso. En la Tabla 5.2 se recogen los péptidos de referencia utilizados en los diferentes estudios de esta tesis doctoral.

Tabla 5.2: Péptidos de referencia utilizados para el cálculo de las áreas normalizadas en los diferentes estudios.

Estudio	Péptido de referencia	Glicoproteína	Artículo
Alcoholismo	DGAGDVAFVK	Transferrina humana	2.1
Defectos congénitos de la glicosilación por CapLC	DYELLCLDGTR	Transferrina humana	2.2
Artritis inducida por colágeno	CDEWSIISEGK	Transferrina de ratón	2.3
Defectos congénitos de la glicosilación por CE	SAGWNIPIGLLYCDLPEPR	Transferrina humana	3.1
Modelos de migración[a]	MEVGQQAVEVWQGLALLSEAVLR	Eritropoyetina humana recombinante	3.2

[a]En este estudio el péptido de referencia se utilizó para calcular el tiempo de migración relativo, pero se seleccionó siguiendo los mismos criterios que en los otros estudios.

Purificación por cromatografía de inmunoafinidad

Uno de los objetivos principales de esta tesis doctoral ha sido el desarrollo de una metodología analítica adecuada para caracterizar la glicosilación de glicoproteínas en muestras de suero. Debido a la gran complejidad de este tipo de muestras, aislar la glicoproteína de interés no es fácil, principalmente, por la presencia de otras proteínas mayoritarias. Dichas proteínas, como por ejemplo la albúmina, pueden interferir en el proceso de digestión o impedir la detección de la glicoproteína de interés por MS, tal y como se demostró anteriormente en nuestro grupo

de investigación para la hTf [291, 292]. Aunque existen *kits* comerciales que permiten eliminar las proteínas mayoritarias presentes en el suero, no suelen ser adecuados cuando se pretende analizar la glicoproteína de interés por MS, ya que la cantidad no eliminada puede interferir. Como ya se ha mencionado en el apartado 1.3, la **cromatografía de inmunoafinidad (IAC)**, debido a su gran especificidad, es la opción más adecuada para llevar a cabo un análisis focalizado (*targeted analysis*) de una proteína concreta en muestras biológicas. En este sentido, en el artículo 2.1 se evaluó la eficacia de dos columnas de inmunoafinidad *home-made*, activadas con el mismo anticuerpo policlonal anti-hTf. Estas dos columnas diferían en el tipo de soporte y en la química de unión entre el anticuerpo y el soporte. La columna de **Sílice-Hidrazida** estaba formada por un soporte de sílice derivatizado con grupos hidrazida para unir el anticuerpo a través de los carbohidratos de las cadenas pesadas [293]. Por otro lado, en la columna de **Sefarosa-CNBr** el soporte de sefarosa se encontraba derivatizado con grupos de bromuro de cianógeno (CNBr), donde el anticuerpo se unía a través de los grupos amino de las lisinas y argininas.

De las dos columnas elaboradas, la de Sefarosa-CNBr fue la que proporcionó mayores recuperaciones cuando se analizaron muestras patrón de 1500 μg mL^{-1} de hTf (∼44% (±5), n=3). Esta columna se empleó para purificar la transferrina humana del suero (artículos 2.1, 2.2 y 3.1), así como la transferrina de ratón (artículos 2.3 y 4.1), ya que ambas glicoproteínas comparten alrededor de un 72% de la secuencia aminoacídica.

A modo de ejemplo, en la Figura 5.4 se muestra la glicoforma H5N4S2 de ambos N-glicopéptidos en el patrón de hTf y en una muestra de suero control donde la hTf ha sido purificada previamente con la columna de inmunoafinidad de Sefarosa-CNBr. A pesar de la elevada complejidad de la muestra de suero, se consiguió aislar satisfactoriamente la hTf, obteniéndose áreas normalizadas muy parecidas para todas las glicoformas de los dos N-

glicopéptidos, tanto en el patrón como en la muestra de suero, demostrando así la eficacia de la columna de Sefarosa-CNBr.

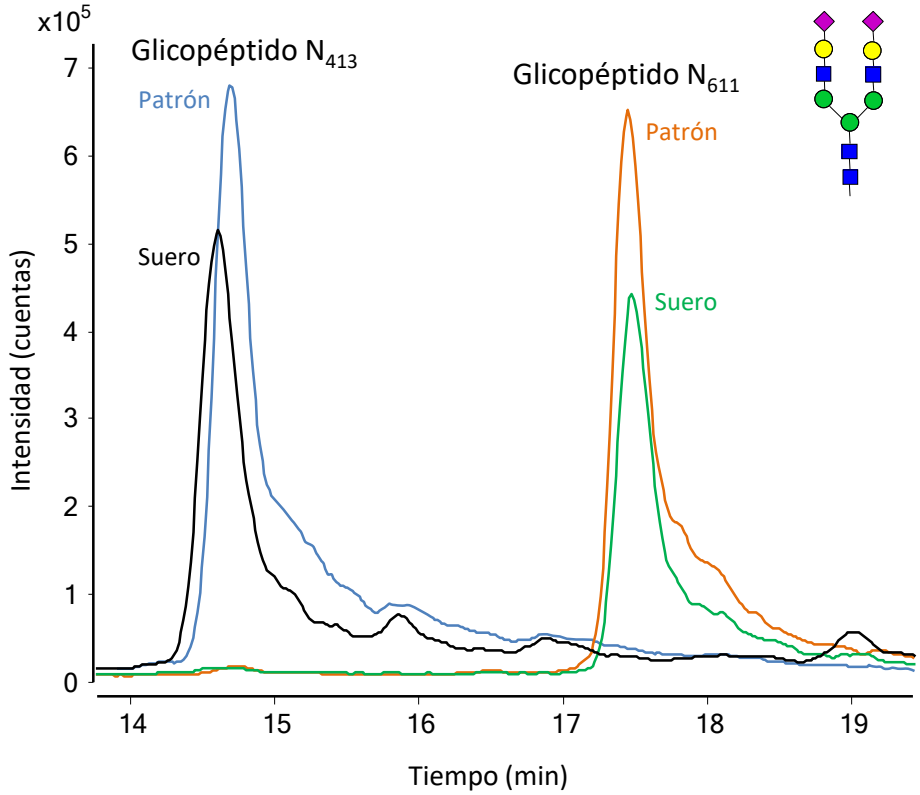


Figura 5.4: EICs de la glicoforma H5N4S2 en ambos glicopéptidos en un patrón de hTf de 1500 µg mL^{-1} y en una muestra de suero control. Los símbolos utilizados para la representación de los glicanos siguen las reglas sugeridas por el *Consortium for Functional Glycomics* (CFG) [13].

5.1.2. Transferrina humana en suero. Diagnóstico del alcoholismo crónico y los defectos congénitos de la glicosilación

Una vez establecido el método de purificación, digestión y análisis por CapLC-MS de los glicopéptidos de la hTf en muestras de suero, se procedió a evaluar el efecto que el **alcoholismo crónico** y los **defectos congénitos de la glicosilación (CDG)** causan en la glicosilación de esta proteína.

Análisis de muestras de suero de individuos con diferentes grados de alcoholismo

Como ya se ha comentado, la ingesta elevada de alcohol se traduce en un aumento de las glicoformas de la hTf menos glicosiladas. Existe cierta discrepancia en torno al mecanismo que provoca la alteración de la glicosilación de la hTf en individuos con un cierto grado de alcoholismo [31–33]. En concreto, algunos autores sugieren que el etanol afecta directamente al traslado o a la unión del glicano a la proteína cuando se está formando, hecho que provocaría que la N-glicosilación no se llegase a iniciar, dando lugar a una hTf deficiente en cadenas completas de N-glicanos [31, 32, 47, 48]. En cambio, otros autores sugieren que el etanol afecta al proceso de N-glicosilación, provocando alteraciones en las cadenas de N-glicanos de las glicoproteínas que se van formando en el aparato de Golgi, lo cual daría lugar a una hTf con cadenas de glicanos truncadas o desialiladas [31, 33, 45, 46]. Aunque aún en debate, se cree que ambos efectos son posibles en el alcoholismo crónico [32]. Con el estudio llevado a cabo y detallado en el artículo 2.1, se pretende caracterizar las glicoformas de la hTf de muestras de individuos con distintos grados de consumo de etanol para, por un lado determinar su validez para el diagnóstico y el seguimiento del alcoholismo crónico, y, por otro, aportar información sobre el mecanismo dominante responsable de la alteración de la glicosilación de la hTf.

Para realizar este estudio, se analizaron varias muestras de suero, aplicando la metodología desarrollada y explicada en el apartado anterior: 1 muestra control de un individuo abstemio, 1 muestra control de alcoholismo leve, 1 muestra control de alcoholismo grave y dos muestras de individuos cuyo grado de alcoholismo era desconocido. Los resultados obtenidos se recogen en la Tabla 5.3.

Tabla 5.3: Áreas normalizadas para las diferentes glicoformas de ambos N-glicopéptidos de la hTf en diferentes muestras de suero para el estudio del alcoholismo.

Alcoholismo		Control	Alcohólico leve	Alcohólico grave	Individuo 1	Individuo 2
N_{413}	Péptido	-	-	3.0	-	2.5
	H5N4	2.5	1.4	1.6	1.7	1.3
	H5N4S1	15.7	12.3	11.4	11.7	11.1
	H5N4S2	66.3	60.9	55.0	62.4	55.3
	H5N4F1S2	1.3	1.1	1.4	1.1	1.2
	H6N5S3	1.1	1.0	1.0	1.0	0.8
N_{611}	Péptido	0.8	2.9	8.5	2.0	7.0
	H5N4	0.5	0.5	0.4	0.5	0.4
	H5N4S1	13.8	11.1	12.0	11.0	12.0
	H5N4S2	74.1	71.9	67.0	72.5	68.6
	H5N4F1S2	4.2	3.6	4.4	3.9	4.6
	H6N5S3	1.9	1.0	1.5	1.8	1.2

Las áreas normalizadas se calcularon como: (Área glicoforma / Área péptido DGAGDVAFVK) x 100

Si se comparan las Tablas 5.1 y 5.3, el área normalizada de las glicoformas del suero control son muy parecidas a las obtenidas para las glicoformas del patrón de hTf, ya que ambas muestran una distribución de las glicoformas típica de una hTf no alterada (que corresponderían a las glicoformas de la proteína intacta predominantes: S2-S6, ver Figura 1.8 del apartado 1.2.1). Sin embargo, en las muestras de individuos con diferentes grados de alcoholismo, se pueden observar variaciones principalmente en dos glicoformas. En concreto, para la muestra con un grado de alcoholismo leve se observa una disminución de la glicoforma mayoritaria H5N4S2 y un aumento de la glicoforma péptido (aquella en la que no hay glicano), aunque únicamente en el glicopéptido N_{611} ya que el glicopéptido N_{413} se detecta con menor intensidad. Este cambio en la glicosilación se traduce, a nivel de glicoproteína intacta, en un aumento de las glicoformas intactas S0 y S2 (CDT, Figura 1.8 del apartado 1.2.1). Este mismo patrón de alteración se repite en las muestras con un grado de alcoholismo grave, pero con una disminución más pronunciada del área normalizada de la glicoforma H5N4S2 y un aumento mayor de la glicoforma péptido, que esta vez se detecta en ambos glicopéptidos. Por ejemplo, la glicoforma péptido para el glicopéptido N_{611} aumenta de un 0.8% a un 8.5% al pasar de un control a un alcohólico grave. Estos resultados refuerzan la teoría de que el etanol

afecta principalmente al transporte del glicano hacia la proteína impidiendo que se inicie la N-glicosilación, aumentando así las glicoformas CDT S0 y S2 de la glicoproteína intacta (ver Figura 1.8 apartado 1.2.1). Si el etanol afectase predominantemente al proceso de N-glicosilación una vez ya iniciado, hubiera habido un aumento de glicoformas truncadas o desialiladas como son la glicoforma H5N4S1 o H5N4. Tal como se observa en la Tabla 5.3 no es el caso, puesto que ambas glicoformas presentan áreas normalizadas muy similares en todos los casos.

Por otro lado, la metodología desarrollada también se aplicó para estimar el grado de alcoholismo de dos muestras desconocidas. Tal y como se puede observar en la Tabla 5.3, la primera muestra desconocida (Individuo 1) debe corresponder a un individuo con un grado de alcoholismo leve ya que la disminución de la glicoforma H5N4S2 es reducida, al igual que el aumento de la glicoforma péptido, y los valores son similares a los obtenidos para la muestra de suero control de individuos con un grado de alcoholismo leve. Sin embargo, en la muestra de Individuo 2 el aumento de la glicoforma péptido y la disminución de la glicoforma H5N4S2 son más acusados, al igual que en la muestra de suero control de individuos con un grado de alcoholismo severo. Por lo tanto, el método desarrollado no solo nos ha permitido determinar, de manera general, el grado de alcoholismo de muestras desconocidas, sino que también nos ha permitido aportar datos que pueden resultar útiles para conocer el mecanismo responsable de la alteración de la glicosilación de la hTf debido al alcoholismo. No obstante, sería necesario un número mayor de muestras para poder validar adecuadamente la metodología establecida como técnica para el diagnóstico y seguimiento del alcoholismo.

Estudio de muestras de suero de pacientes con diferentes tipos de CDG

En el siguiente trabajo de esta tesis doctoral, que se recoge en el artículo 2.2, se aplicó la metodología descrita anteriormente para estudiar el efecto de los CDG sobre la glicosilación de la hTf. Este nuevo estudio pretendía encontrar las glicoformas principales que permiten distinguir entre muestras de suero de individuos sanos y con CDG-I o CDG-II, con el fin de

obtener un panel novedoso de biomarcadores glicopeptídicos. En la Tabla 5.4 se recogen los resultados obtenidos tras analizar varias muestras de suero de individuos sanos (9 muestras control) y muestras de pacientes con CDG (5 muestras de CDG-I y 5 muestras de CDG-II), utilizando la metodología CapLC-MS desarrollada y explicada en el apartado anterior. En la Figura 5.5 se muestran, a modo de ejemplo, los EICs de las glicoformas H5N4S2 y H5N4S1 de los glicopéptidos N_{413} y N_{611} de la hTf en tres muestras de suero: control, CDG-I y CDG-II. También se incluye en esta figura los espectros de masas de las glicoformas mencionadas para la muestra de suero control.

Los resultados obtenidos demuestran la robustez del método, ya que al aplicar la metodología desarrollada a estas nuevas muestras de suero, fue posible aislar la hTf, digerirla y obtener información sobre las diferentes glicoformas de ambos glicopéptidos. A pesar de no obtener separación entre glicoformas de un mismo glicopéptido (igual que ocurría en el estudio del alcoholismo), ambos glicopéptidos se separan perfectamente pudiendo cuantificar de manera relativa las diferentes glicoformas en cada punto de N-glicosilación.

Tabla 5.4: Áreas normalizadas para las diferentes glicoformas de ambos glicopéptidos de la hTf en diferentes muestras de suero para el estudio de los CDG.

CDG		Glicopéptido N_{413}							Glicopéptido N_{611}						
Muestra		Péptido	H5N4	H5N4S1	H5N4S2	H5N4 F1S2	H6N5S2	H6N5S3	Péptido	H5N4	H5N4S1	H5N4S2	H5N4 F1S2	H6N5S2	H6N5S3
1	Control_1	0.3	0.8	6.1	76.3	1.3	1.8	1.2	3.5	0.5	5.7	91.0	4.5	0.7	2.8
2	Control_2	2.2	0.6	4.6	73.0	1.8	1.5	3.4	4.7	0.8	3.9	92.0	5.5	0.7	2.2
3	Control_3	2.8	1.1	4.3	103.7	1.4	2.0	4.7	6.0	1.3	4.9	116.1	6.0	0.9	3.9
4	Control_4	0.5	0.7	7.3	75.7	1.5	1.8	2.3	6.1	0.5	6.2	87.3	4.7	0.9	3.4
5	Control_5	0.3	0.4	5.8	77.3	1.8	1.6	2.5	5.5	0.5	5.6	94.7	5.1	0.8	3.4
6	Control_6	2.9	1.2	4.3	55.8	1.4	1.2	3.4	7.6	1.1	5.5	101.0	7.5	0.8	3.4
7	Control_7	2.9	0.4	4.6	75.6	1.9	1.6	3.9	5.2	0.9	5.5	93.6	6.1	0.8	2.7
8	Control_8	4.4	0.8	4.4	58.6	1.1	1.8	4.1	6.2	1.0	5.4	98.8	6.2	0.9	4.4
9	Control_9	4.4	1.3	3.7	68.3	1.5	1.5	4.7	6.1	1.2	4.6	95.5	6.2	1.2	4.4
10	PGM1-CDG	14.6	1.3	7.3	67.8	2.5	1.5	1.6	14.6	0.5	6.2	63.6	5.9	0.7	1.4
11	DPM1-CDG	19.1	-	8.3	51.5	1.5	0.7	1.6	20.8	0.3	5.0	52.2	7.1	0.7	1.0
12	RFT1-CDG	59.4	-	3.3	45.1	3.3	1.9	1.7	46.7	0.4	4.7	45.0	4.6	0.4	1.3
13	DPAGT1-CDG	21.7	0.7	8.0	55.6	0.9	0.8	1.6	20.5	0.5	5.8	59.2	3.9	0.8	2.1
14	PMM2-CDG	66.3	-	6.0	46.8	2.2	0.5	2.0	59.5	0.2	2.9	34.3	3.6	0.6	0.5
15	ATP6V0A2-CDG	2.1	2.9	19.4	43.7	0.6	1.0	1.7	5.9	0.3	10.4	52.6	4.8	1.0	1.7
16	CDG-IIx (sepsis)	-	2.0	14.4	72.6	1.6	2.2	1.1	5.9	0.6	9.5	68.6	7.6	1.1	2.3
17	CDG-IIx	14.0	1.6	17.4	48.6	1.1	1.4	1.4	12.1	0.7	10.5	66.9	6.7	1.1	1.9
18	COG8-CDG_1	0.7	1.5	18.1	61.2	1.1	1.0	1.7	6.1	0.7	11.9	81.5	2.3	0.9	1.9
19	COG8-CDG_2	0.7	1.2	17.3	60.9	1.0	1.1	1.4	5.7	0.6	11.5	83.4	2.6	0.7	2.0

Las áreas normalizadas se calcularon como: (Área glicoforma / Área péptido DYELLCLDGTR) x 100. Naranja: Controles. Verde: CDG-I. Rojo: CDG-II.

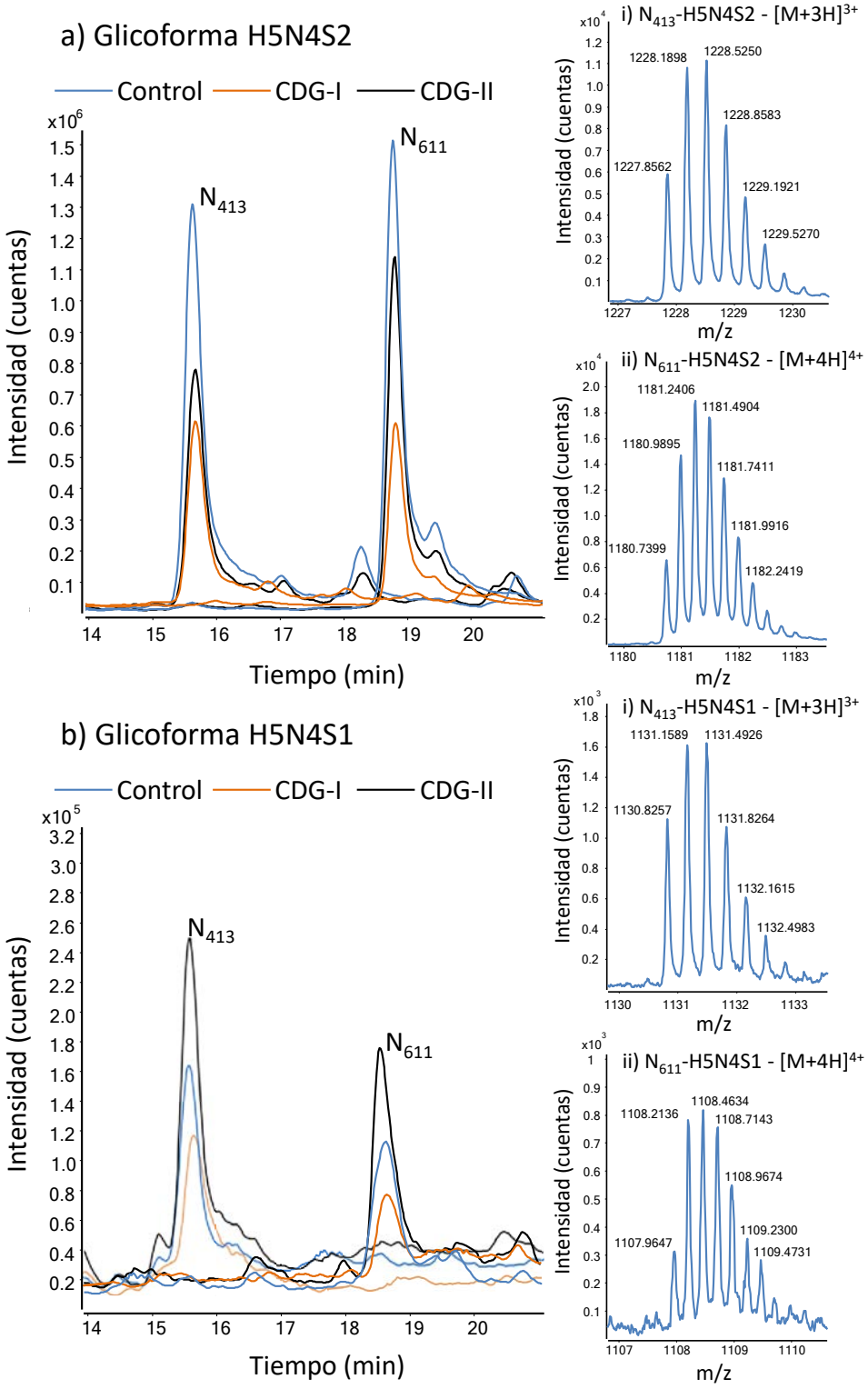


Figura 5.5: EIC de las glicoformas a) H5N4S2 y b) H5N4S1 de ambos glicopéptidos de la hTf en una muestra de suero control, de CDG-I y de CDG-II. i, ii) Espectros de masas de las glicoformas H5N4S2 y H5N4S1 para ambos N-glicopéptidos en una muestra control.

Para facilitar la interpretación de los resultados recogidos en la Tabla 5.4, correspondiente a un número considerable de muestras y de glicoformas identificadas, en la Figura 5.6 se muestra un gráfico de barras con el área normalizada media y la desviación estándar (±s) de las diferentes glicoformas para las muestras control, de CDG-I y de CDG-II.

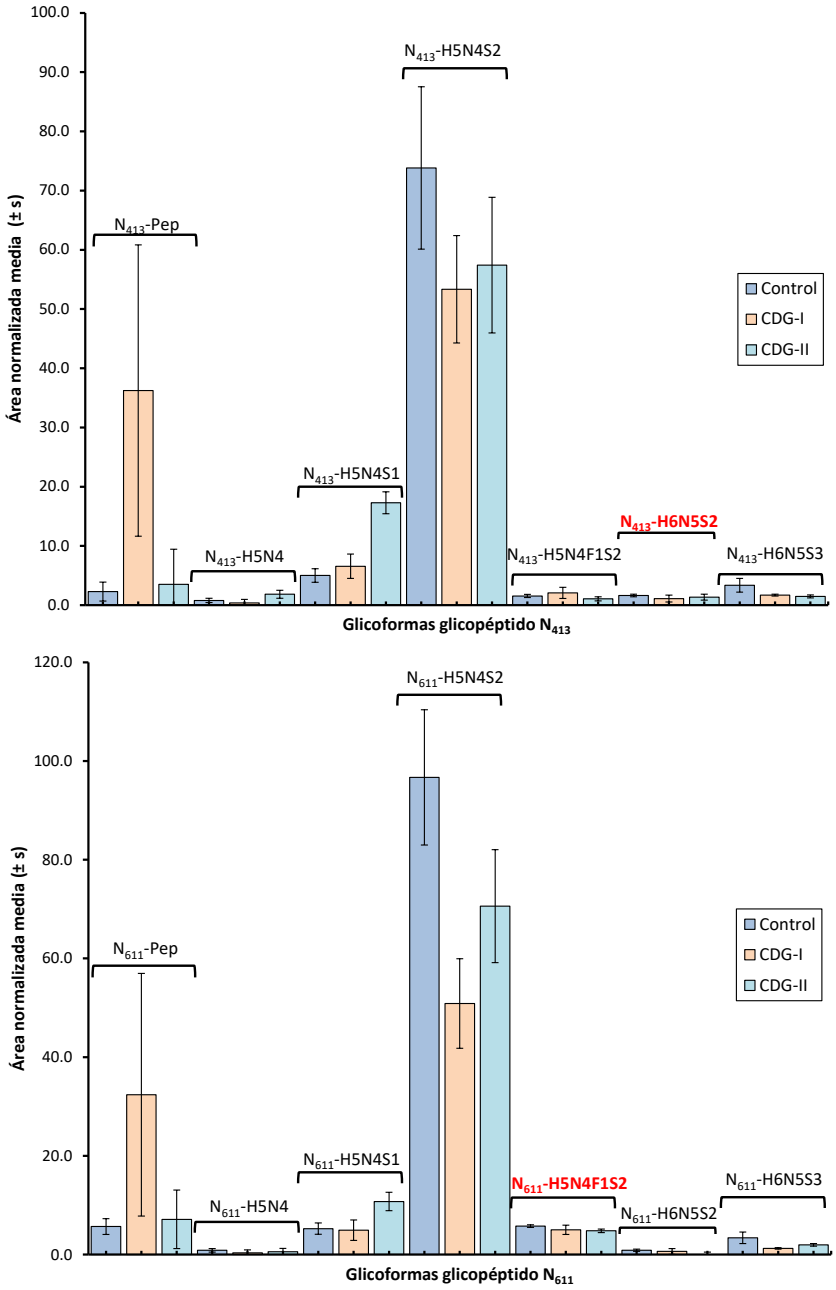


Figura 5.6: Gráfico de barras con el área normalizada media, y la correspondiente desviación estándar, de las diferentes glicoformas del a) glicopéptido N_{413} y b) el glicopéptido N_{611} de la hTf, para el conjunto de muestras control, de CDG-I y de CDG-II. En rojo se marcan las glicoformas cuya área normalizada media es considerada similar tras realizar un ANOVA de un factor.

Como se puede observar, en algunas glicoformas las desviaciones estándar obtenidas son elevadas, dada la gran heterogeneidad de las muestras analizadas dentro del mismo grupo (diferentes individuos, mutaciones, presencia de otras patologías, etc.). A pesar de ello, a partir de los datos disponibles, se puede obtener cierta información relevante. Por ejemplo, la glicoforma péptido (sin glicosilación) aumenta claramente en el caso de las muestras de CDG-I, mientras que no lo hace en las de CDG-II, lo cual es razonable, puesto que los CDG-I engloban todos aquellos defectos que tienen lugar **antes** de que llegue a iniciarse la N-glicosilación de las glicoproteínas. En cambio, en muestras de CDG-II, la glicoforma H5N4S1 aumenta, lo que también parece lógico si se tiene en cuenta que los CDG-II engloban todos aquellos defectos en la remodelación o síntesis del glicano **durante** la N-glicosilación, dando lugar a un aumento de glicoformas desialiladas o truncadas (La Figura 5.5 también permite observar las mismas tendencias).

Estudio estadístico y análisis multivariante del conjunto de datos

Aunque se pueden apreciar diferencias entre algunas glicoformas al observar la Figura 5.5 y 5.6, dada la complejidad de los datos obtenidos y la heterogeneidad de las muestras estudiadas, es aconsejable la utilización de estudios estadísticos y quimiométricos con el objetivo de conocer que glicoformas pueden ser las significativas para la diferenciación entre los tres grupos de muestras.

En primer lugar, para comparar las áreas normalizadas de cada glicoforma entre los diferentes grupos de muestras, se realizó un análisis de la varianza (ANOVA) de un solo factor. Con este tratamiento estadístico, se pudo detectar que dos de las catorce glicoformas analizadas (H6N5S2 del glicopéptido N_{413} y H5N4F1S2 del glicopéptido N_{611}) no eran importantes a la hora de diferenciar entre los tres grupos, debido a que su área normalizada permanecía prácticamente constante y, por tanto, no eran interesantes como posibles biomarcadores (glicoformas marcadas en rojo en la Figura 5.6).

Seguidamente, se decidió utilizar dos métodos quimiométricos que tuvieran en cuenta las relaciones entre las diferentes glicoformas. Inicialmente, se aplicó un método quimiométrico más sencillo, como es el **análisis por componentes principales (PCA)**. Con el PCA se pretendía estudiar de manera no supervisada si existían tendencias o grupos entre las diferentes muestras y, también, detectar la presencia de posibles *outliers* [214, 215].

En la Figura 5.7-a, se muestran los resultados obtenidos tras realizar el PCA al conjunto de datos. En dicha figura se muestra el grafico de *scores* (Figura 5.7-a (i)) para los dos componentes principales (PC). El PC 1 explica alrededor del 41% de toda la varianza, mientras que el PC 2 explica un 29%. A pesar de la elevada heterogeneidad entre las muestras de un mismo tipo (controles, CDG-I y CDG-II), en este estudio no supervisado, se pueden observar claramente tres grupos, los cuales corresponden precisamente a los tres tipos indicados anteriormente. Estos resultados demuestran, en primer lugar, que las diferencias entre los tres grupos son mayores que las variaciones observadas entre muestras de un mismo grupo. Adicionalmente, tal y como se observa en el gráfico de *scores*, el PC 1 permite diferenciar entre las muestras control y CDG-I, mientras que el PC 2 permite separar las muestras de CDG-II del resto. También es posible ver en el gráfico de scores que no hay ninguna muestra que se encuentre muy alejada del resto, confirmándose así la ausencia de *outliers*. Finalmente, el gráfico de *loadings* de la Figura 5.7-a (ii) proporciona información interesante de las variables, es decir, de las glicoformas de los glicopéptidos en este caso, y como éstas están relacionadas con los componentes principales (PC).

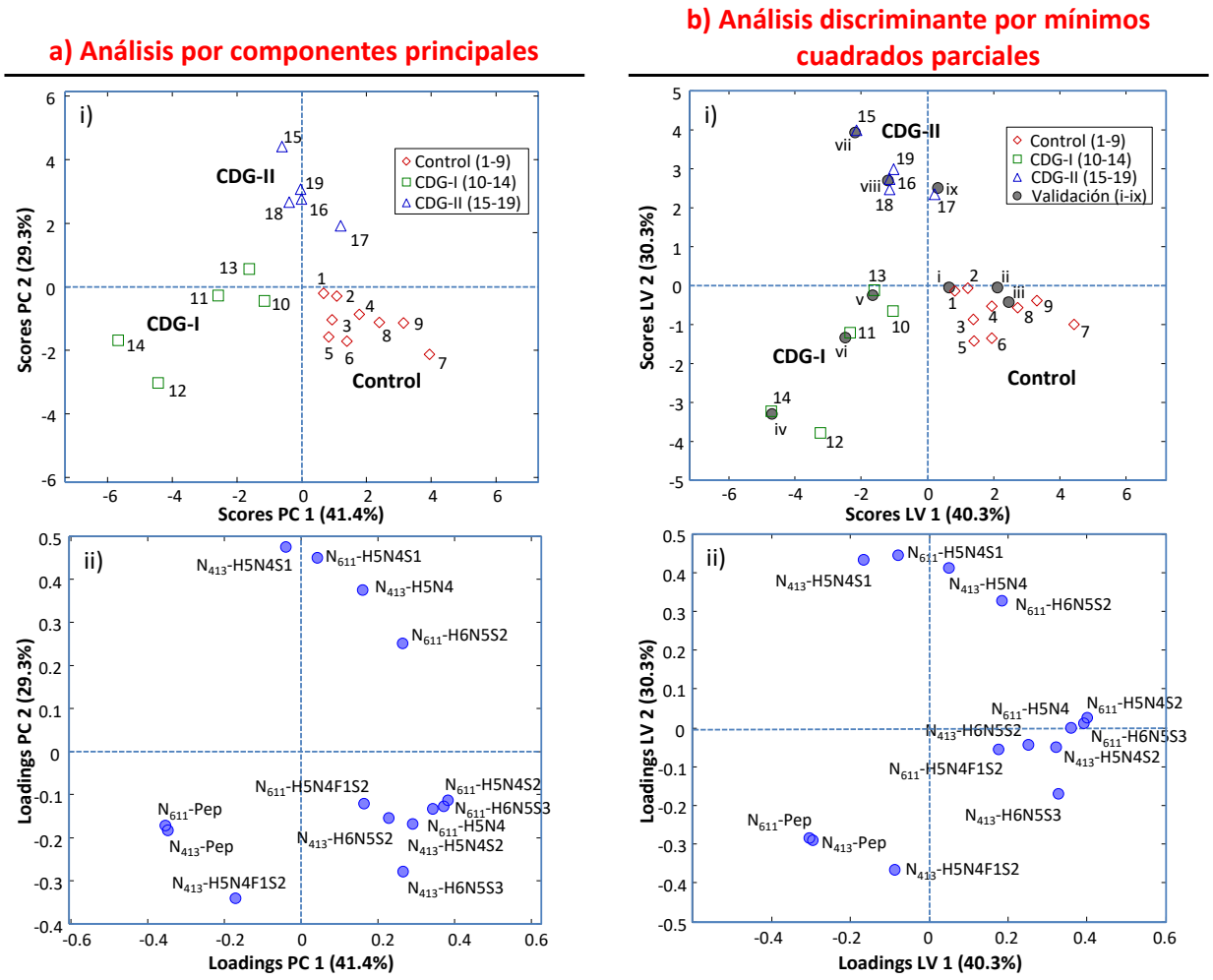


Figura 5.7: Gráficos de *scores* (i) y de *loadings* (ii) para a) el análisis de componentes principales (PCA) y b) el análisis discriminante por mínimos cuadrados parciales (PLS-DA), en el estudio de las muestras de CDG.

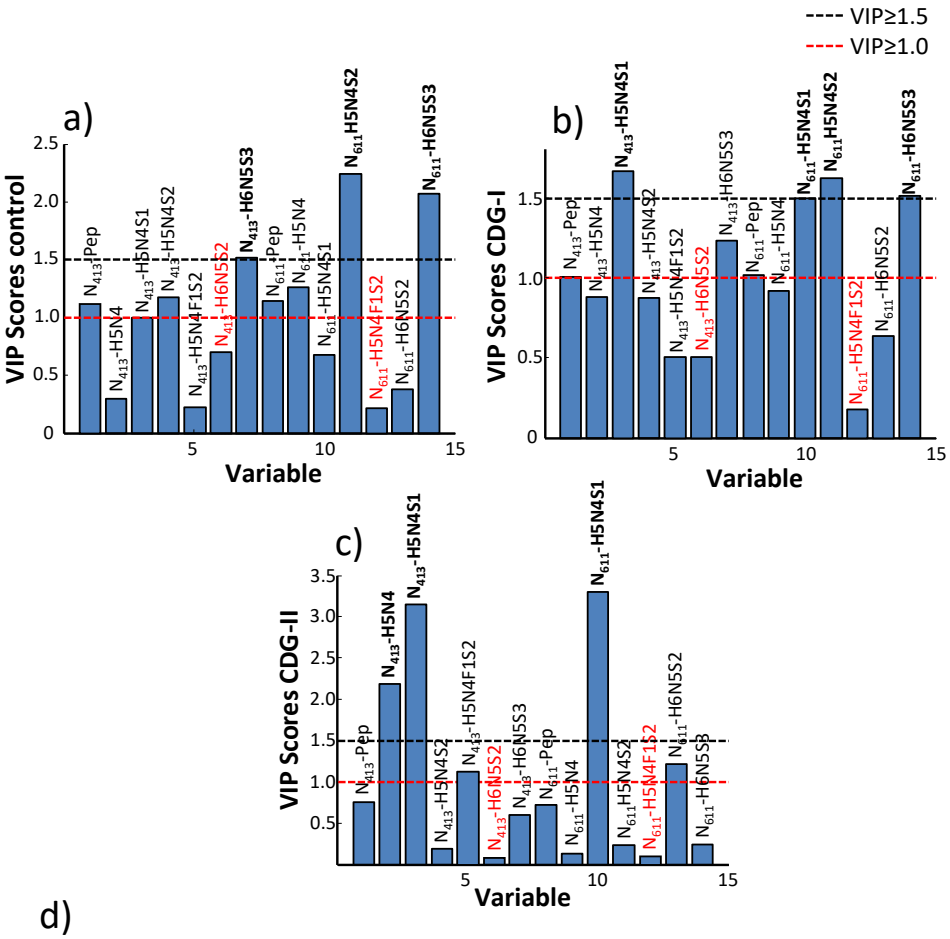
Por ejemplo, se puede deducir que las glicoformas H5N4S2 y H6N5S3, ambas del glicopéptido N_{611}, contribuyen a diferenciar entre las muestras control y CDG-I, debido a que su proyección en el eje de las X, lo que sería el PC 1, es elevada. Del mismo modo, las glicoformas péptido de ambos glicopéptidos también serían destacables en la diferenciación entre CDG-I y muestras controles (situadas en el lado negativo del eje de las X y con una proyección elevada en este eje). Por otro lado, la glicoforma H5N4S1 de ambos glicopéptidos sería relevante en la diferenciación de las muestras de CDG-II del resto, debido a que su proyección en el eje de las Y es elevada. Sin embargo, al ser un método exploratorio, PCA no es capaz de indicar con

exactitud que glicoformas son las que más contribuyen en la diferenciación de los tres grupos. Por este motivo, se optó por realizar un análisis multivariante supervisado del conjunto de datos y así poder elaborar un modelo más refinado que permitiese una mejora en la clasificación de los diferentes grupos. En concreto, se optó por realizar un **análisis discriminante por mínimos cuadrados parciales (PLS-DA)** [294].

Para elaborar el modelo de PLS-DA, se tuvo en cuenta la clasificación obtenida por PCA de las diferentes muestras en los tres grupos mencionados. En la Figura 5.7-b se presentan los gráficos de *scores* (i) y de *loadings* (ii) obtenidos con el modelo de PLS-DA. Tal y como se observa en el gráfico de *scores*, dos variables latentes (LV) permiten discriminar entre los tres grupos diferentes. Concretamente, la LV 1 y la LV 2 permiten explicar aproximadamente un 40% y un 30% de la varianza, respectivamente. Si se compara el gráfico de *loadings* obtenido para el PLS-DA con el de PCA, se puede ver como son prácticamente idénticos, con ligeros cambios en la posición de algunas variables. Adicionalmente, se puede apreciar cómo, en general, las glicoformas con un contenido alto en ácido siálico se sitúan en el lado positivo del eje de las X (H6N5S3, H5N4S2 para ambos glicopéptidos), zona en la que se encuentran las muestras control si observamos el gráfico de *scores*. Precisamente, esto indica que estas glicoformas tienen cierta relevancia en la diferenciación entre muestras control y CDG, y, en efecto, las glicoformas más sialiladas se encuentran en mayor concentración en las muestras control. Contrariamente, en líneas generales, son las glicoformas desialiladas o, incluso, sin glicosilación (péptido), las que se encontraban en mayor concentración en las muestras de CDG, y por ello se sitúan o bien en el lado negativo del eje de las X (zona que corresponde a las muestras de CDG-I si observamos el gráfico de *scores*, Figura 5.7-b (i)), o en lado positivo del eje de las Y (zona que corresponde a las muestras de CDG-II). Si exploramos los resultados en más detalle, se puede observar como son las glicoformas H5N4S1 y H5N4, para el glicopéptido N_{413}, y las glicoformas H5N4S1 y H6N5S2, para el glicopéptido N_{611}, las que están relacionadas

con las muestras de CDG-II. Por otro lado, las glicoformas péptido de ambos glicopéptidos parecen ser importantes en el caso de las muestras de CDG-I.

No obstante, y, aunque la información obtenida de ambos gráficos es muy útil para un primer análisis del conjunto de datos, sólo es posible obtener información cualitativa. Gracias al gráfico de **VIP** (Importancia de la Variable en la Proyección – *Variable Importance in Projection*) *scores*, es posible cuantificar la influencia de cada glicoforma en la diferenciación de los grupos. Este gráfico incluye una estimación de la importancia de cada variable en la proyección del modelo de PLS-DA construido, es decir, es un indicador de las variables originales que más contribuyen a la diferenciación entre grupos. En general, aquellas variables con un VIP≥1 suelen ser consideradas como relevantes [218]. Sin embargo, debido a que varias variables tienen un VIP≥1, se estimó oportuno aumentar el umbral (*threshold*) y considerar como variables más importantes sólo aquellas con un VIP≥1.5. En los gráficos de barras de la Figura 5.8, se muestran los VIP *scores* de cada glicoforma cuando se pretende diferenciar controles de las muestras de CDG (5.8-a), las muestras de CDG-I de los controles y CDG-II (5.8-b) y las muestras de CDG-II de los controles y CDG-I (5.8-c). Si observamos los tres gráficos en conjunto, se aprecia como únicamente existen dos glicoformas de las 14 identificadas que presentan un valor de VIP inferior a uno en todos los casos (marcadas en rojo).



Glicoformas de los glicopéptidos de la hTf con VIP≥1.0 en el modelo de PLS-DA

CONTROL	CDG-I	CDG-II
N_{413}-Pep	N_{413}-Pep	
		N_{413}-H5N4
N_{413}-H5N4S1	**N_{413}-H5N4S1**	**N_{413}-H5N4S1**
N_{413}-H5N4S2		
		N_{413}-H5N4F1S2
N_{413}-H6N5S3	N_{413}-H6N5S3	
N_{611}-Pep	N_{611}-Pep	
N_{611}-H5N4		
	N_{611}-H5N4S1	**N_{611}-H5N4S1**
N_{611}-H5N4S2	**N_{611}-H5N4S2**	
		N_{611}-H6N5S2
N_{611}-H6N5S3	**N_{611}-H6N5S3**	

* En negrita se resaltan las glicoformas con VIP≥1.5

Figura 5.8: Gráfico de VIP *scores* para las diferentes glicoformas de ambos glicopéptidos de la hTf cuando se considera la separación entre: a) controles y CDG; b) CDG-I y controles/CDG-II; c) CDG-II y controles/CDG-I. d) Tabla resumen de las diferentes glicoformas con VIP≥1 (en negrita se indican las glicoformas con VIP≥1.5).

Estas glicoformas son, precisamente, las que ya con el test ANOVA, resultaban irrelevantes a la hora de diferenciar entre los tres grupos de muestras (H6N5S2 del glicopéptido N_{413} y H5N4F1S2 del glicopéptido N_{611}). Del resto de glicoformas, se deben destacar aquellas que tienen un valor de VIP superior a 1.5 en alguno de los tres gráficos, ya que son éstas las más importantes para diferenciar entre los tres grupos de muestras. Tal y como se muestra en la tabla resumen de la Figura 5.8-d, donde aparecen todas las glicoformas con VIP superior o igual a 1 para cada uno de los gráficos, son las glicoformas H6N5S3, H5N4S1 y H5N4 del glicopéptido N_{413} y H6N5S3, H5N4S2 y H5N4S1 del glicopéptido N_{611}, las seis que tienen un VIP mayor a 1.5 (marcadas en negrita). Entre estas seis, las glicoformas N_{413}-H6N5S3, N_{611}-H6N5S3 y N_{611}-H5N4S2, cuya área normalizada disminuye en las muestras patológicas, son las más importantes para diferenciar controles de las muestras de CDG. De manera similar, las glicoformas N_{413}-H5N4S1, N_{611}-H6N5S3, N_{611}-H5N4S2 y N_{611}-H5N4S1 son las más influyentes para discriminar las muestras de CDG-I del resto. Las glicoformas péptido resultan ser importantes (VIP>1) pero no las más destacadas (VIP<1.5), seguramente debido a la elevada heterogeneidad de las muestras de CDG-I (diferentes individuos, mutaciones, etc.). Finalmente, las glicoformas desialiladas y truncadas N_{413}-H5N4S1, N_{413}-H5N4 y N_{611}-H5N4S1 son las más discriminantes entre las muestras de CDG-II y el resto.

Finalmente, aunque a la hora de construir el modelo de PLS-DA se realizó una validación cruzada *leave-one-out*, con el fin de realizar una validación externa del modelo, se construyó un nuevo set de nueve muestras adicionales (3 controles (i-iii), 3 CDG-I (iv-vi) y 3 CDG-II (vii-ix)) que se analizaron y procesaron del mismo modo que las muestras incluidas en el modelo de PLS-DA. Tal y como se puede observar en la Figura 5.7-b (i) (muestras i-ix, etiquetadas como muestras de validación), el modelo de PLS-DA es capaz de clasificar, con errores muy reducidos, cada muestra dentro del grupo correspondiente.

Gracias a este trabajo, fue posible establecer un método de análisis multivariante sencillo pero que permite obtener información relevante de un set de datos complejos, como el obtenido tras el análisis de las glicoformas de los glicopéptidos de la hTf por CapLC-MS. Además se puede observar como este tipo de métodos quimiométricos permiten profundizar en la relevancia de las diferentes glicoformas para distinguir entre controles, CDG-I y CDG-II. No obstante, en un futuro se deberían analizar un número superior de muestras para acabar de confirmar si estas seis glicoformas con VIP ≥ 1.5 podrían considerarse biomarcadores potenciales para el diagnóstico de CDG.

Comparación entre el estudio del alcoholismo y CDG-I

Si se comparan los resultados obtenidos en ambos estudios, cabe destacar la similitud existente entre la alteración de la glicosilación de la hTf producida en individuos con CDG-I y en los consumidores crónicos de alcohol, motivo por el cual, las glicoformas CDT de la hTf se pueden considerar biomarcadores para ambas enfermedades. Sin embargo, en el caso de neonatos, en la bibliografía se detalla que un excesivo consumo de alcohol durante el embarazo puede dar lugar a alteraciones del feto y en los años posteriores al nacimiento del niño, muy parecidas a las producidas por una CDG-I [32]. También se han publicado algunos casos de falsos positivos en el diagnóstico de CDG-I o alcoholismo debido a estas similitudes [32, 295].

Para hacer más sencilla la comparación entre ambos estudios, en la Tabla 5.5 se incluyen las áreas normalizadas de las diferentes glicoformas de los glicopéptidos de la hTf para una muestra de suero control, y de suero de alcohólico leve y grave. También se incluyen las áreas normalizadas de un suero control y de dos muestras de suero de CDG-I, en concreto, la DPAGT1-CDG y la RFT1-CDG. Para realizar una comparación más rigurosa las áreas normalizadas del estudio de CDG-I se recalcularon con el péptido de referencia utilizado en el estudio del alcoholismo (*DGAGDVAFVK*).

Tabla 5.5: Área normalizada de diferentes glicoformas detectadas para el glicopéptido N_{413} y N_{611} de la hTf en muestras de suero control y patológicas siguiendo la metodología utilizada en el estudio del alcoholismo (azul) y el estudio de los CDG (naranja).

Glicopéptido	Glicoforma	Estudio alcoholismo			Estudio CDG-I		
		Control	Alcohólico leve	Alcohólico grave	Control	DPAGT1 (CDG-I_j)	RFT1 (CDG-I_n)
N_{413}	Péptido	-	-	3.0	0.4	18.9	56.8
	H5N4	2.5	1.4	1.6	0.7	0.6	-
	H5N4S1	15.7	12.3	11.4	6.2	6.8	3.1
	H5N4S2	66.3	60.9	55.0	74.7	47.4	42.4
	H5N4F1S2	1.3	1.1	1.4	1.5	0.8	3.1
	H6N5S3	1.1	1.0	1.0	1.9	1.3	1.6
N_{611}	Péptido	0.8	2.9	8.5	4.9	17.0	44.6
	H5N4	0.5	0.5	0.4	0.8	0.4	0.4
	H5N4S1	13.8	11.1	12.0	5.7	5.0	4.5
	H5N4S2	74.1	71.9	67.0	89.0	49.4	42.3
	H5N4F1S2	4.2	3.6	4.4	4.7	3.3	4.4
	H6N5S3	1.9	1.0	1.5	3.1	1.7	1.3

Áreas normalizadas calculadas como: (Área glicoforma / Área péptido DGAGDVAFVK) x 100.

Si comparamos todas las muestras, se puede observar como ambas enfermedades presentan un aumento de las glicoformas péptido y una disminución de las glicoformas H5N4S2 en ambos glicopéptidos, aunque en proporciones distintas. Para poder apreciar mejor estas diferencias, en la Figura 5.9 se incluye un gráfico de barras con el aumento de área normalizada respecto al correspondiente control de las diferentes muestras de alcohólicos y CDG-I. Se puede observar cómo, en líneas generales, el aumento de la glicoforma péptido para ambos glicopéptidos de la hTf es bastante inferior en alcohólicos, incluso en aquellos ya considerados severos (ejemplo: aproximadamente 9% de área normalizada para el péptido N_{611} para un alcohólico severo, en comparación a un valor del 17% para la muestra de DPAGT1).

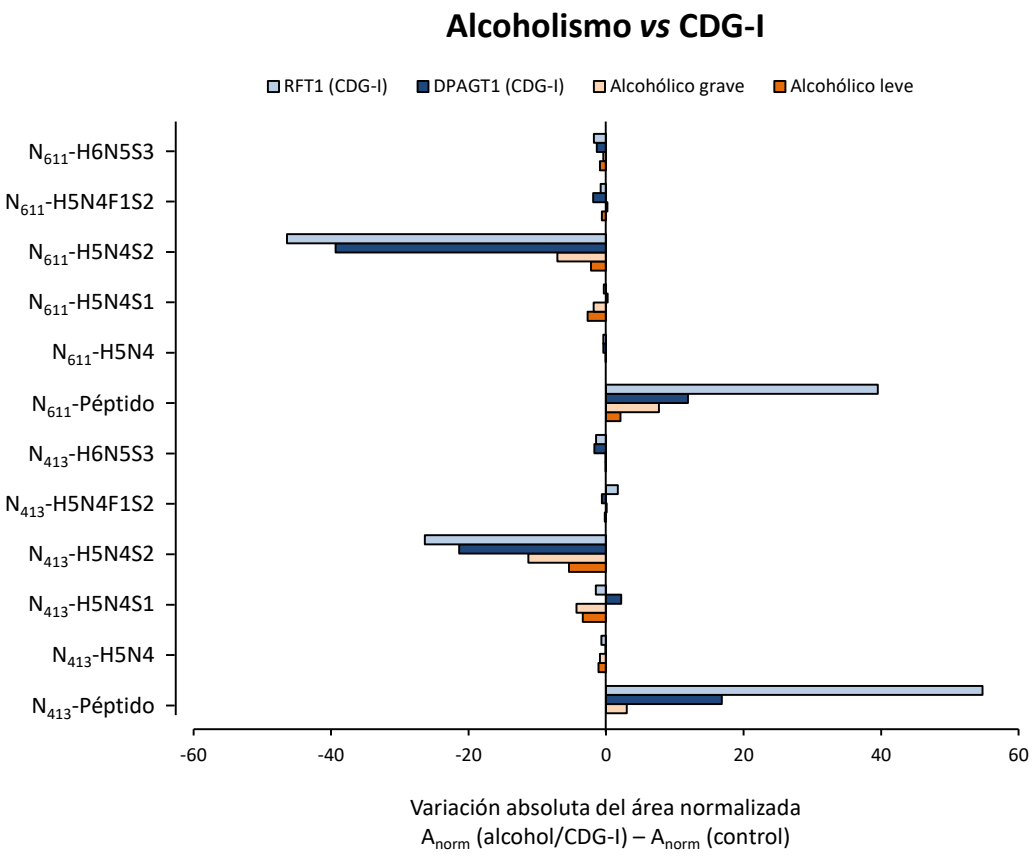


Figura 5.9: Aumento o disminución, con respecto al suero control, del área normalizada de las diferentes glicoformas de ambos N-glicopéptidos de la hTf en diferentes muestras de suero tanto para el estudio del alcoholismo como el de los CDG-I.

También existe una diferencia considerable entre alcohólicos y CDG-I en cuanto a la disminución de la glicoforma H5N4S2 de ambos glicopéptidos, siendo mucho más acusada en individuos que sufren una CDG-I que en aquellos con alcoholismo. En cuanto al resto de glicoformas, es más complicado extraer una conclusión, debido, principalmente, a que no hay una disminución o aumento absoluto de las áreas normalizadas tan notable. Sin embargo, la conclusión más importante que se puede extraer de estos resultados es que parece que los CDG-I afectan más a la glicosilación de la hTf, produciéndose una mayor alteración que en el caso del alcoholismo crónico.

No obstante, aunque se pueden observar diferencias interesantes entre ambas enfermedades, sería necesario hacer un estudio más exhaustivo con un mayor número de muestras y realizar

tratamientos quimiométricos, como un PLS-DA, para poder verificar las conclusiones extraídas. Una vez realizado este estudio, la metodología desarrollada podría considerarse como una alternativa adecuada para discriminar entre alcoholismo y CDG-I en aquellos casos donde el diagnóstico no sea claro con los métodos de referencia actuales.

5.1.3. Transferrina de ratón y su implicación en la artritis inducida por colágeno (CIA)

Como se comentó en la introducción de esta tesis doctoral, la hTf también está implicada en procesos de inflamación al ser una proteína de fase aguda negativa, es decir, su concentración experimenta una disminución de más del 25% como respuesta inicial a un proceso inflamatorio. Estos procesos pueden ser causados por varios motivos, siendo uno de los más comunes las enfermedades autoinmunes, que no son más que procesos inflamatorios debidos a la acción del sistema inmune del organismo en contra de una célula u órgano propio del individuo, que es reconocido como agente extraño. Además de la alteración en la concentración de ciertas proteínas, también se ha descrito que algunas glicoproteínas pueden ver modificada su glicosilación en un proceso inflamatorio [38, 40, 70]. Por ejemplo, la glicosilación de la hTf se ve alterada en la artritis reumatoide (RA), una enfermedad autoinmune que provoca un aumento de las glicoformas más sialiladas y una disminución de aquellas menos sialiladas [41].

El estudio de enfermedades de origen humano en modelos de ratones es una opción muy interesante en glicoproteómica, como estudio preliminar a la investigación en humanos. En concreto, la artritis inducida por colágeno (CIA) en ratones comparte muchas características con la RA en humanos, tanto en la evolución de la enfermedad, en cómo afecta a los tejidos de las articulaciones o en su respuesta a diferentes fármacos [75, 76]. Esta similitud hace prever que la glicosilación de la transferrina de ratón (mTf) también podría verse alterada en presencia de CIA. Por todos estos motivos, en el siguiente estudio (artículo 2.3) se aplicó la

metodología desarrollada de purificación, digestión y análisis de los glicopéptidos por CapLC-MS para estudiar la alteración de la mTf en CIA.

El trabajo llevado a cabo en este tercer artículo fue posible gracias a la colaboración con el Instituto de Parasitología y Biomedicina "López-Neyra" (IPBLN) del Centro Superior de Investigaciones Científicas (CSIC) de Granada, que se encargaron de realizar los experimentos de electroforesis en gel y de obtener las muestras de bazo y suero de los ratones.

El estudio de la glicosilación de la mTf surgió a partir de un estudio previo más general donde se pretendía observar diferencias de abundancia en las proteínas del bazo de ratones con y sin CIA. Este estudio se realizó por electroforesis diferencial en gel (DiGE) y se compararon las proteínas detectadas en cuatro tipos de muestras de bazo de diferentes ratones:

- Wild-type (WT) control (ratones sanos sin CIA).
- Wild-type (WT) con CIA[a].
- CD38-KO[b] control.
- CD38-KO[b] con CIA[a].

Se aplicaron diferentes métodos estadísticos y quimiométricos (ANOVA de dos factores y PCA) con el fin de detectar varios *spots* cuya diferencia de concentración fuese significativa entre las diferentes muestras de ratón con CIA respecto a los controles, tanto en ratones WT como CD38-KO. Muchos de estos *spots* se identificaron, mediante su digestión con tripsina y detección de los péptidos resultantes por MALDI-TOF/TOF-MS, como *spots* pertenecientes a diferentes especies de la mTf y por este motivo se decidió analizar la glicosilación de esta proteína en más profundidad a partir del análisis de los glicopéptidos con la metodología desarrollada para el estudio del alcoholismo y los CDG.

[a]Los ratones utilizados (tipo C57BL/6 o B6) desarrollan CIA al ser inmunizados con colágeno tipo II de pollo (Col-II) juntamente con un inmunopotenciador.
[b]Ratones modificados genéticamente que no presentan el gen CD38, descritos como resistentes a desarrollar CIA [296].

Análisis de un patrón de mTf por CapLC-MS

Previamente al análisis de las muestras de bazo, se analizó un patrón de 1500 µg mL^{-1} de mTf comercial, además de un suero control (WT) previamente purificado por IAC. Mediante este análisis se pudo: 1) confirmar que el método de digestión desarrollado también era adecuado para la mTf; 2) verificar que las condiciones cromatográficas y del espectrómetro de masas eran también adecuadas para el análisis del glicopéptido de la mTf; 3) detectar y caracterizar las diferentes glicoformas del glicopéptido de la mTf presentes en un patrón de mTf; 4) comprobar que la columna de inmunoafinidad anti-hTf también retenía y purificaba la mTf de las muestras de suero. Tal y como se puede observar con los resultados obtenidos (Figura 5.10, Tabla 5.6), no fue necesario modificar ninguno de los procesos mencionados anteriormente, probablemente debido a la gran similitud entre la hTf y la mTf (glicanos de tipo complejo, 72% de semejanza entre sus secuencias aminoacídicas, glicosilación reducida, estructura compacta con varios puentes disulfuro). A modo de resumen, la Figura 5.10 muestra los EICs de las glicoformas más abundantes del glicopéptido en el patrón de mTf, así como el espectro de masas para dos de ellas, la H5N4S2 y la H5N4F1S3.

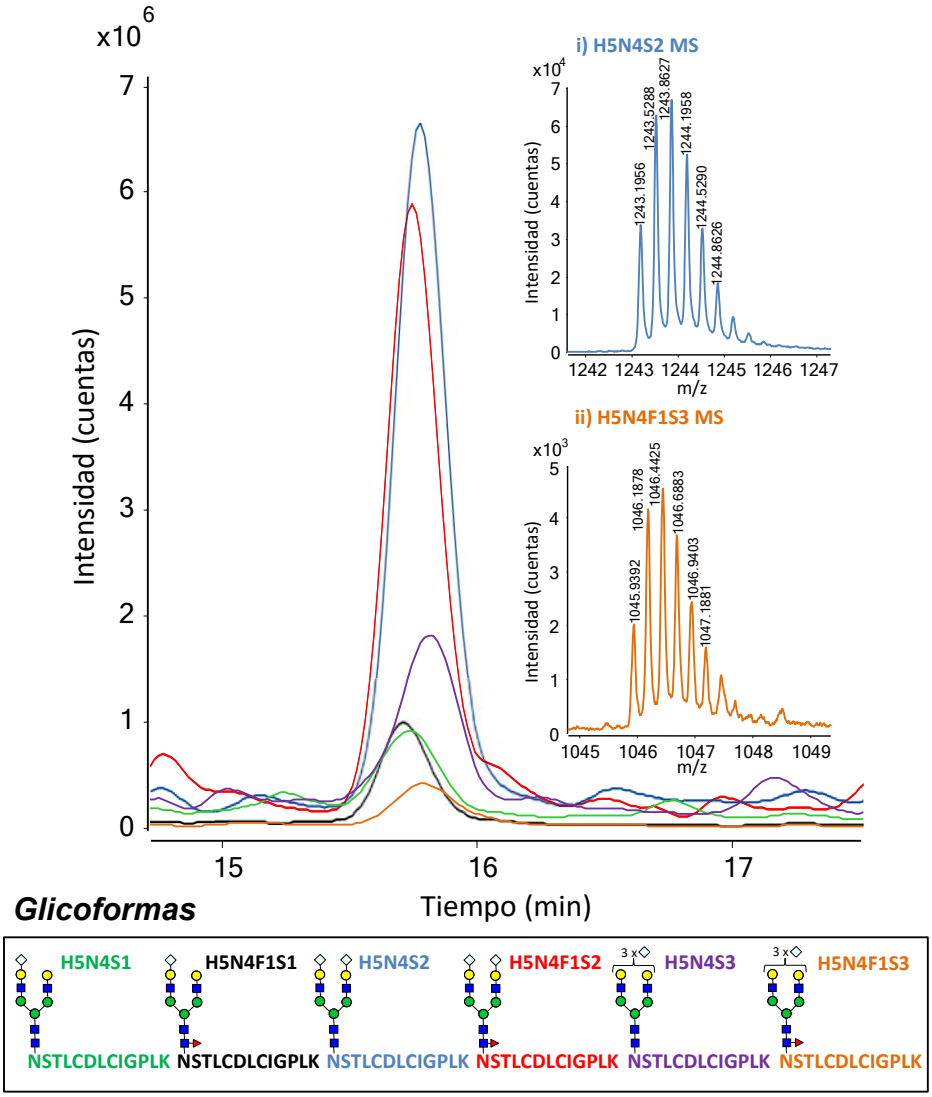
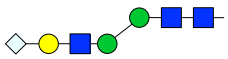
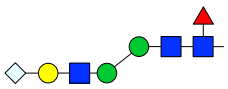
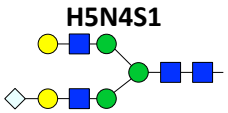
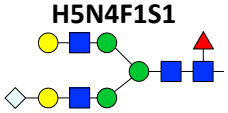
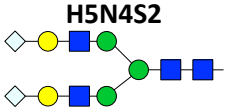
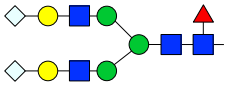
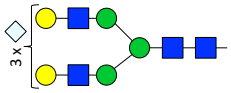
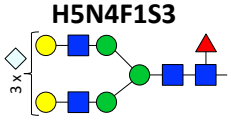
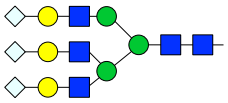


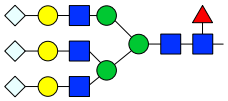
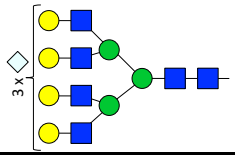
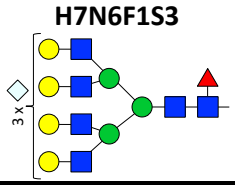
Figura 5.10: EICs de las glicoformas más abundantes en un patrón de mTf y espectro de masas de las glicoformas i) H5N4S2 y ii) H5N4F1S3. El método de digestión y análisis utilizado para el estudio del patrón de mTf es el mismo que el utilizado para la hTf en los artículos anteriores (artículo 2.1 y 2.2). Los símbolos utilizados para la representación de los glicanos siguen las reglas sugeridas por el *Consortium for Functional Glycomics* (CFG) [13].

Además también se incluye, en la Tabla 5.6, el área normalizada, la masa teórica y experimental, el error de masa y las cargas observadas de todas las glicoformas detectadas, siendo las glicoformas más abundantes la H5N4S2 y H5N4F1S2. Cabe destacar que la cobertura de la secuencia aminoacídica, a partir del análisis de los péptidos y del glicopéptido del digesto de la mTf patrón por CapLC-MS, fue de casi el 91%. También se debe tener cuenta, que a diferencia de la hTf, las glicoformas detectadas de la mTf contenían N-glicolilneuramínico

como ácido siálico, y no se detectó ninguna glicoforma con N-acetilneuramínico. Además, también se puede observar como la fucosilación de esta glicoproteína es muy acusada, detectándose siempre un homólogo fucosilado de cada glicoforma, aunque en general es menos abundante que el no-fucosilado. Otra diferencia destacable respecto a la hTf es el hecho de que se detectaron glicoformas con un mayor grado de sialilación, como la glicoforma H5N4S3, o glicoformas más ramificadas, como, por ejemplo, la H7N6S3.

Tabla 5.6: Masa teórica (M_{teo}) y experimental (M_{exp}), error de masa, cargas observadas y área normalizada (A_{norm}) para las glicoformas detectadas del glicopéptido N_{494} en un patrón de mTf.

Glicoforma	A_{norm}	RSD (%)	M_{teo}	M_{exp}	Error (ppm)	Cargas observadas
H3N3S1 	**5.0**	7.4	2892.2138	2892.2443	10.5	+2,+3
H3N3F1S1 	**1.6**	7.7	3038.2717	3038.3080	11.9	+2
H5N4S1 	**12.5**	5.8	3419.3988	3419.4394	11.9	+2,+3,+4,+5
H5N4F1S1 	**6.6**	6.3	3565.4568	3565.4955	10.9	+2,+3,+4
H5N4S2 	**88.2**	2.6	3726.4892	3726.5352	12.3	+2,+3,+4,+5
H5N4F1S2 	**77.7**	2.6	3872.5471	3872.5958	12.6	+2,+3,+4,+5
H5N4S3 	**28.7**	6.6	4033.5795	4033.6386	14.6	+2,+3,+4
H5N4F1S3 	**30.8**	5.6	4179.6374	4179.6860	11.6	+2,+3,+4
H6N5S3 	**6.9**	4.2	4398.7117	4398.7553	9.9	+3,+4

H6N5F1S3 	**5.9**	2.0	4544.7696	4544.8172	10.5	+3,+4
H7N6S3 	**4.3**	9.2	4763.8439	4763.7646	16.7	+3
H7N6F1S3 	**2.6**	6.8	4909.9018	4909.9535	10.5	+3

Las áreas normalizadas se calcularon como: (Área glicoforma / Área péptido CDEWSIISEGK) x 100.

Estudio de la alteración de la glicosilación de la mTf debido a CIA por CapLC-MS

El estudio de la alteración de la glicosilación de la mTf en ratones que desarrollan CIA, se llevó a cabo de dos maneras diferentes. En primer lugar se analizó la mTf de las muestras de bazo analizadas anteriormente por DiGE, a partir de la digestión en gel de los spots correspondientes a esta glicoproteína. Por otro lado, puesto que en muchas enfermedades autoinmunes se producen alteraciones de la concentración de proteínas en el suero y puesto que la mTf es una proteína abundante en el suero, se aplicó la metodología establecida anteriormente a muestras de suero de ratones que desarrollaban CIA, para comprobar si la mTf también se encontraba alterada en este fluido biológico.

El estudio en muestras de bazo no resultó ser demasiado satisfactorio dado que la concentración de mTf en bazo es reducida y, por tanto, no se disponía de sensibilidad suficiente para detectar correctamente todas las glicoformas del glicopéptido en los diferentes *spots*. En cambio, el análisis de las muestras de suero sí que permitió detectar con éxito las diferentes glicoformas del glicopéptido de la mTf. En la Tabla 5.7, se recogen las glicoformas detectadas en las cuatro muestras de suero de ratón: WT control, WT con CIA, CD38-KO control y CD38-KO con CIA.

Tabla 5.7: Área normalizada (A_{norm}) para las glicoformas detectadas del glicopéptido N_{494} de la mTf en muestras de suero de ratones control y ratones con CIA, tanto en ratones WT como en ratones alterados genéticamente (CD38-KO).

Glicoforma	WT control		WT CIA		CD38-KO control		CD38-KO CIA	
	A_{norm}	RSD (%)	A_{norm}	RSD (%)	A_{norm}	RSD (%)	A_{norm}	RSD (%)
H3N3S1	**7.2**	9.7	**3.0**	2.4	**4.9**	5.0	**3.4**	2.8
H3N3F1S1	**1.3**	9.9	**1.6**	4.5	**1.9**	4.3	**2.4**	5.6
H5N4S1	**20.4**	0.9	**16.1**	5.9	**20.7**	1.6	**19.9**	4.6
H5N4F1S1	**6.0**	4.3	**4.6**	1.2	**6.9**	4.5	**8.4**	4.8
H5N4S2	**142.6**	1.5	**135.4**	2.6	**153.6**	2.1	**147.6**	0.2
H5N4F1S2	**71.5**	3.1	**60.8**	3.2	**58.9**	1.2	**64.5**	1.0
H5N4S3	**46.1**	1.2	**38.3**	3.5	**42.9**	2.2	**35.1**	4.6
H5N4F1S3	**3.9**	5.0	**4.1**	6.2	**4.0**	7.4	**4.5**	10.0
H6N5S3	**2.2**	10.1	**4.1**	5.0	**2.7**	2.5	**4.6**	7.1
H6N5F1S3	**0.5**	9.0	**0.7**	4.7	**0.6**	12.2	**0.6**	12.9
H7N6S3	**4.3**	4.6	**3.9**	12.2	**5.6**	1.6	**5.2**	6.0
H7N6F1S3	**2.4**	7.6	**3.0**	9.7	**2.4**	5.7	**3.4**	5.9

Las áreas normalizadas se calcularon como: (Área glicoforma / Área péptido CDEWSIISEGK) x 100.

Si se comparan las Tablas 5.6 y 5.7, se puede observar que se detectaron las mismas glicoformas en el patrón de mTf y en las muestras de suero, siendo también las más abundantes las glicoformas H5N4S2 y H5N4F1S2. Sin embargo, existe cierta diferencia entre el área normalizada de algunas glicoformas del glicopéptido en el suero respecto al patrón, como, por ejemplo, la glicoforma H5N4S2 o la H5N4F1S3. Aunque no afecta al objetivo de este trabajo, esta diferencia de área puede ser debida al tipo de ratón utilizado en el patrón comercial respecto a los ratones utilizados en este estudio (ratones C57BL/6, machos de 8-12 semanas de edad). Es posible que, al igual que la glicosilación en humanos puede variar en función de la raza, edad o sexo del individuo, ratones de diferentes tipos tengan una glicosilación distinta.

Para visualizar mejor los resultados obtenidos y observar con más claridad los cambios en la glicosilación de la mTf, en la Figura 5.11, se representa en un gráfico de barras el $\log \frac{A_{norm}\ CIA}{A_{norm}\ control}$ tanto para las muestras de WT como de CD38-KO.

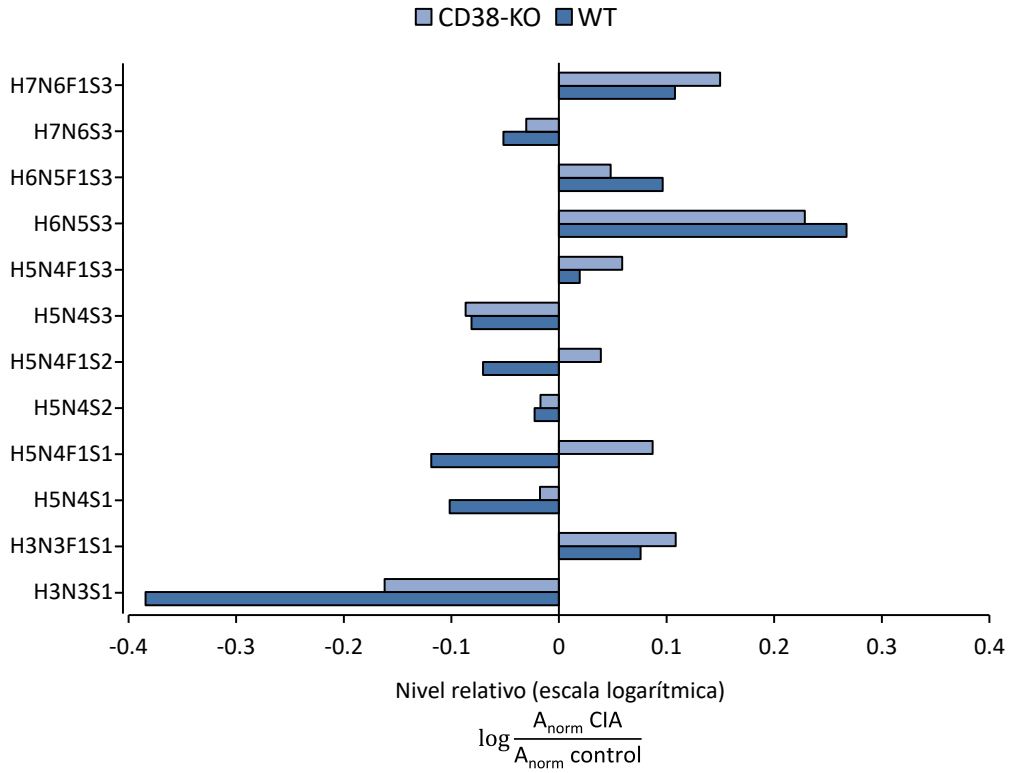


Figura 5.11: Gráfico de barras del logaritmo en base 10 del cociente de las áreas normalizadas de las diferentes glicoformas del glicopéptido de la mTf entre una muestra de suero de ratón con CIA y una muestra de suero control, tanto para los ratones WT como para los ratones CD38-KO.

Como se puede deducir de esta figura, la glicosilación de la mTf se ve claramente alterada en ratones con CIA, tanto si son WT como CD38-KO. Se puede observar también como el perfil de glicosilación también es diferente entre los dos tipos de ratones, WT o CD38-KO. Además, a excepción de la glicoforma H7N6S3, el resto de glicoformas más ramificadas, las glicoformas tri- y tetraantenarias (H7N6F1S3, H6N5S3 y H6N5F1S3), se ven aumentadas en muestras con CIA, tanto en ratones WT como en ratones CD38-KO. También se puede apreciar que, en líneas generales, las glicoformas menos ramificadas disminuyen, en especial la glicoforma H3N3S1, y que las glicoformas fucosiladas también parecen aumentar en ratones con CIA.

Aunque con el estudio llevado a cabo en este trabajo es razonable afirmar que la metodología establecida también puede ser útil para el diagnóstico y estudio de enfermedades inflamatorias en las que exista una alteración en la glicosilación de la transferrina, se debería realizar un estudio más amplio, con un mayor número de muestras, para poder extraer conclusiones fiables y evaluar el papel de la mTf en CIA y el efecto del gen CD38 en el desarrollo de dicha enfermedad.

5.2. Estudio de la glicosilación de glicoproteínas por electroforesis capilar acoplada a la espectrometría de masas

5.2.1. Desarrollo y optimización de la metodología analítica

Con el objetivo de desarrollar una metodología analítica complementaria a CapLC-MS capaz de separar las diferentes glicoformas de los glicopéptidos de la hTf, en esta tesis doctoral se decidió evaluar la electroforesis capilar acoplada a la espectrometría de masas (CE-MS). Para desarrollar dicha metodología, se partió de un método establecido previamente en nuestro grupo de investigación para el análisis de los glicopéptidos de otras glicoproteínas, como por ejemplo, la eritropoyetina humana recombinante (rhEPO) [20, 82], empleando un electrolito de separación de 50 mM HAc / 50 mM HFor (pH 2.2-2.3) y un líquido coaxial auxiliar (*sheath liquid*) de iPrOH:H_2O (50:50, v/v) con 0.05% (v/v) de HFor. No obstante, en el caso de la hTf los picos obtenidos fueron muy anchos y distorsionados tanto para péptidos como glicopéptidos, además de obtener una intensidad de la señal y una reproducibilidad entre inyecciones muy baja. Para comprobar si este efecto podía ser causado por el uso de RapiGest®, se decidió digerir, con y sin RapiGest®, y analizar por CE-MS otra glicoproteína bastante caracterizada por el grupo de investigación, como es la **alfa-1-glicoproteína ácida bovina (bAGP).** Tal y como se puede observar en la Figura 5.12-a, que muestra a modo de ejemplo los electroferogramas de iones extraídos (EIEs) de las glicoformas H5N4S2 y H5N4S3 del glicopéptido N_{118} de la bAGP, se detectaron las diferentes glicoformas sin ninguna distorsión en los picos electroforéticos cuando no se usó RapiGest® en el proceso de digestión. Por el contrario, cuando se utilizó RapiGest®, tal y como ocurría con la hTf, se obtuvieron picos distorsionados, confirmando que el fragmento residual del surfactante que queda en solución tras la hidrólisis ácida, parecía interaccionar, de alguna manera, con las paredes internas del capilar de CE, dando lugar a estos picos distorsionados, no solo para los glicopéptidos pero también para los péptidos (ver Figura 5.12-b). Cabe destacar además, que este efecto no es exclusivo del RapiGest®, puesto

que se obtuvieron resultados muy similares en referencia a la baja reproducibilidad y a la distorsión de los picos, al emplear otros surfactantes aniónicos como el ProteaseMAX™ (Promega).

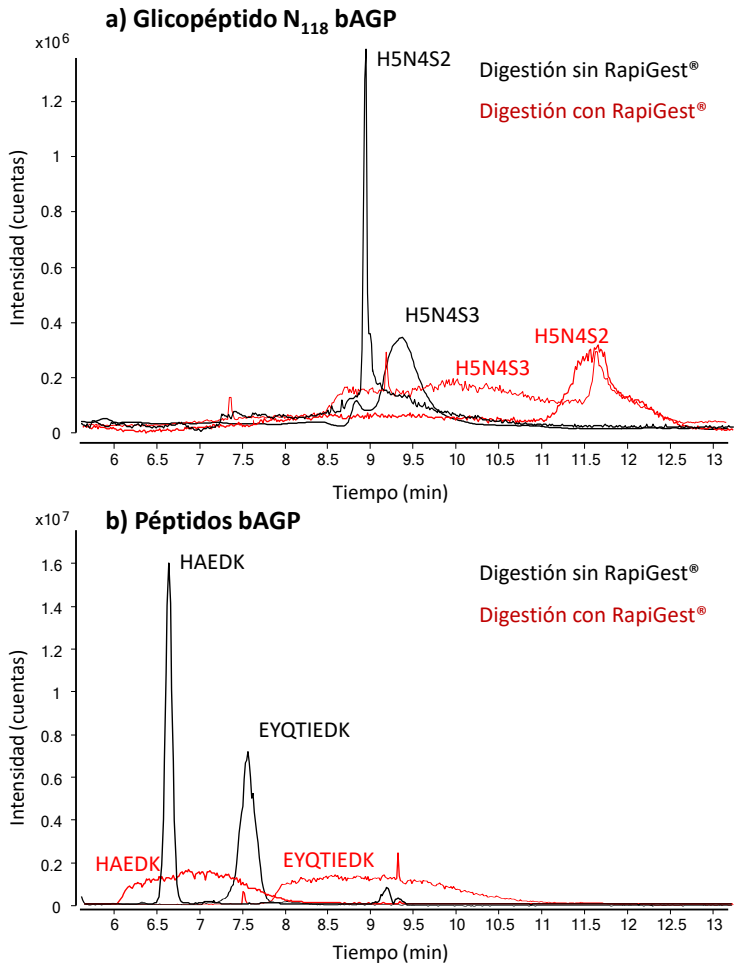


Figura 5.12: Electroferogramas de iones extraídos (EIEs) de a) las glicoformas H5N4S2 y H5N4S3 del glicopéptido N_{118} y b) dos péptidos diferentes de un patrón de bAGP digerido sin (negro) y con (rojo) RapiGest®.

Como el uso del RapiGest® es imprescindible en el análisis de los péptidos y glicopéptidos de la hTf, en esta tesis doctoral (artículo 3.1) se puso a punto un método de *clean-up* de la muestra para eliminar el residuo de RapiGest® que interfiere en el análisis por CE-MS.

Para lograr este objetivo, se introdujo un proceso de desalado posterior a la digestión enzimática, que consistía en el uso de cartuchos de extracción en fase sólida en una microplaca

de elución, los cuales contenían una fase estacionaria polimérica Oasis® HLB (WATERS® S.A). Tal y como se puede observar en la Figura 5.13, la eliminación del RapiGest® (residuo con grupo sulfónico, ver Figura 5.1) provocó una mejora sustancial de la señal y de la forma de los picos electroforéticos de la hTf, tanto para glicopéptidos como para péptidos.

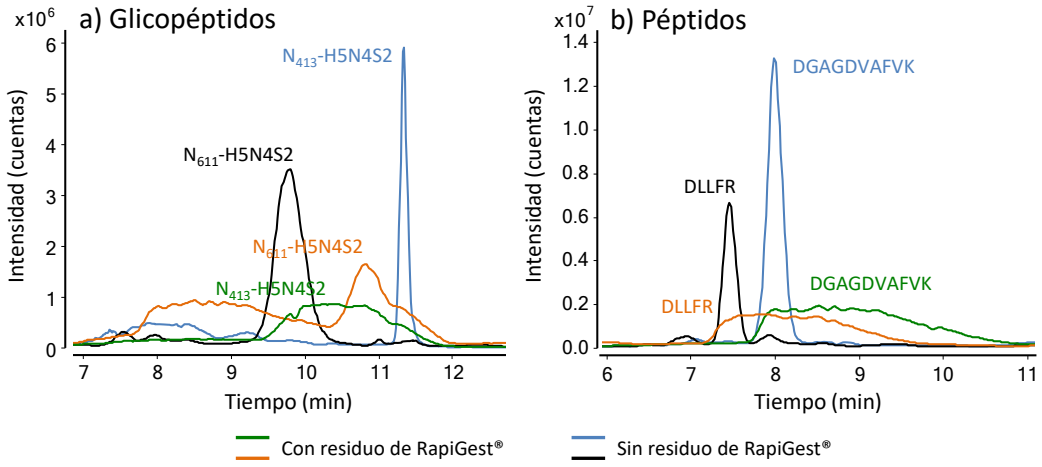


Figura 5.13: EIEs de a) la glicoforma mayoritaria H5N4S2 de ambos glicopéptidos de la hTf y b) de dos péptidos diferentes del digesto, en muestras con residuo de RapiGest® y en muestras purificadas con las microplacas de elución Oasis® HLB.

Este tratamiento permitió compatibilizar el uso de surfactantes aniónicos como el RapiGest® con el análisis de digestos de glicoproteínas por CE-MS. A modo de ejemplo, en la Figura 5.14 se incluyen los EIEs de varias glicoformas detectadas de los glicopéptidos N_{413} y N_{611}, del O-glicopéptido y de algunos péptidos del digesto tríptico de un patrón de hTf aplicando la metodología establecida. Como se puede observar la forma de los picos es adecuada y, a diferencia de lo que ocurre en CapLC, las glicoformas de los N-glicopéptidos se separan.

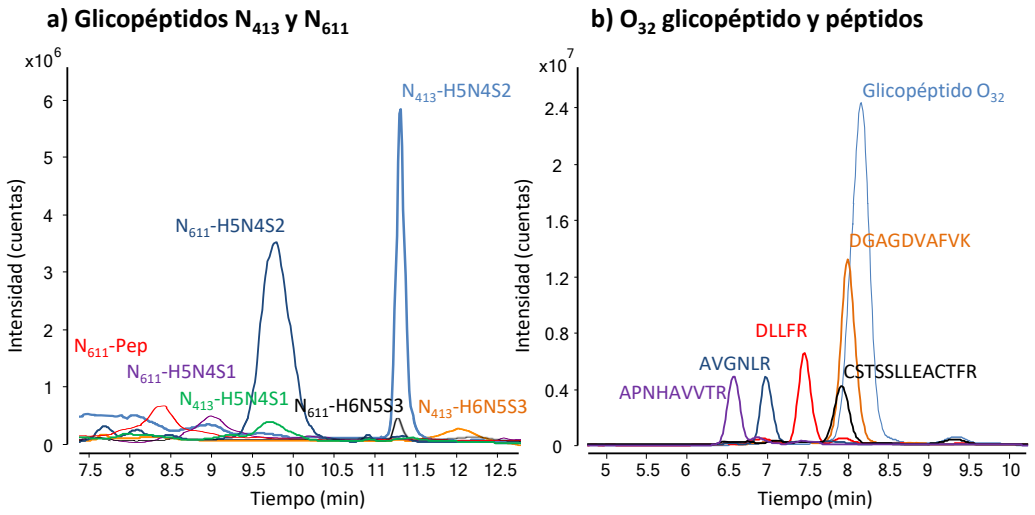


Figura 5.14: EIEs de a) gran parte de las glicoformas detectadas de ambos N-glicopéptidos y b) EIEs del O-glicopéptido y varios péptidos de un digesto de un patrón de hTf por CE-MS una vez eliminado el RapiGest® por SPE. Electrolito de separación: 50 mM HFor / 50 mM HAc pH: 2.2-2.3. Líquido coaxial auxiliar: iPrOH:H_2O (60:40, v/v) con 0.05% (v/v) HFor.

5.2.2. Análisis de muestras de suero

Finalmente, para evaluar la robustez de la metodología desarrollada al analizar muestras biológicas, se procedió al análisis de muestras de suero control y de CDG (artículo 3.1), previa purificación de la hTf mediante IAC (apartado 5.1.1).

En la Figura 5.15, se muestran tres glicoformas del glicopéptido N_{611} en una muestra de suero control y en tres muestras diferentes de CDG (PMM2, MPI y COG8). En general, si comparamos los resultados de las muestras reales con los obtenidos con el patrón de hTf (Figura 5.14-a), se puede observar que la separación entre las diferentes glicoformas es muy parecida, demostrando la aplicabilidad de la metodología en muestras reales. La diferencia más destacable es la intensidad observada, siendo más reducida para las muestras de suero, debido a las pérdidas de glicoproteína después de la inmunopurificación dado que las recuperaciones con las columnas de IAC eran de alrededor un 45% (ver apartado 5.1.1, artículo

2.1). En el caso de las muestras de CDG además se observa una clara disminución de la glicoforma H5N4S2, debida a la propia enfermedad.

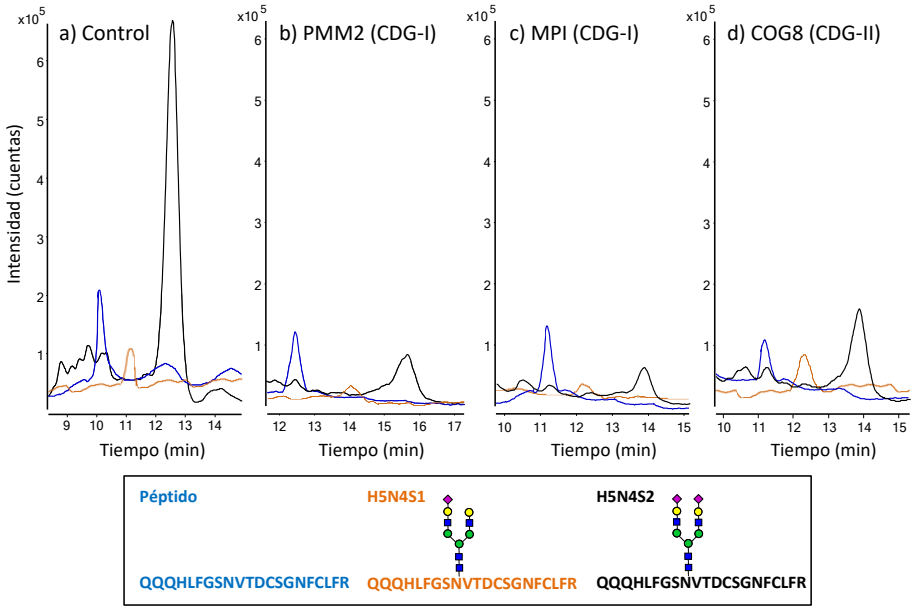


Figura 5.15: EIEs de las tres glicoformas principales del glicopéptido N_{611} de la hTf en diferentes muestras de suero: a) suero control; b) PMM2-CDG (CDG-I); c) MPI-CDG (CDG-I); d) COG8-CDG (CDG-II). Los símbolos utilizados para la representación de los glicanos siguen las reglas sugeridas por el *Consortium for Functional Glycomics* (CFG) [13].

En la Tabla 5.8, se recogen las áreas normalizadas de las diferentes glicoformas para las cuatro muestras de suero. Dado que, en general, los péptidos migran antes que los glicopéptidos por CE, la elección de un péptido de referencia fue más complicada. Los péptidos DGAGDVAFVK y DYELLCLDGTR escogidos para CapLC-MS fueron descartados en este estudio ya que no migraban a un tiempo similar ni tenían una intensidad similar a las glicoformas de los glicopéptidos. En este caso se seleccionó el péptido SAGWNIPIGLLYCDLPEPR, ya que era el más próximo y de intensidad parecida a la glicoforma mayoritaria H5N4S2 de ambos N-glicopéptidos, con el fin de poder cuantificar de manera relativa y poder comparar entre muestras.

Tabla 5.8: Área normalizada para las diferentes glicoformas detectadas para ambos N-glicopéptidos de la hTf en diferentes muestras de suero control y CDG (n=3).

Muestras									
Glicoforma		Control		PMM2-CDG (CDG-I)		MPI-CDG (CDG-I)		COG8-CDG (CDG-II)	
		A_{norm} (%)	RSD (%)	A_{norm} (%)	RSD (%)	A_{norm} (%)	RSD (%)	A_{norm} (%)	RSD (%)
N_{413}	Péptido	**0.0**	0.0	**77.3**	1.2	**76.5**	4.5	**0.0**	0.0
	H5N4S1	**3.5**	8.0	**2.5**	7.8	**1.7**	1.5	**13.4**	6.3
	H5N4S2	**60.2**	1.9	**25.8**	4.6	**25.1**	3.2	**28.4**	0.5
	H5N4F1S2	**2.2**	2.0	**3.3**	7.5	**2.5**	5.0	**2.8**	2.0
	H6N5S3	**7.1**	7.5	**4.0**	6.1	**3.5**	6.5	**5.4**	8.1
N_{611}	Péptido	**13.0**	3.7	**31.6**	6.1	**46.3**	1.1	**9.9**	5.3
	H5N4S1	**6.3**	9.0	**4.4**	6.1	**4.1**	7.4	**15.6**	10.2
	H5N4S2	**83.3**	3.1	**31.9**	9.1	**23.2**	1.5	**36.0**	7.4
	H5N4F1S2	**6.9**	2.9	**6.4**	4.9	**6.4**	2.0	**6.7**	3.4
	H6N5S3	**3.0**	1.7	**2.3**	4.3	**1.8**	6.6	**1.8**	7.9

Las áreas normalizadas se calcularon como: (Área glicoforma / Área péptido SAGWNIPIGLLYCDLPEPR) x 100.

Si se comparan las áreas normalizadas de esta tabla con las de la Tabla 5.4, se puede observar que las áreas normalizadas obtenidas por CapLC-MS y CE-MS no son directamente comparables. Esto se debe a que se utiliza un péptido de referencia distinto en cada caso y a que la separación es muy diferente. No obstante, la distribución observada de las diferentes glicoformas es la misma en CE-MS que la obtenida por CapLC-MS. Además, con los resultados obtenidos es posible observar que ambas técnicas analíticas son equiparables por lo que respecta a los resultados y conclusiones obtenidas al comparar entre muestras. Así, por CE-MS también se observó una disminución de la glicoforma mayoritaria (H5N4S2) en todas las muestras de CDG, un aumento de la glicoforma no glicosilada (péptido) en las muestras de CDG-I y un aumento de la H5N4S1 en la muestra de CDG-II.

Comparación entre CapLC-MS y CE-MS y ventajas en el estudio de la glicosilación de proteínas

Después de analizar la glicosilación de la hTf tanto en muestras patrón como en muestras de suero, a partir de las dos metodologías desarrolladas en esta tesis, CapLC-MS y CE-MS, es

interesante poder comparar de manera general las ventajas e inconvenientes de cada una y destacar su aportación al estudio de la glicosilación en el campo de la glicoproteómica.

En la Figura 5.16 se muestran los EIEs y EICs de las glicoformas principales del glicopéptido N_{611} en un patrón de hTf, utilizando ambas metodologías.

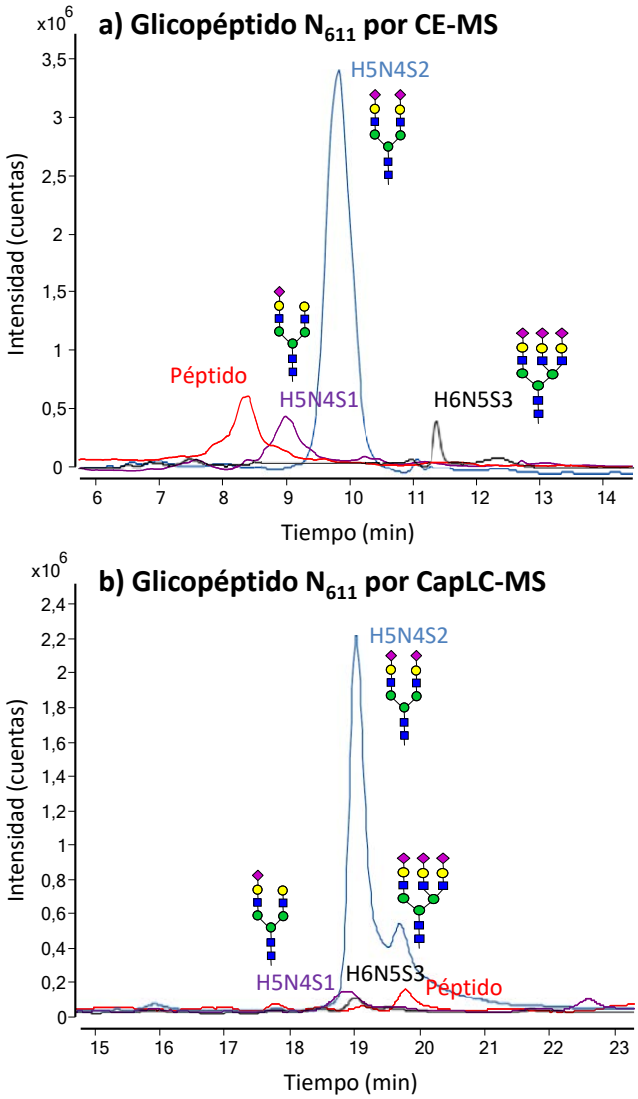


Figura 5.16: EIEs y EICs de las principales glicoformas del glicopéptido N_{611} detectadas en un patrón de hTf analizado por a) CE-MS y b) CapLC-MS. Los símbolos utilizados para la representación de los glicanos siguen las reglas sugeridas por el *Consortium for Functional Glycomics* (CFG) [13].

Una de las primeras ventajas que se puede observar al analizar los resultados obtenidos con CE es que los tiempos de análisis son más reducidos, no tan solo debido a que la separación tiene lugar en menor tiempo (14 min en CE *versus* 23 min en CapLC), sino que también los tiempos de análisis totales son inferiores (30 min en CE respecto a los 75 min en CapLC), principalmente debido a que los tiempos de lavado y equilibrado necesarios en CapLC son muy superiores. No obstante, la gran ventaja de la CE como técnica de separación es la posibilidad de separar las diferentes glicoformas de un mismo glicopéptido, tal y como se muestra en la Figura 5.16-a. La separación de las glicoformas aumenta la sensibilidad debido a su mejor ionización, ya que al estar separadas entre sí, no se suprimen la ionización mutuamente (la intensidad de las diferentes glicoformas es similar en CE y CapLC a pesar de que el volumen de inyección en CE era inferior, 83 nL en CE *versus* 150 nL en CapLC). Otra ventaja añadida de la separación entre glicoformas es que se detectan con mayor fiabilidad ya que se descarta que alguna de las glicoformas detectadas sea debida a la fragmentación en la fuente de ionización, de otra glicoforma, ya que, de ser así, tanto la glicoforma fragmento como la original comigrarían.

Sin embargo la reproducibilidad entre inyecciones es inferior en CE, pudiendo llegar a migrar una misma glicoforma a tiempos muy dispares, incluso de más de 3 min. de diferencia. En cambio, en general las diferentes glicoformas se obtienen en un mismo tiempo de retención en CapLC, incluso en muestras muy diferentes.

No obstante, en muchas ocasiones, para poder completar el mapa glicopeptídico de una proteína, es necesario el uso de ambas técnicas de separación al considerarse técnicas complementarias. Además, el acoplamiento de ambas técnicas con la espectrometría de masas, permite obtener una identificación inequívoca de las diferentes glicoformas. Esto es extremadamente útil en aquellos casos donde los métodos de análisis actuales para CDT, basados en IEF o HPLC / CE con detección UV, no permiten dar un diagnóstico totalmente fiable. El ejemplo más destacado es la confusión en el diagnóstico de CDG, cuando la hTf

presenta diferentes **variantes genéticas** (modificaciones de la secuencia aminoacídica debidas a una variación genética). Hasta ahora se han descrito más de treinta variantes genéticas de la hTf tan solo en individuos caucásicos, de las cuales, la variante C (sobretodo C1 y C2) es la predominante en todas las razas [297, 298]. Sin embargo, otras variantes son relativamente comunes, sobretodo en la población no caucásica, por ejemplo la variante B y la variante D. Combinaciones heterocigóticas de estas variantes (BC o CD), dan lugar a una distribución de bandas o picos de las glicoformas intactas de la hTf que se asemeja a la obtenida para ciertos tipos de CDG [299–302]. Esto es debido al cambio en la secuencia aminoacídica que provoca que la banda correspondiente a la variante comigre o coeluya en la zona de las glicoformas CDT [55, 298–302]. Para solventar este problema, normalmente es necesario hacer análisis adicionales, hecho que complica el diagnóstico debido a la limitada cantidad de muestra de la que se dispone en la mayoría de casos. Con los métodos desarrollados en esta tesis doctoral, en los que el análisis de la glicosilación se realiza a nivel glicopeptídico y utilizando la espectrometría de masas, es posible identificar, con tan solo un análisis, si la hTf presenta una modificación de la secuencia aminoacídica o si existe alteración en su glicosilación, incluso si la variante genética afecta a uno de los puntos de glicosilación.

Adicionalmente, el estudio de la glicosilación a nivel glicopeptídico llevado a cabo en esta tesis, no sólo permite detectar cómo se altera la glicosilación de una proteína debido a una enfermedad, sino también saber si esta alteración afecta de igual manera a todos los puntos de glicosilación. En general, tanto en el estudio del alcoholismo como de los CDG, ya sea por CapLC o CE, se pudo comprobar que ambos puntos de N-glicosilación se veían alterados de igual manera.

5.2.3. Aplicabilidad de los modelos teóricos de migración al estudio del comportamiento electroforético de péptidos y glicopéptidos por CE-MS

Otra de las grandes ventajas de la CE es la posibilidad de predecir la movilidad electroforética de los analitos utilizando las denominadas relaciones semiempíricas clásicas entre la movilidad electroforética y la relación carga/masa. Como se introdujo anteriormente, la movilidad electroforética (**m_e**) está relacionada con la carga (**q**) y la masa de la especie analizada (**M**), según la siguiente ecuación (apartado 1.3.4.2):

$$m_e = A\frac{q}{M^{\alpha}} \qquad \textbf{Ec. 1}$$

Este tipo de modelizaciones presentan un gran interés ya que la predicción de la movilidad electroforética permite acelerar el desarrollo y optimización de métodos basados en CE, así como confirmar las asignaciones estructurales basadas en las masas moleculares medidas por MS.

En esta tesis doctoral se han evaluado tres métodos de modelización clásicos para predecir la movilidad electroforética de los glicopéptidos y péptidos presentes en un digesto. Para llevar a cabo este estudio, que se recoge en el artículo 3.2, se escogió la **eritropoyetina humana recombinante (rhEPO)** como glicoproteína modelo, dado su especial interés como biofármaco y en el control antidopaje. Otro motivo importante por el que se seleccionó esta glicoproteína fue por el gran número de glicoformas de glicopéptidos que se obtienen tras su digestión enzimática, lo que se debe a su elevada heterogeneidad, ocasionada por un alto porcentaje de glicosilación.

Primeramente, la rhEPO se digirió con tripsina (rhEPO-T) en 50 mM NH_4HCO_3 pH 7.9 a 37˚C durante 18h, previa reducción y alquilación con DTT e IAA. El digesto se analizó mediante CE-MS para obtener los valores de tiempo de migración experimentales tanto de péptidos como de glicopéptidos (O_{126}, $N_{24\text{-}36}$ y N_{83}). Debido a que la mayoría de las glicoformas de los

glicopéptidos presentan ácidos siálicos, el cual puede influir considerablemente en la movilidad electroforética, al contener un ácido carboxílico y consecuentemente carga negativa en el pH de la separación (BGE: 50 mM HFor / 50 mM HAc pH: ~2.2), se decidió digerir también la rhEPO con tripsina y neuraminidasa (rhEPO-TN), y así aumentar el número de glicopéptidos sin ácidos siálicos (S o SiA) presentes en el modelo. Para la muestra de rhEPO-TN, tras la digestión tríptica de la rhEPO, se digirió la muestra con neuraminidasa en 50 mM NH_4Ac (pH 5.0) a 37˚C durante 18h.

En la Tabla 5.9, se recogen todos los péptidos y glicopéptidos que se han utilizado para llevar a cabo la modelización, su número de referencia (**N**), así como su masa teórica monoisotópica ($\mathbf{M_{teo}}$), carga (**q**), tiempo de migración relativo (**t'**) y fracción glicosilada. Para obtener la carga de péptidos y glicopéptidos al pH de la separación electroforética (pH ~2.2), se utilizó la ecuación de Sillero y Ribeiro [303], utilizando los valores de pKa para cada aminoácido proporcionados por Rickard [206] y un valor de pKa de 2.6 para el ácido siálico. Debido al elevado número de glicoformas que pueden presentar los diferentes glicopéptidos de la rhEPO, para poder realizar una modelización correcta sólo se utilizaron aquellas glicoformas que presentaban una relación señal/ruido adecuada. En cambio, como los péptidos generalmente se ionizan mejor, todos los péptidos detectados (~97% de la cobertura de la proteína) se tuvieron en cuenta en la modelización.

Tabla 5.9: Lista de péptidos y glicopéptidos incluidos en el estudio de los modelos clásicos de relación semiempírica (**set de modelización**) de un digesto de rhEPO-T y rhEPO-TN.

N	Péptidos	M_{teo} (Da)	*q*	t'		Fracción glicosilada (%)[a]
				Media	s (n=3)	
1	APPR	439.2543	1.8882	0.759	0.005	-
2	LICDSR	762.3694	1.8288	0.845	0.003	-
3	VLER	515.3067	1.8819	0.792	0.004	-
4	YLLEAK	735.4167	1.8819	0.832	0.003	-
5	VNFYAWK	926.4650	1.8882	0.836	0.003	-
6	MEVGQQAVEVWQGLALLSEAVLR[c]	2525.3311	1.8694	1.00	-	-
7	AVSGLR	601.3547	1.8882	0.806	0.005	-
8	SLTTLLR	802.4912	1.8882	0.837	0.004	-
9	ALGAQK	586.3438	1.8882	0.805	0.004	-
10	TITADTFR	923.4712	1.8288	0.867	0.003	-

11	LFR	434.2641	1.8882	0.761	0.005	-
12	VYSNFLR	897.4708	1.8882	0.841	0.004	-
13	GK	203.1270	1.8882	0.686	0.003	-
14	LK	259.1896	1.8882	0.718	0.005	-
15	LYTGEACR	968.4385	1.8819	0.867	0.004	-
N-glicopéptidos en rhEPO-T						
N	N_{83} (GQALLVNSSQPWEPLQLHVDK)[b]					
16	H6N5F1S2	5074.1958	2.1547	1.14	0.003	53.5
17	H6N5F1S3	5365.2911	1.8209	1.26	0.009	56.0
18	H7N6F1S2	5439.3280	2.1547	1.14	0.001	56.6
19	H7N6F1S3	5730.4234	1.8209	1.26	0.009	58.8
20	H7N6F1S4	6021.5188	1.4870	1.30	0.01	60.8
21	H8N7F1S2	5804.4602	2.1547	1.15	0.003	59.4
22	H8N7F1S3	6095.5556	1.8209	1.26	0.009	61.3
23	H8N7F1S4	6386.6510	1.4870	1.31	0.01	63.1
24	H9N8F1S3	6460.6878	1.8209	1.26	0.009	63.5
25	H9N8F1S4	6751.7832	1.4870	1.31	0.01	65.1
26	H10N9F1S4	7116.9154	1.4870	1.31	0.06	66.9
N	N_{24}-N_{38} (EAENITTGCAEHCSLNENITVPDTK)[b]					
27	H14N12F2S7	9837.7194	0.4666	1.49	0.02	71.5
28	H14N12F2S8	10128.8149	0.1327	1.56	0.02	72.3
29	H15N13F2S7	10202.8516	0.4666	1.49	0.02	72.5
30	H15N13F2S8	10493.9471	0.1327	1.55	0.02	73.3
N-glicopéptidos en rhEPO-TN						
N	N_{83} (GQALLVNSSQPWEPLQLHVDK)[b]					
31	H6N5F1	4492.0048	2.8224	1.03	0.0006	47.5
32	H7N6F1	4857.1370	2.8224	1.04	0.001	51.4
33	H8N7F1	5222.2692	2.8224	1.05	0.0008	54.8
34	H9N8F1	5587.4014	2.8224	1.06	0.002	57.8
35	H10N9F1	5952.5336	2.8224	1.08	0.0005	60.4
N	N_{24}-N_{38} (EAENITTGCAEHCSLNENITVPDTK)[b]					
36	H14N12F2	7800.0515	2.8036	1.12	0.001	64.1
37	H15N13F2	8165.1837	2.8036	1.13	0.001	65.7
O-glicopéptidos en rhEPO-T						
N	O_{126} (EAISPPDAASAAPLR)[b]					
38	H1N1	1829.8894	1.8226	0.980	0.003	20.0
39	H1N1S1	2120.9851	1.4887	1.09	0.002	30.9
40	H1N1S2	2412.0808	1.1548	1.25	0.009	39.3
O-glicopéptidos en rhEPO-TN						
N	O_{126} (EAISPPDAASAAPLR)[b]					
41	H1N1	1829.8894	1.8226	1.01	0.0006	20.0

[a]Fracción glicosilada: (masa del glicano / masa del glicopéptido) x 100.

[b]Secuencia peptídica de los glicopéptidos N_{83}, N_{24}-N_{38}, O_{126}.

[c]El péptido número 6 se utilizó como referencia en el cálculo del tiempo de migración relativo (tiempo de migración medio t=7.68 min.).

Tal y como se puede observar en la Tabla 5.9, existe una gran heterogeneidad dentro de cada grupo. En el caso de los péptidos, éstos se diferencian principalmente en la longitud de la cadena aminoacídica y, por consiguiente, en su masa molecular. En cambio, en los glicopéptidos, su elevada heterogeneidad se debe al tipo (O u N-glicopéptido), la complejidad

o ramificación de la cadena hidrocarbonada y a la presencia o no de ácidos siálicos. Para aumentar la reproducibilidad entre replicados y poder comparar entre diferentes muestras, por ejemplo, entre aquellas digeridas con tripsina (rhEPO-T) y las digeridas con tripsina y neuraminidasa (rhEPO-TN), se decidió modelizar a partir del tiempo de migración relativo de cada especie respecto a un péptido del digesto. El péptido de referencia escogido fue el *MEVGQQAVEVWQGLALLSEAVLR*, de intensidad parecida al de las glicoformas abundantes y un tiempo de migración próximo a la zona donde migraban la mayoría de los glicopéptidos.

Tal y como se comentó en la introducción, a fin de obtener la mejor correlación lineal entre $\mathbf{m_e}$ y $\mathbf{q/M^{\alpha}}$, se evaluaron tres modelos clásicos: ***Classical polymer model*** ($\alpha = 1/2$), ***Stoke's law*** ($\alpha = 1/3$) y ***Offord's Surface law*** ($\alpha = 2/3$). Cada uno de estos modelos asume que la especie analizada tiene una forma concreta y que se comporta de una manera específica bajo las fuerzas que se aplican durante la separación electroforética [206–208]. Con el set de péptidos y glicopéptidos de la Tabla 5.9, se construyeron los tres modelos lineales, uno para cada valor de α, separando péptidos y glicopéptidos, pues se observó que su movilidad era muy diferente para poderlos englobar en un mismo modelo. Los parámetros del ajuste de cada modelo con ambos grupos se muestran en la Tabla 5.10.

Tabla 5.10: Parámetros obtenidos tras la aplicación de los modelos clásicos de relación semiempírica a los péptidos y glicopéptidos de un digesto de rhEPO-T y rhEPO-TN (y = b + mx).

	$q/M^{1/2}$			$q/M^{1/3}$			$q/M^{2/3}$		
	b (±s)	**m (±s)**	**R^2**	**b (±s)**	**m (±s)**	**R^2**	**b (±s)**	**m (±s)**	**R^2**
***N*- y *O*-glicopéptidos (con y sin SiA)**	1.557 (±0.008)	-13.3 (±0.3)	0.9891	1.57 (±0.02)	-3.33 (±0.14)	0.9569	1.527 (±0.015)	-49 (±2)	0.9598
Péptidos	1.054 (±0.015)	-3.1 (±0.2)	0.9120	1.18 (±0.02)	-1.65 (±0.08)	0.9443	0.989 (±0.015)	-6.4 (±0.5)	0.8756

s: desviación estándar.

Si nos fijamos en los valores de R^2 de la tabla, podemos observar que el c*lassical polymer model* es el modelo que presenta mejor linealidad para glicopéptidos (R^2=0.9891), y en cambio, los péptidos parecen ajustarse mejor al modelo de *Stoke's law* (R^2=0.9443). Para intentar mejorar los resultados obtenidos, se consideró el denominado efecto de supresión de la carga electroestática, que simplemente indica que, cuando la carga de una especie es muy elevada, se puede producir un efecto de supresión, que hace que una parte de su carga no influya en su movilidad electroforética. Este efecto se puede evaluar con relativa facilidad si se sustituye en la ecuación 1 (Ec.1), la **q** por **ln(1+q)**. Adicionalmente, se consideró que la fricción dieléctrica podía tener un efecto en la movilidad electroforética de las diferentes especies tal y como sugiere el modelo de Hubbard-Onsager [304, 305]. No obstante, ninguna de estas dos modificaciones mejoró los resultados ya obtenidos con los modelos clásicos.

Debido a la gran heterogeneidad existente dentro de cada grupo (péptidos y glicopéptidos) las diferencias entre los tres modelos no son muy pronunciadas, hecho que impide extraer conclusiones fiables respecto a la forma molecular que adoptan los diferentes compuestos. Pese a ello, parece claro que son necesarios dos modelos diferentes para péptidos y glicopéptidos para predecir correctamente su comportamiento electroforético. Este hecho podría atribuirse principalmente a la diferencia de masa más que a la de carga (Tabla 5.9). Por un lado, la mayor masa molecular de los glicopéptidos provoca un aumento significativo de los **t'**, respecto a los péptidos. Además, pese a la gran variedad de cargas que presentan los glicopéptidos, éstos se ajustan bien al mismo modelo lineal, lo que refuerza la idea de que la masa molecular es el elemento diferenciador.

A continuación, en la Figura 5.17 se muestra un gráfico donde se representa **t'** respecto a la $\mathbf{q/M^{\alpha}}$, teniendo en cuenta el mejor modelo para cada grupo. En esta figura además se incluye, en forma de línea negra, la recta que se obtiene al calcular el tiempo de migración relativo con el modelo correspondiente para cada grupo.

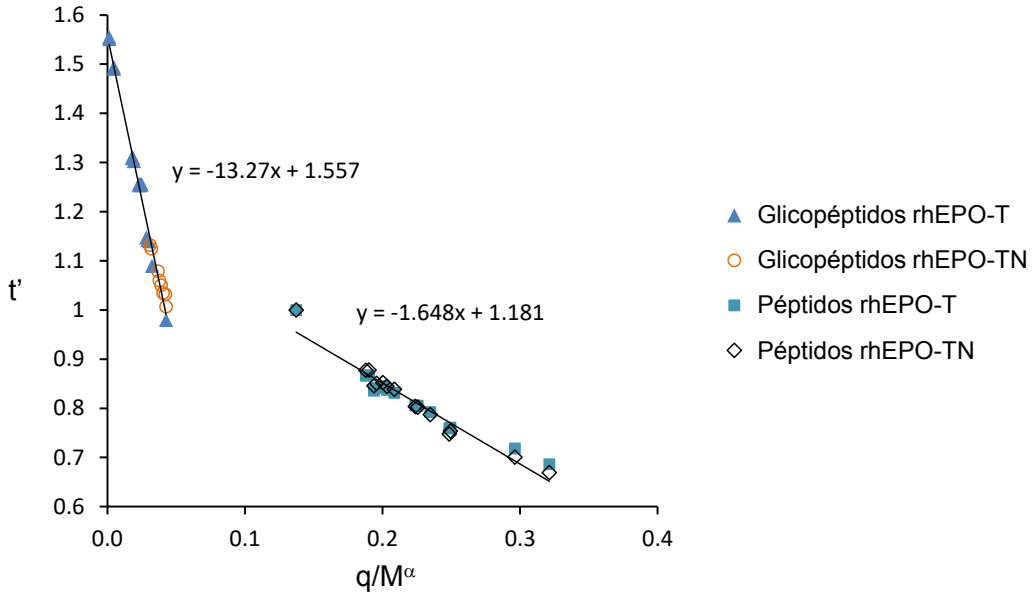


Figura 5.17: Gráfico del tiempo de migración relativo (**t'**) experimental (símbolos) y estimado (línea) al aplicar los modelos clásicos de relación semiempírica: **t'** vs $\mathbf{q/M^{1/3}}$ (*Stoke's law*) y $\mathbf{q/M^{1/2}}$ (*classical polymer model*) para péptidos y glicopéptidos, respectivamente, de un digesto de rhEPO-T y rhEPO-TN utilizando la mejor correlación lineal.

Para comprobar la idoneidad de los modelos en la predicción del tiempo de migración se realizaron dos validaciones: a) una interna, donde los modelos se aplicaron para predecir la $\mathbf{m_e}$ de un conjunto de glicopéptidos de la propia rhEPO, no incluidos en la elaboración del modelo y con una relación señal ruido suficiente; b) una externa, donde los modelos se aplicaron para predecir la $\mathbf{m_e}$ de un conjunto de péptidos y glicopéptidos de otra glicoproteína, en concreto, la hTf. Para visualizar mejor la predicción de la movilidad electroforética y de los tiempos de migración relativos, se simularon los electroferogramas de las especies modelizadas y se compararon con los electroferogramas experimentales obtenidos por CE-MS. En la Figura 5.18, se recogen los electroferogramas experimentales y simulados para una muestra de: a) rhEPO-T y b) rhEPO-TN, donde se simulan algunos péptidos y glicopéptidos incluidos en la elaboración del modelo, c) rhEPO-T, donde se simulan algunos glicopéptidos no incluidos en la elaboración del modelo y d) hTf-T (transferrina humana digerida con tripsina), donde se simulan algunos de sus péptidos y glicopéptidos. Debido a que la digestión de rhEPO-T y rhEPO-TN seguían dos

procedimientos diferentes, se tuvieron que realizar dos electroferogramas separados, uno para cada muestra.

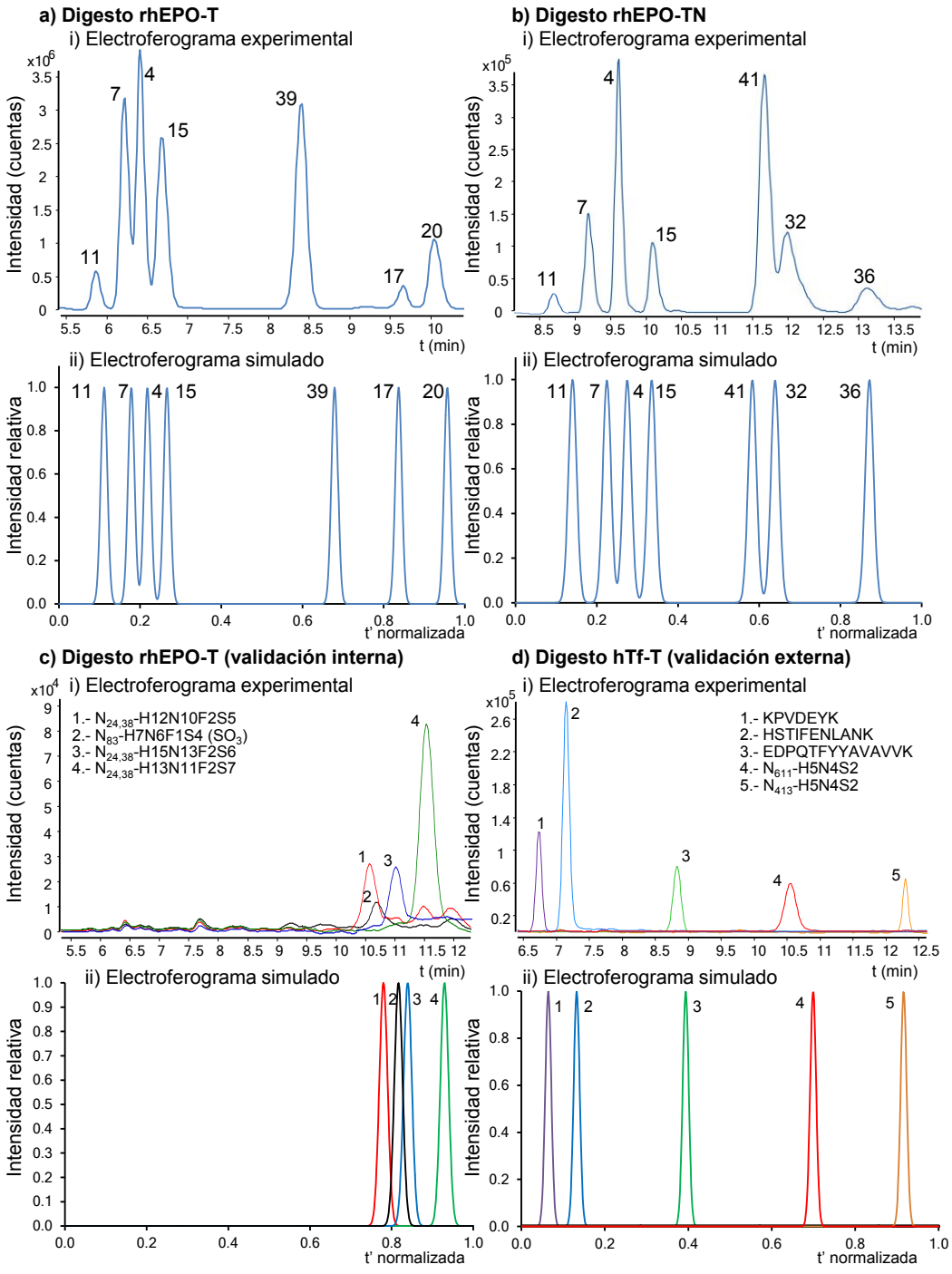


Figura 5.18: Electroferogramas experimentales y simulados para: a) una muestra de rhEPO-T; b) una muestra de rhEPO-TN; c) una muestra de rhEPO-T de validación interna; d) una muestra de hTf-T de validación externa. Los péptidos y glicopéptidos de los gráficos a y b se incluyeron en el set de modelización, en cambio, los de los gráficos c y d no se incluyeron en el set de modelización.

Para facilitar la interpretación de los electroferogramas simulados, como se puede observar en la Figura 5.18 (ii), a cada pico simulado se le asignó una intensidad relativa fija de 1 y una anchura a media altura de 0.005. Además, la escala del electroferograma simulado fue normalizada de 0 a 1 $((t_i'-t_{min}')/(t_{max}'-t_{min}'))$. En todos los casos la correlación entre el electroferograma experimental y el simulado es muy buena, teniendo en cuenta la gran diversidad de péptidos y glicopéptidos, incluso de diferentes glicoproteínas, lo que demuestra la capacidad de los modelos establecidos para predecir correctamente la movilidad electroforética. También cabe destacar como los péptidos y glicopéptidos salen en ventanas de tiempo separadas, tanto en los electroferogramas experimentales como en los simulados, lo que refuerza el hecho que, para cada grupo, un modelo diferente es el más apropiado. Las ligeras diferencias que se pueden observar en algún caso son, probablemente, debidas a diferentes factores: inexactitud en algunos valores de pK_a lo que puede afectar al cálculo de la carga o, por ejemplo, los efectos de difusión longitudinal y conductividad diferente entre los tampones de las muestras y el electrolito de separación, los cuales pueden afectar ligeramente a la separación de las especies y no se tuvieron en cuenta en la predicción. También es importante destacar que, en el caso de la validación interna, se incluyó un glicopéptido con un grupo sulfónico (N_{83}-H7N6F1S4(SO_3)), y aunque ninguno de los glicopéptidos incluidos en la elaboración del modelo tenía este grupo, la concordancia entre el electroferograma experimental y el simulado es igualmente buena. Este hecho demuestra que, incluso con grupos funcionales no presentes en la elaboración del modelo, se puede predecir la movilidad electroforética de ciertos compuestos siempre y cuando no sean grupos muy diferentes, y que el pK_a utilizado para el cálculo de la carga sea lo más exacto posible.

En conclusión, los modelos de migración establecidos en esta tesis doctoral pueden resultar una herramienta útil, rápida y sencilla para predecir a priori los tiempos de migración y la separación electroforética por CE-MS de péptidos y glicopéptidos de glicoproteínas de interés biológico.

5.3. Estudio de la glicosilación de glicoproteínas por movilidad iónica y espectrometría de masas

En esta tesis se ha demostrado como las técnicas de separación miniaturizadas acopladas a la MS (CapLC-MS y CE-MS) pueden resultar muy útiles en el estudio de la glicosilación de proteínas. Sin embargo, detectar cambios en la abundancia y en la composición de los glicanos unidos a la proteína puede no ser suficiente para diagnosticar ciertas enfermedades y buscar dianas terapéuticas. A veces un proceso patológico puede afectar sólo a un isómero del glicano y no a los demás, siendo necesario, en este caso, el análisis de los isómeros de los diferentes glicanos para entender la extensión de la enfermedad [251–254].

Tal y como se mencionó en la introducción, los isómeros de glicanos de tipo complejo están relacionados, generalmente, con el tipo de enlace de los ácidos siálicos (SiA) terminales (α-2→3 o α-2→6), o de las fucosas, que pueden estar unidas a la antena o al *core* del glicano. En estos casos, la CapLC-MS y la CE-MS no permiten diferenciar entre los diferentes isómeros. En la literatura, algunos autores han propuesto el uso de fases estacionarias como la de interacción hidrofílica zwitteriónica (*Zwitterionic Hydrophilic Interaction Liquid Chromatography* - ZIC-HILIC) o el carbono grafitizado poroso (*Porous Graphitic Carbon* - PGC) en cromatografía de líquidos o el uso de la espectrometría de masas en tándem (MS/MS) para identificar y caracterizar isómeros de glicoconjugados [147–151]. Sin embargo, estas metodologías están todavía en fase de desarrollo y en muchos casos son laboriosas y complejas.

La Movilidad Iónica (IM) es una técnica de separación en fase gas que permite separar especies por su forma y tamaño, además de su carga, y que junto con la detección por MS (IM-MS) ha conseguido popularidad en los últimos años gracias a su capacidad para separar compuestos isoméricos [140–143, 152–155]. Por estos motivos, en esta tesis (artículo 4.1) se

ha evaluado la eficacia de la ***travelling wave ion mobility spectrometry*** **(TWIMS)**, para estudiar los isómeros de los glicanos de la mTf a tres niveles: glicoproteína intacta, glicopéptidos y glicanos.

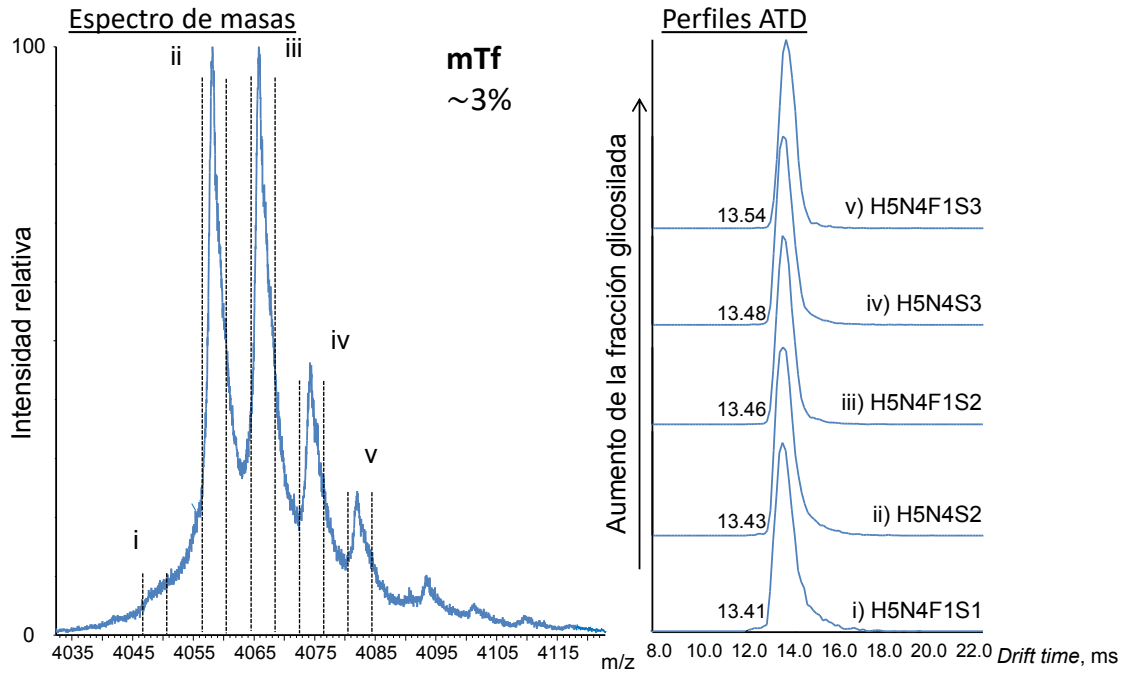
5.3.1. Glicoproteína intacta

El estudio de la glicoproteína intacta por MS se puede llevar a cabo de dos maneras: a partir de la proteína desnaturalizada (utilizando tampones ácidos) o en condiciones nativas, donde se utiliza un tampón neutro para conservar su estructura terciaria y sus interacciones no covalentes. Una de las principales diferencias al estudiar una glicoproteína en condiciones nativas por MS es que adquiere menor número de cargas, lo que permite que las diferentes glicoformas se puedan resolver con mayor facilidad.

En este trabajo se estudiaron diferentes glicoproteínas (mTf, hTf y hAGP) y se optimizaron los parámetros de la fuente de nanoelectroespray (nanoESI) y la concentración del tampón neutro de acetato de amonio (NH_4Ac) para obtener un espray estable y poder trabajar en condiciones nativas. Además, a fin de aumentar la sensibilidad y mejorar la resolución de las glicoformas, se evaluaron tres tratamientos de desalado de la muestra: 1) diálisis; 2) exclusión por tamaño y 3) ultracentrifugación, siendo este último el más efectivo y con el que se obtuvieron mejores resultados.

En la Figura 5.19, se muestra el espectro de masas del ion de carga +19 de la mTf intacta (MS ampliado en la zona de masas correspondiente a esta carga), obtenido mediante nanoESI-IM-MS en estas condiciones óptimas de NH_4Ac y de desalado (100 mM NH_4Ac y ultracentrifugación, respectivamente). En concreto, se puede observar como cada pico coincide con la masa de una glicoforma distinta, tal y como se muestra en la tabla incluida en la Figura 5.19. Una vez obtenida esta separación de las glicoformas intactas, se midió el **tiempo de deriva (*drift time*)** de cada glicoforma. Este *drift time* es el tiempo que tarda la molécula en

recorrer la celda de movilidad y está relacionado con la forma y tamaño de la molécula, además de su carga. En concreto, un *drift time* mayor indica que la estructura o forma de la molécula es más grande, o, dicho de otro modo, que su sección transversal de colisión (***collision cross section* – CCS**) es mayor. En el caso de moléculas de *drift time* menor ocurre justo lo contrario. Las distribuciones del tiempo de llegada (***arrival time distribution* - ATD**) de cada glicoforma y los valores de *drift time* correspondientes también se incluyen en la Figura 5.19 (i-v).



Glicoforma mTf	m/z experimental nominal $[M+19H]^{19+}$	Masa experimental nominal glicoproteína	Masa teórica nominal glicoproteína
i) H5N4F1S1	4048	76893	76908
ii) H5N4S2	4058	77083	77069
iii) H5N4F1S2	4066	77235	77216
iv) H5N4S3	4074	77387	77377
v) H5N4F1S3	4082	77539	77523

Figura 5.19: Espectro de masas del ion de carga +19 de la proteína intacta (MS ampliado en las masas correspondientes a esta carga) en un patrón de mTf y perfil de la distribución del tiempo de llegada (ATD) por nanoESI-IM-MS de cada glicoforma (i-v). El valor aproximado incluido en la figura indica el porcentaje de glicosilación de la glicoproteína.

Como se puede observar en la figura, las glicoformas de la mTf dan valores de *drift time* muy similares, aunque se observa una cierta tendencia: el *drift time* aumenta a medida que la glicosilación también aumenta, es decir, cuando la glicoforma está más ramificada. Este comportamiento fue más acusado al analizar la hTf y la hAGP, glicoproteínas con más porcentaje de glicosilación que la mTf (ver Figura 5.20). Concretamente, en el caso de la hAGP el aumento del *drift time* fue más perceptible, demostrando que un cambio en la glicosilación afecta a la forma, y, por consiguiente, a la CCS de aquellas glicoproteínas con mayor porcentaje de glicosilación.

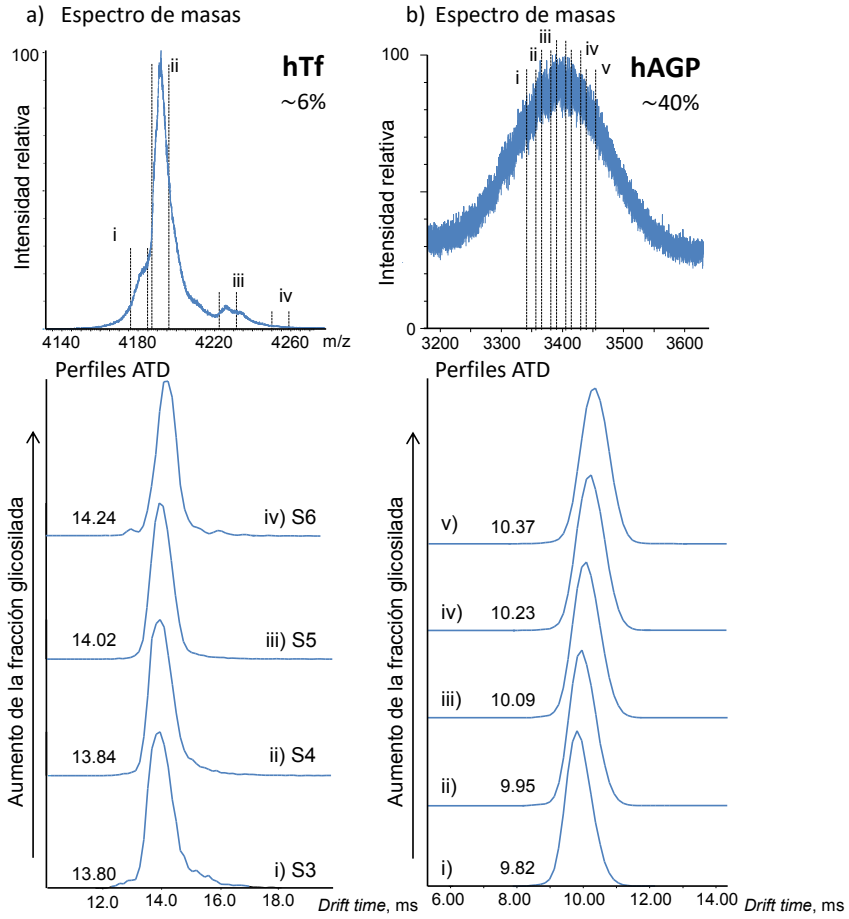


Figura 5.20: a) Espectro de masas del ion de carga +19 (MS ampliado en las masas correspondientes a esta carga) de la proteína intacta en un patrón de hTf y perfil ATD por nanoESI-IM-MS de cada glicoforma (i-iv). b) Espectro de masas del ion de carga +12 de la proteína intacta en un patrón de hAGP y perfil ATD por nanoESI-IM-MS de cada sección del espectro de masas (i-v). El valor aproximado incluido en la figura indica el porcentaje de glicosilación de la glicoproteína.

No obstante, los resultados obtenidos a nivel de glicoproteína intacta, únicamente permitieron evaluar el efecto global que provocaba un cambio considerable de glicosilación en la CCS de la proteína. Otras modificaciones más sutiles, como, por ejemplo, un cambio en el enlace del ácido siálico de las glicoformas, no parecían provocar una modificación de la CCS. Por este motivo, se decidió continuar con el análisis de glicoconjugados de menor tamaño, como los glicopéptidos y los glicanos.

5.3.2. Glicopéptidos y glicanos

El estudio por nanoESI-IM-MS con glicopéptidos y glicanos también se inició con la mTf, que presenta un único punto de glicosilación, y por lo tanto un único glicopéptido. Dada la baja complejidad de esta glicoproteína, se decidió realizar el análisis haciendo infusión directa, sin utilizar una técnica de separación previa. Sin embargo, el gran número de péptidos originados en la digestión suprimieron la ionización del glicopéptido correspondiente y por este motivo se utilizó **nano cromatografía de líquidos (nanoLC)** como técnica de separación, ya que permitía un acoplamiento más rápido y sencillo con la fuente de nanoESI. Para el análisis del digesto tríptico de la mTf se empleó una nanocolumna C18 (Peptide BEH C18, 1.7 µm diámetro de partícula, 130 Å tamaño de poro, 100 x 0.1 mm L_T x i.d.; WATERS®) y una fase móvil H_2O:ACN con 0.1% (v/v) de HFor. En este caso, se detectaron prácticamente las mismas glicoformas ya observadas anteriormente para el estudio de la mTf por CapLC-MS (Tabla 5.6, apartado 5.1.3).

A continuación, se obtuvieron los diferentes ATD por IM-MS de cada glicoforma del glicopéptido, observándose un único valor de *drift time* en cada caso. Al igual que sucede con la mTf intacta, estos resultados nos indican que, si en la mTf existen isómeros debidos al tipo de enlace del SiA, éstos no alteran significativamente la CCS del glicopéptido. Para poder solventar este problema, se decidió reducir el tamaño de la molécula fragmentando el glicopéptido por MS/MS y analizar por movilidad iónica un fragmento que todavía conservase unido el SiA. En la Figura 5.21, se muestra el espectro de MS/MS de la glicoforma más intensa

del glicopéptido de la mTf (H5N4S2), así como el ATD del fragmento que aún conserva el ácido siálico, H1N1S1. Como se puede observar en la Figura 5.21-c, el fragmento H1N1S1 muestra dos picos con valores de drift time diferentes, pudiendo ser esta diferencia de *drift time* ocasionada por el tipo de enlace del SiA. Estos resultados coinciden con los descritos por otros autores en el análisis de glicopéptidos por IM-MS [143, 306].

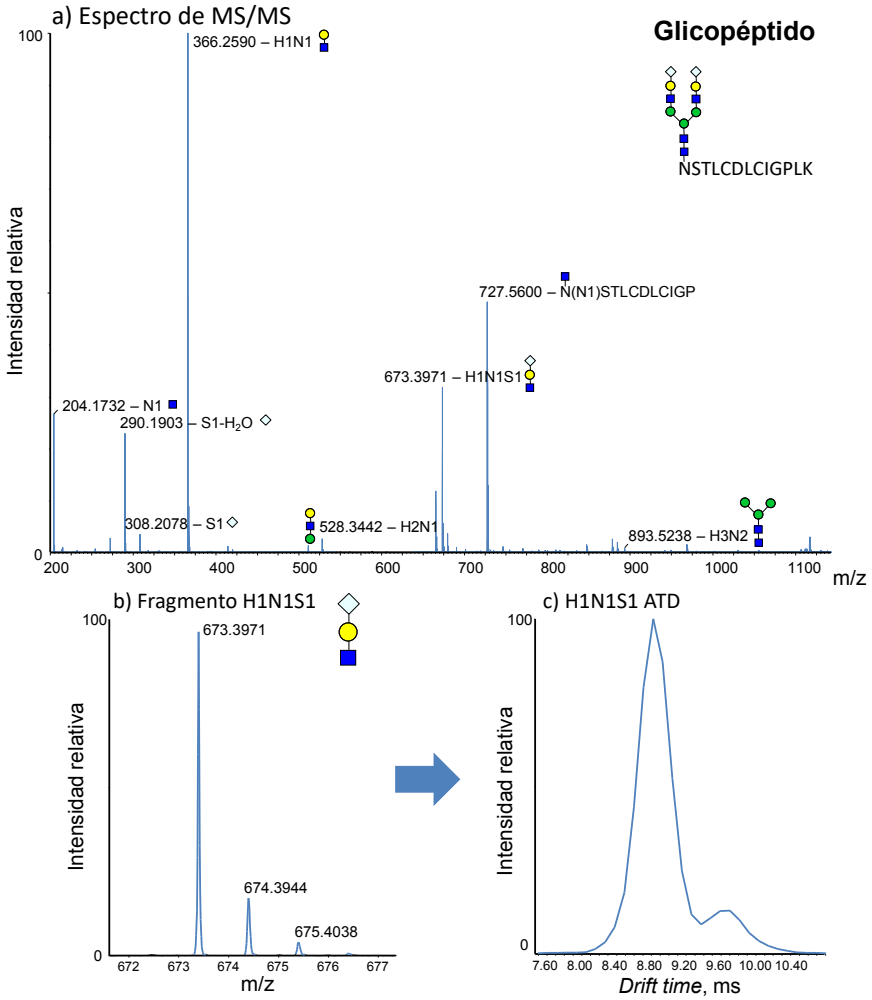


Figura 5.21: a) Espectro de MS/MS por CID del glicopéptido N(H5N4S2)STLCDLCIGPLK de la mTf; b) espectro de masas del fragmento que aún conserva el ácido siálico (H1N1S1) y tiene una intensidad adecuada; c) ATD del fragmento H1N1S1 (rango de masa: 673.3-673.5). Los símbolos utilizados para la representación de los glicanos siguen las reglas sugeridas por el *Consortium for Functional Glycomics* (CFG) [13].

Con el fin de comprobar si se observaban también valores de *drift time* diferentes a nivel de glicano, éstos se analizaron por nanoESI-IM-MS tras la digestión de la mTf con PNGasa F (0.1% β-ME (v/v) y 0.1% SDS (m/v) en 50 mM NH_4HCO_3 (pH 7.9) con 1% NP-40 (v/v), incubación de 18h a 37˚C). En este caso, los glicanos se analizaron por infusión directa y en polaridad negativa, tras su purificación por extracción en fase sólida (SPE) con cartuchos de carbono grafitizado poroso (HyperSep™ Hypercarb™, Thermo Scientific™), con el fin de aislarlos de la proteína intacta. La Tabla 5.11 muestra los glicanos detectados, los cuales coinciden con las glicoformas detectadas del glicopéptido por nanoLC-IM-MS.

Tabla 5.11: Masa molecular (M_r) teórica, M_r experimental, error de masa (ppm) y carga observada para cada glicano detectado de la mTf.

	M_r teórica	M_r experimental	Error de masa (ppm)	Carga observada
Glicanos				
H3N3S1	1420.4975	1420.4659	22	-1
H3N3F1S1	1566.5554	1566.5692	9	-1
H5N4S1	1947.6825	1947.7031	11	-1
H5N4F1S1	2093.7404	2093.7891	23	-1
H5N4S2	2254.7729	2254.7650	4	-2
H5N4F1S2	2400.8308	2400.8240	3	-2
H5N4S3	2561.8632	2561.9150	20	-2
H5N4F1S3	2707.9211	2707.9470	10	-2
H6N5S3	2926.9954	2926.9334	21	-2, -3
H6N5F1S3	3073.0533	3073.0036	16	-3

En la Figura 5.22, se muestra la separación obtenida para diferentes glicanos tras optimizar la altura y la velocidad de la onda de la celda de movilidad, WH y WV, respectivamente. Para poder visualizar mejor la presencia de diferentes isómeros, se realizó, una reconstrucción gaussiana de cada uno de los picos observados en cada ATD utilizando el programa Origin (versión 8, OriginLab Corp., Northampton, MA, USA). En los casos donde no fue posible utilizar el programa mencionado, la reconstrucción gaussiana se realizó manualmente.

En el caso del glicano mayoritario (H5N4S2), la separación es suficientemente buena como para observar que existen dos isómeros, siendo el de mayor *drift time* y mayor CCS el más intenso, contraponiéndose a lo observado en el estudio del glicopéptido (Figura 5.21-c), donde el pico más intenso correspondía al isómero de menor *drift time*. Esto puede ser debido a que se tratan de moléculas completamente diferentes cuya CCS puede variar de manera diferente al cambiar el tipo de enlace del SiA.

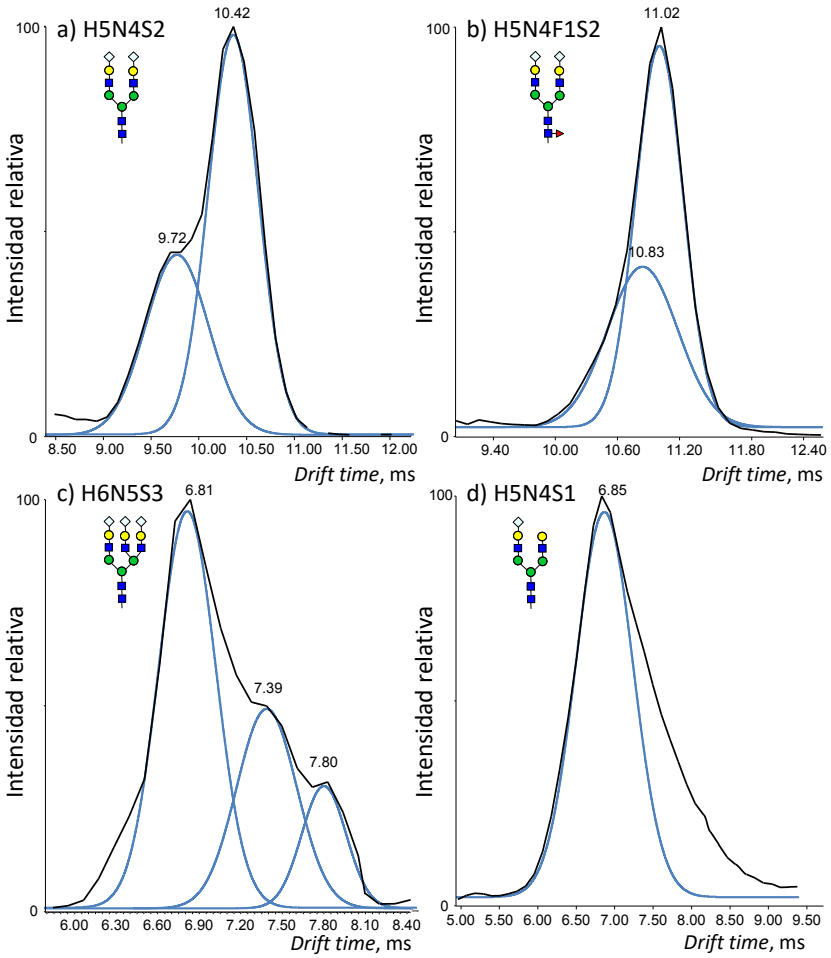


Figura 5.22: ATD de diferentes glicanos de la mTf obtenidos por nanoESI-IM-MS: a) H5N4S2, b) H5N4F1S2, c) H6N5S3 y d) H5N4S1. Los símbolos utilizados para la representación de los glicanos siguen las reglas sugeridas por el *Consortium for Functional Glycomics* (CFG) [13]. Los picos gaussianos simulados se realizaron con el programa Origin (versión 8, OriginLab Corp.).

Con respecto a otros glicanos, se obtuvieron tres *drift times* diferentes para el glicano H6N5S3, un *drift time* para el glicano H5N4F1S2 y un *drift time* para el glicano H5N4S1. En el caso del glicano H5N4F1S2, aunque sería razonable pensar que también existen dos isómeros (por eso se han simulado dos picos gaussianos), realmente sólo se observa un pico. Podría sugerirse que la resolución entre los dos isómeros es significativamente peor, haciendo imposible su distinción. Esto puede ser debido a que la presencia de una fucosa adicional impide que un cambio en el enlace del SiA afecte tanto a la CCS del glicano. En cuanto al glicano H5N4S1, parece que solo existe un isómero ya que se observa un único pico. Aun así, no se puede asegurar completamente ya que es posible que ambos isómeros tengan una CCS igual o muy parecida impidiendo su correcta separación; por lo tanto, la ausencia de un segundo pico no descarta por completo la presencia de más de un isómero.

Para estudiar en más profundidad si el tipo de enlace del SiA puede afectar o no a la forma global del glicano y observar así diferencias entre isómeros por movilidad iónica, la Figura 5.23 muestra de manera esquemática las estructuras 3D obtenidas en la aplicación online de *Molecular Networks GmbH and Altamira, LLC.* [307] para el fragmento H1N1S1, obtenido por MS/MS de la glicoforma H5N4S2 del glicopéptido, y para el glicano H5N4S2. Las estructuras mostradas son las más estables (menor energía) para cada caso.

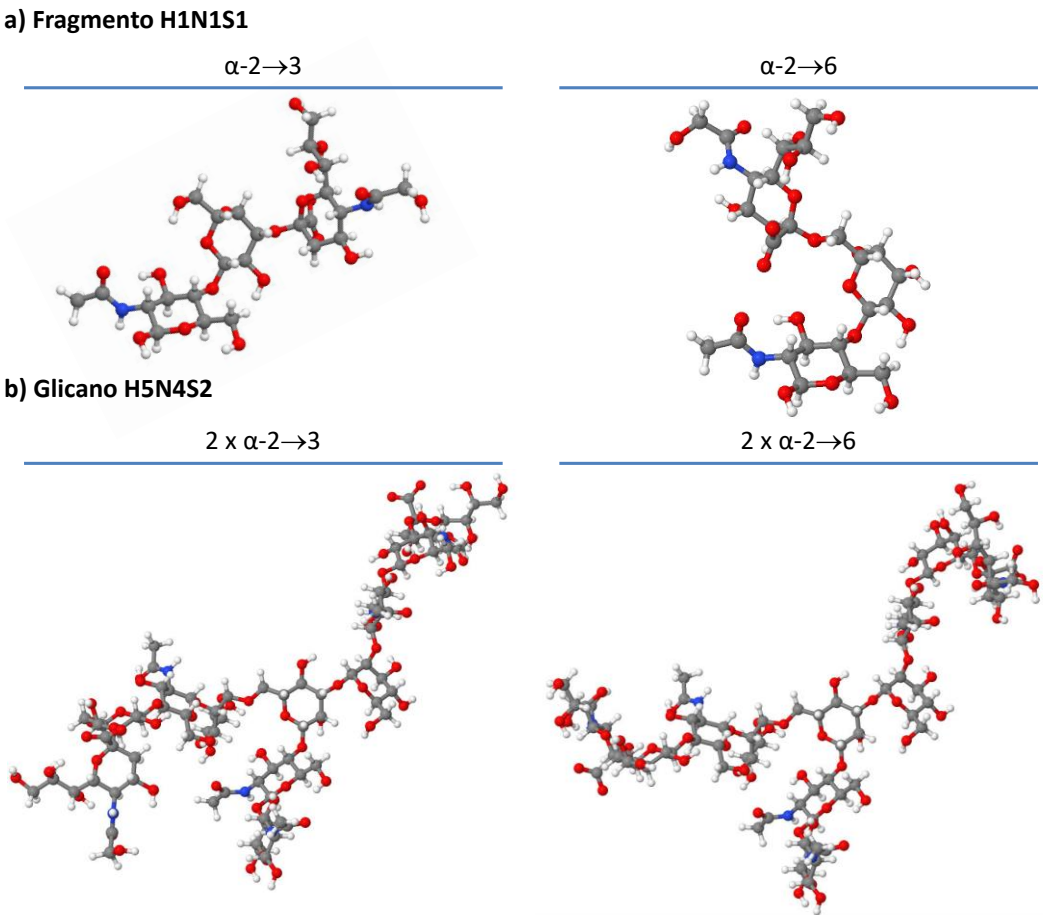


Figura 5.23: Representación 3D del a) fragmento H1N1S1 del glicopéptido N(H5N4S2)STLCDLCIGPLK y b) del glicano H5N4S2, con diferentes enlaces del ácido siálico, utilizando la aplicación online de *Molecular Networks GmbH and Altamira, LLC* [307].

En el caso del fragmento H1N1S1, cuando el enlace del SiA es α-2→6 la estructura parece ser más compacta que con el enlace α-2→3 y por este motivo la CCS del isómero α-2→3 quizás sea mayor. En cambio, en el caso del glicano H5N4S2, al comparar la forma de los isómeros 2 x α-2→3 y 2 x α-2→6, se puede observar que la estructura es muy parecida, siendo difícil extraer conclusiones. Quizás podría parecer que el glicano con dos enlaces α-2→6 tiene una estructura ligeramente más compacta que el glicano con los dos enlaces α-2→3 lo que daría lugar a valores de CCS menores. Dado que únicamente con la visualización de la forma de las moléculas era muy complicado extraer conclusiones fiables, se decidió aplicar modelos de cálculo teórico de CCS utilizando el software **IMoS** [308, 309], que se encuentra disponible gratuitamente en *imospedia*.com, para el fragmento H1N1S1 y el glicano H5N4S2. En la Tabla

5.12 se muestran los valores de CCS obtenidos. Cabe destacar que, para que los cálculos de las CCS fueran los más exactos posibles, se utilizaron las estructuras más estables, las de mínima energía, tanto para el fragmento H1N1S1 como para el glicano H5N4S2, obtenidas a partir de la aplicación ***carbohydrate builder*** disponible gratuitamente en *glycam.org* [310].

Tabla 5.12: CCS teóricas obtenidas con el software IMoS para los diferentes isómeros debidos al ácido siálico del fragmento H1N1S1 y el glicano H5N4S2.

		Isómero	CCS (Å^2)
Fragmento	H1N1S1	α-2→3	243.7
	H1N1S1	α-2→6	233.4
Glicano	H5N4S2	2 x α-2→3	652.3
	H5N4S2	1 x α-2→3 1 x α-2→6	635.4
	H5N4S2	2 x α-2→6	623.9

En primer lugar, los resultados obtenidos para el fragmento H1N1S1 concuerdan con los obtenidos por otros autores [143, 306], donde el primer pico, aquel con un *drift time* menor, y por lo tanto una CCS menor, correspondería al SiA α-2→6 y el de mayor *drift time* y mayor CCS correspondería al SiA α-2→3. Además, si observamos la intensidad relativa entre los dos picos (Figura 5.21-c), se puede ver como parece ser bastante más intenso el SiA α-2→6, lo cual también coincide con los resultados obtenidos por ciertos autores [306], que sugieren que, en general, las glicoformas biantenarias tienen un contenido de SiA α-2→3 menor.

En el caso del glicano H5N4S2, las diferencias de CCS observadas entre los diferentes isómeros son suficientemente grandes como para que fuese posible separarlos por IM. Teniendo en cuenta que anteriormente se observó que la abundancia del isómero α-2→6 era mayor que la del α-2→3 (en el estudio del fragmento H1N1S1), y que el isómero con dos enlaces α-2→3 sería menos probable según la literatura [306], podríamos asignar el primer pico (el menos abundante) de la Figura 5.22-a al glicano con dos enlaces α-2→6, y el segundo pico al glicano con un enlace α-2→6 y otro α-2→3. Sin embargo, esta asignación es tentativa, y solo sería

completamente fiable si se complementase con algún estudio más exhaustivo de MS/MS o donde se utilicen sialidasas específicas [144]. En conclusión, la metodología desarrollada en esta tesis doctoral permite separar isómeros debidos al enlace del SiA a nivel glicopeptídico y de glicano, aunque su asignación debería ser confirmada con estudios complementarios. No obstante, en el caso de los glicopéptidos, es necesario separarlos previamente de los péptidos del digesto y fragmentarlos por MS/MS para poder obtener información sobre su isomería, siendo esta vía mucho más laboriosa. Por ello, el análisis de los glicanos se puede considerar como la opción más adecuada para el estudio de los isómeros debidos al enlace del SiA.

Para comprobar que la metodología propuesta se puede aplicar a otras glicoproteínas, se analizaron los glicanos obtenidos de la hAGP. En general, se obtuvieron resultados muy similares a los obtenidos para la mTf: dos *drift times* diferentes para el glicano H5N4S2 y tres *drift times* para el glicano H6N5S3, aunque la resolución que se obtuvo fue ligeramente peor en esta glicoproteína. Aun así, los resultados obtenidos permiten concluir que la metodología establecida puede ser utilizada para el estudio de otras glicoproteínas.

5.3.3. Análisis de muestras de suero

Finalmente, para comprobar la robustez del método al analizar muestras más complejas, se analizaron dos muestras de suero de ratón: WT control y WT con CIA, donde la mTf había sido previamente purificada por IAC, como en un estudio anterior (artículo 2.3). A modo de ejemplo, la Figura 5.24 muestra los ATD de los glicanos H5N4S2 y H6N5S3 en las dos muestras de suero (en esta figura también se realizó una reconstrucción gaussiana de los diferentes picos). Si se comparan estos resultados con la Figura 5.22, se puede observar que se obtienen los mismos resultados que con el patrón de mTf para ambos glicanos (dos *drift times* para el glicano H5N4S2 y tres para el glicano H6N5S3).

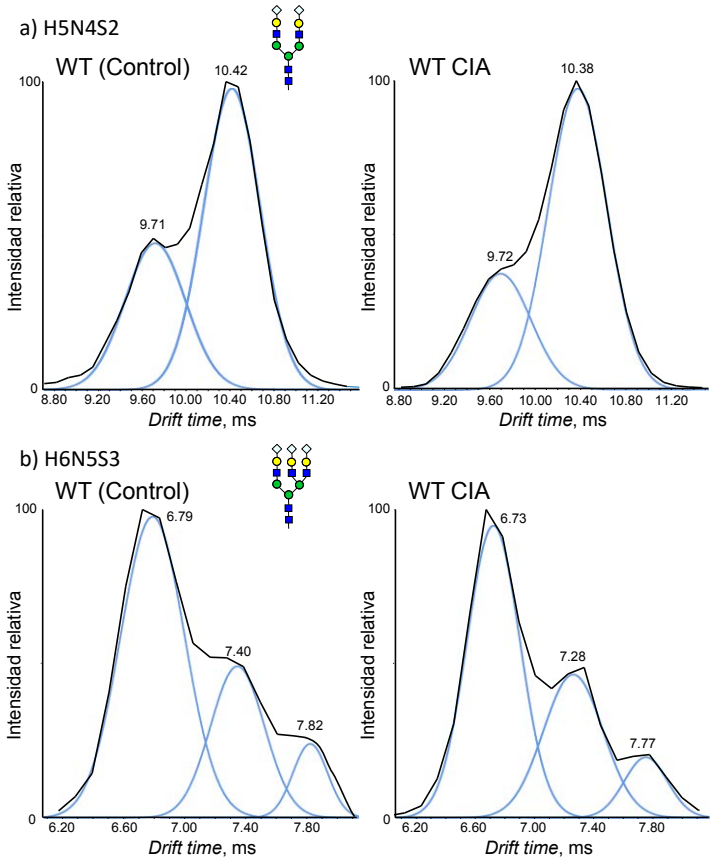


Figura 5.24: ATD de diferentes glicanos de la mTf, obtenidos por nanoESI-IM-MS, en una muestra de suero de ratón WT control y de ratón WT con CIA: a) H5N4S2 y b) H6N5S3. Los símbolos utilizados para la representación de los glicanos siguen las reglas sugeridas por el *Consortium for Functional Glycomics* (CFG) [13]. Los picos gaussianos simulados se realizaron con el programa Origin (versión 8, OriginLab Corp.).

Aunque existe una ligera diferencia en cuanto a la intensidad relativa de los diferentes isómeros entre ambas muestras de suero, sería necesario un número superior de muestras para poder llegar a discernir si CIA afecta o no a la abundancia relativa de los isómeros. Aun así, los resultados obtenidos demuestran que la metodología propuesta es aplicable al análisis de la glicosilación y al estudio de los isómeros de glicanos de glicoproteínas en muestras de suero.

En conclusión, con la instrumentación disponible y con su potencial actual, IM se puede considerar una técnica importante en la separación de compuestos isoméricos y, complementada con otras técnicas, es posible llegar a obtener información fiable sobre la

comparación de isómeros en glicoconjugados. En concreto, IM permite separar glicanos isoméricos que difieran en el tipo de enlace del SiA. En cuanto a glicopéptidos y glicoproteína intacta, la capacidad actual de esta técnica es insuficiente para poder separar los diferentes isómeros debidos a un cambio en el enlace del SiA. Aun así, existe la alternativa de fragmentar los glicopéptidos y analizar por IM fragmentos menores, como glicanos, de tal manera que una modificación del enlace del SiA sí tenga efecto en la separación. Aunque, en realidad, con esta alternativa se mide el *drift time* de un fragmento y no directamente del glicopéptido. De todos modos, el interés de la comunidad científica en IM ha crecido considerablemente en los últimos años, lo cual, seguramente, llevará a nuevos avances tecnológicos que mejorarán la capacidad de los equipos de IM. Es posible que en un futuro, con una nueva generación de instrumentos, sea posible, no sólo mejorar la separación entre glicanos isoméricos, sino que incluso se pueda llegar a observar separación de isómeros debidos al tipo de enlace del SiA analizando glicopéptidos o glicoproteínas intactas.

Chapter 6. Conclusions

The following general conclusions can be extracted from the work carried out and explained in this thesis:

- A **CapLC-MS** methodology has been developed for the **glycosylation** characterization of glycopeptides obtained from the proteolytic digestion of a glycoprotein.
    - The use of an anionic **surfactant** (RapiGest®) was necessary to enhance the enzymatic activity and, consequently, efficiently digest **proteolytic resistant glycoproteins**, such as **human transferrin (hTf)**.
        - The conditions for the surfactant hydrolysis recommended by the manufacturer were optimized to minimize sialic acid loss.
        - A concentration of 5% (v/v) of HFor and reaction times of 30 min were selected as the optimum conditions for the surfactant hydrolysis.
- hTf glycopeptide glycoforms have been successfully characterized by CapLC-MS. The effect of chronic alcoholism and congenital disorders of glycosylation (CDG) in the relative abundance of these glycoforms has been assessed:
    - **Chronic alcoholism:**
    - The non-glycosylated peptide variant corresponding to each glycopeptide, that is, only the amino acid chain without any glycan (referred to as peptide glycoform), was observed to be increased in chronic alcoholism, whereas other major glycoforms, mainly the H5N4S2 for both glycopeptides, were reduced. Additionally, other glycoforms, such as the H5N4S1, decreased just slightly or were barely affected.
    - The obtained results suggest that the mechanism causing the abnormality in the glycosylation of hTf in heavy alcohol consumption impedes the initiation of N-linked glycosylation in the pre-Golgi step. Hence, the effect of ethanol in the second step of N-glycosylation is minimal.

- **Congenital disorders of glycosylation:**

- **PCA** and **PLS-DA** were correctly applied as potent exploratory and classificatory tools to properly study the huge data set originated in this study.
- Only six out of the fourteen detected glycopeptide glycoforms were observed to be enough to accurately classify the samples in three groups: control, CDG-I and CDG-II. This small glycoform set could be regarded as a novel biomarker panel for CDG.
- Unexpectedly, peptide glycoforms were not included in this biomarker panel. Even though their relative abundance was increased in CDG-I samples, the high heterogeneity of the analyzed samples might be the cause why peptide glycoforms are not part of this biomarker panel.
- The most glycosylated glycoforms (H5N4S2 and H6N5S3) were essential for the differentiation of healthy and CDG samples.
- Some important less sialylated glycoforms (H5N4S1 and H5N4) were present in the biomarker panel, as they were increased in CDG-II samples.
- The alteration of hTf glycosylation seems to be more pronounced in CDG-I than in chronic alcoholism; at least, this is the case with the number of samples available in each study.

- **Mouse transferrin (mTf)** was also observed to be altered in mice with **collagen induced arthritis (CIA)**, an autoimmune inflammatory disease homologous to human rheumatoid arthritis (RA) in some aspects:
  - Increased fucosylated and branched glycoforms were observed, which was in agreement with the reported alteration in other inflammatory diseases. A decrease in less branched glycoforms was also observed.

  - The observed alteration in mTf glycosylation profile was different between WT and CD38-KO mice, demonstrating that a defect in CD38 gene can have an effect in the development of CIA.

- The developed CapLC-MS methodology can be regarded as an excellent tool not only for the characterization of the glycosylation at the glycopeptide level, but also to study the effect of a certain disease in the glycosylation of a protein. Even it could be considered an alluring option to assess the contribution of a given gene or protein to the development of certain diseases. Additionally, it could be useful to ensure the quality of a biopharmaceutical due to the characterization of its glycosylation.
- Complete elimination of the surfactant required to properly digest proteolytic resistant glycoproteins, such as hTf, was achieved, which permitted the use of CE-MS and led to the study of the glycopeptide glycosylation of hTf by this technique, which was otherwise impossible:
  - Using CE as separation technique prior to MS provided several benefits, for instance, all glycopeptide glycoforms were conveniently separated in a single run.
  - The developed CE-MS methodology could be regarded as a valid complementary approach to CapLC-MS for the glycosylation characterization of proteins, each with its own benefits and drawbacks.
- Additionally, another benefit of using CE is that, theoretically, the migration behavior of peptides and glycopeptides originated from the tryptic digest of a glycoprotein can be predicted and modelled:
  - This theory was successfully corroborated with **recombinant human erythropoietin (rhEPO)**, as the migration behavior of both, its peptides and glycopeptides, was predicted.

- Peptides were observed to follow the **Stoke's law** ($q/M^{1/3}$), whereas glycopeptides fitted better with the **classical polymer model** ($q/M^{1/2}$). Moreover, peptides and glycopeptides predicted electropherograms perfectly matched the experimental electropherograms.
- The migration behavior of other glycopeptide glycoforms of rhEPO, or even other glycoproteins, was also correctly predicted and theoretical electropherograms and experimental electropherograms were equivalent.

- **IM-MS** has been demonstrated to be an enticing and powerful technique for the separation of isomeric glycoconjugates, at least, regarding those isomers owing to the type of sialic acid linkage:
  - No differentiation between isomers was observed in IM-MS when analyzing mTf at the intact glycoprotein level or at the glycopeptide level.
  - Sialic acid linkage type was successfully studied at the glycopeptide level only when the drift time of a glycopeptide fragment that still retained the sialic acid was measured. Specifically, two different drift times were observed for the glycan H1N1S1, a relative intense fragment of the most abundant glycopeptide of mTf.
  - Straightforward separation was possible with the analysis of the free glycans from mTf.
  - Two and three different drift times were observed for the H5N4S2 and the H6N5S3 glycans, respectively.
  - The aforementioned distinction was not only observed in standard samples but also in serum samples from control mice and mice with CIA, which demonstrates the robustness of the developed method.

- It can be safe to assume that, in the near future, IM-MS technology will improve considerably to be able to enhance isomeric separation between glycans or, even, distinguish the type of sialic acid linkage at the glycopeptide level.

Chapter 7. References

[1] E. Gubb, R. Matthiesen, Chapter 1. Introduction to omics, in: R. Matthiesen (Ed.), Bioinformatics Methods in Clinical Research, Humana Press, 2010, pp. 1–23.

[2] R.P. Horgan, L.C. Kenny, " Omic " technologies: proteomics and metabolomics, Obstet. Gynaecol. 13 (2011) 189–195.

[3] J.C. Venter, M.D. Adams, E.W. Myers, P.W. Li, R.J. Mural, G.G. Sutton, H.O. Smith, et al., The sequence of the human genome, Science 291 (2001) 1304–1351.

[4] C.R. Bertozzi, R. Sasisekharan, Chapter 48. Glycomics, 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 of Glycobiology, Cold Spring Harbor Laboratory Press, 2009.

[5] 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.

[6] M.J. Kailemia, D. Park, C.B. Lebrilla, Glycans and glycoproteins as specific biomarkers for cancer, Anal. Bioanal. Chem. 409 (2016) 395–410.

[7] Y. Zhang, J. Jiao, P. Yang, H. Lu, Mass spectrometry-based N-glycoproteomics for cancer biomarker discovery, Clin. Proteomics 11 (2014) 18.

[8] R.C. Hughes, Glycoproteins, Chapman and Hall, 1983.

[9] R.A. Dwek, Glycobiology: towards understanding the function of sugars, Chem. Rev. 96 (1996) 683–720.

[10] H.J. Allen, E.C. Kisailus, Glycoconjugates: Composition, Structure, and Function, CRC Press, 1992.

[11] 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 of Glycobiology, Cold Spring Harbor Laboratory Press, 2009.

[12] C. Bertozzi, D. Rabuka, Chapter 2. Structural Basis of Glycan Diversity, 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 of Glycobiology, Cold Spring Harbor Laboratory Press, 2009.

[13] Consortium for Functional Glycomics. Functional Glycomics Gateway, http://www.functionalglycomics.org/.

[14] 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.), Current Protocols in Protein Science, Wiley, 2009.

[15] R.G. Spiro, Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds, Glycobiology 12 (2002) 43R–56R.

[16] S.B. Levery, C. Steentoft, A. Halim, Y. Narimatsu, H. Clausen, S.Y. Vakhrushev, Advances in mass spectrometry driven O-glycoproteomics, Biochim. Biophys. Acta - Gen. Subj. 1850 (2015) 33–42.

[17] P. Stanley, H. Schachter, N. Taniguchi, Chapter 8. N-Glycans, 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 of Glycobiology, Cold Spring Harbor Laboratory Press, 2009.

[18] 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 of Glycobiology, Cold Spring Harbor Laboratory Press, 2009.

[19] H.J. Gabius, S. Gabius, Eds., Glycosciences. Status and perspectives., Glycosciences. Status and perspectives., WILEY-VCH, 1997.

[20] 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.

[21] M. Devasahayam, Factors affecting the expression of recombinant glycoproteins, Indian J. Med. Res. 126 (2007) 22–27.

[22] M.J. Betenbaugh, N. Tomiya, S. Narang, Chapter 4.32. Glycoengineering: Recombinant Glycoproteins, in: J.P. Kamerling (Ed.), Comprehensive Glycoscience, Elsevier B.V., 2007, pp. 607–642.

[23] N. Sethuraman, T.A. Stadheim, Challenges in therapeutic glycoprotein production, Curr. Opin. Biotechnol. 17 (2006) 341–346.

[24] M.W. Spellman, Carbohydrate Characterization of Recombinant Glycoproteins of Pharmaceutical Interest, Anal. Chem 62 (1990) 1714–1722.

[25] O. Gornik, J. Dumic, M. Flogel, G. Lauc, Glycoscience - a new frontier in rational drug design, Acta Pharm 56 (2006) 19–30.

[26] R.T. MacGillivray, E. Mendez, S.K. Sinha, M.R. Sutton, J. Lineback-Zins, K. Brew, The complete amino acid sequence of human serum transferrin, Proc. Natl. Acad. Sci. U. S. A. 79 (1982) 2504–2508.

[27] M. Ridinger, P. Köhl, E. Gäbele, N. Wodarz, G. Schmitz, P. Kiefer, C. Hellerbrand, Analysis of carbohydrate deficient transferrin serum levels during abstinence, Exp. Mol. Pathol. 92 (2012) 50–53.

[28] W. Oberrauch, A.C. Bergman, A. Helander, HPLC and mass spectrometric characterization of a candidate reference material for the alcohol biomarker carbohydrate-deficient transferrin (CDT), Clin. Chim. Acta 395 (2008) 142–145.

[29] F. Bortolotti, G. De Paoli, F. Tagliaro, Carbohydrate-deficient transferrin (CDT) as a marker of alcohol abuse: A critical review of the literature 2001-2005, J. Chromatogr. B Anal. Technol. Biomed. Life Sci. 841 (2006) 96–109.

[30] O. Niemelä, Biomarker-based approaches for assessing alcohol use disorders, Int. J. Environ. Res. Public Health 13 (2016) 166.

[31] N. Waszkiewicz, S.D. Szajda, A. Zalewska, A. Szulc, A. Kepka, A. Minarowska, M. Wojewódzka-Zelezniakowicz, B. Konarzewska, S. Chojnowska, J.R. Ładny, K. Zwierz, Alcohol abuse and glycoconjugate metabolism, Folia Histochem. Cytobiol. 50 (2012) 1–11.

[32] M. Binkhorst, S.B. Wortmann, S. Funke, T. Kozicz, R.A. Wevers, E. Morava, Glycosylation defects underlying fetal alcohol spectrum disorder: A novel pathogenetic model, J. Inherit. Metab. Dis. 35 (2012) 399–405.

[33] C.A. Casey, G. Bhat, M.S. Holzapfel, A. Petrosyan, Study of Ethanol-Induced Golgi Disorganization Reveals the Potential Mechanism of Alcohol-Impaired N-Glycosylation, Alcohol. Clin. Exp. Res. 40 (2016) 2573–2590.

[34] J. Jaeken, Chapter 179. Congenital disorders of glycosylation, in: E.F.M. Wijdicks, A.H. Kramer (Eds.), Handbook of Clinical Neurology, Elsevier B.V., 2013, pp. 1737–1743.

[35] R. Barone, A. Fiumara, J. Jaeken, Congenital disorders of glycosylation with emphasis on cerebellar involvement, Semin. Neurol. 34 (2014) 357–366.

[36] L.A. Wolfe, D. Krasnewich, Congenital disorders of glycosylation and intellectual disability, Dev. Disabil. Res. Rev. 17 (2013) 211–225.

[37] Y. Wada, Mass spectrometry of transferrin and apolipoprotein C-III for diagnosis and screening of congenital disorder of glycosylation, Glycoconj. J. 33 (2016) 297–307.

[38] C. McCarthy, R. Saldova, M.R. Wormald, P.M. Rudd, N.G. McElvaney, E.P. Reeves, The role and importance of glycosylation of acute phase proteins with focus on alpha-1 antitrypsin in acute and chronic inflammatory conditions, J Proteome Res 13 (2014) 3131–3143.

[39] 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.

[40] M.A. Connelly, E.G. Gruppen, J.D. Otvos, R.P.F. Dullaart, Inflammatory glycoproteins in cardiometabolic disorders, autoimmune diseases and cancer, Clin. Chim. Acta 459 (2016) 177–186.

[41] R.A. Feelders, G. Vreugdenhil, G. de Jong, A.J.G. Swaak, H.G. van Eijk, Transferrin microheterogeneity in rheumatoid arthritis. Relation with disease activity and anemia of chronic disease, Rheumatol. Int. 12 (1992) 195–199.

[42] J. Jaeken, H. Stibler, B. Hagberg, The carbohydrate-deficient glycoprotein syndrome. A new inherited multisystemic disease with severe nervous system involvement, Acta Paediatr Scand Suppl 375 (1991) 1–71.

[43] O. Niemelä, Biomarkers in alcoholism, Clin. Chim. Acta 377 (2007) 39–49.

[44] P. Sillanaukee, N. Strid, J.P. Allen, R.Z. Litten, Possible reasons why heavy drinking increases carbohydrate-deficient transferrin, Alcohol. Clin. Exp. Res. 25 (2001) 34–40.

[45] C. Ji, Advances and New Concepts in Alcohol-Induced Organelle Stress, Unfolded Protein Responses and Organ Damage, Biomolecules 5 (2015) 1099–1121.

[46] M. Gong, M. Garige, K. Hirsch, M.R. Lakshman, Liver Galβ1,4GlcNAc α2,6-sialyltransferase is down-regulated in human alcoholics: possible cause for the appearance of asialoconjugates, Metabolism 56 (2007) 1241–1247.

[47] M. Welti, A.J. Hülsmeier, Ethanol-induced impairment in the biosynthesis of N-linked glycosylation, J. Cell. Biochem. 115 (2014) 754–762.

[48] J.J. Galligan, K.S. Fritz, H. Tipney, R.L. Smathers, J.R. Roede, C.T. Shearn, L.E. Hunter, D.R. Petersen, Profiling impaired hepatic endoplasmic reticulum glycosylation as a consequence of ethanol ingestion, J. Proteome Res. 10 (2011) 1837–1847.

[49] K. Scott, T. Gadomski, T. Kozicz, E. Morava, Congenital disorders of glycosylation: New defects and still counting, J. Inherit. Metab. Dis. 37 (2014) 609–617.

[50] M. Monticelli, T. Ferro, J. Jaeken, V. dos Reis Ferreira, P.A. Videira, Immunological aspects of congenital disorders of glycosylation (CDG): a review, J. Inherit. Metab. Dis. 39 (2016) 765–780.

[51] D. Rymen, J. Jaeken, Skin manifestations in CDG, J. Inherit. Metab. Dis. 37 (2014) 699–708.

[52] J. Jaeken, G. Matthijs, R. Barone, H. Carchon, Carbohydrate deficient glycoprotein (CDG) syndrome type I, J. Med. Genet. 34 (1997) 73–76.

[53] M. Aebi, A. Helenius, B. Schenk, R. Barone, A. Fiumara, E.G. Berger, T. Hennet, T. Imbach, et al., Carbohydrate-deficient glycoprotein syndromes become congenital disorders of glycosylation : An updated nomenclature for CDG, Glycoconj. J. 16 (1999) 669–671.

[54] D. Krasnewich, Human glycosylation disorders, Cancer Biomarkers 14 (2014) 3–16.

[55] M. Guillard, Y. Wada, H. Hansikova, I. Yuasa, K. Vesela, N. Ondruskova, M. Kadoya, A.

Janssen, L.P.W.J. Van den Heuvel, E. Morava, J. Zeman, R.A. Wevers, D.J. Lefeber, Transferrin mutations at the glycosylation site complicate diagnosis of congenital disorders of glycosylation type I, J. Inherit. Metab. Dis. 34 (2011) 901–906.

[56] B. Xia, W. Zhang, X. Li, R. Jiang, T. Harper, R. Liu, R.D. Cummings, M. He, Serum N-glycan and O-glycan analysis by mass spectrometry for diagnosis of congenital disorders of glycosylation, Anal. Biochem. 442 (2013) 178–185.

[57] J. Jaeken, T. Hennet, G. Matthijs, H.H. Freeze, CDG nomenclature: time for a change!, Biochim. Biophys. Acta 1792 (2009) 825–826.

[58] J. Jaeken, Congenital disorders of glycosylation, Ann. N. Y. Acad. Sci. 1214 (2010) 190–8.

[59] Genetics Home Reference. U.S. National Library of Medicine, ghr.nlm.nih.gov.

[60] Genetic and Rare Diseases Information Center. National Center for Advancing Translational Sciences, https://rarediseases.info.nih.gov/.

[61] A. Helander, J.P. Bergström, Determination of carbohydrate-deficient transferrin in human serum using the Bio-Rad %CDT by HPLC test, Clin. Chim. Acta 371 (2006) 187–190.

[62] A. Helander, J.P.M. Wielders, R. te Stroet, J.P. Bergström, Comparison of HPLC and Capillary Electrophoresis for Confirmatory Testing of the Alcohol Misuse Marker Carbohydrate-Deficient Transferrin, Clin. Chem. 51 (2005) 1528–1531.

[63] J.O. Jeppsson, T. Arndt, F. Schellenberg, J.P.M. Wielders, R.F. Anton, J.B. Whitfield, A. Helander, Toward standardization of carbohydrate-deficient transferrin (CDT) measurements: I. Analyte definition and proposal of a candidate reference method, Clin. Chem. Lab. Med. 45 (2007) 558–562.

[64] J.R. Delanghe, M.L. De Buyzere, Carbohydrate deficient transferrin and forensic medicine, Clin. Chim. Acta 406 (2009) 1–7.

[65] F. Parente, N. Ah Mew, J. Jaeken, B.M. Gilfix, A new Capillary Zone Electrophoresis method for the screening of Congenital Disorders of Glycosylation (CDG), Clin. Chim. Acta 411 (2010) 64–66.

[66] E. Quintana, R. Montero, M. Casado, A. Navarro-Sastre, M.A. Vilaseca, P. Briones, R. Artuch, Comparison between high performance liquid chromatography and capillary zone electrophoresis for the diagnosis of congenital disorders of glycosylation, J. Chromatogr. B Anal. Technol. Biomed. Life Sci. 877 (2009) 2513–2518.

[67] R. Medzhitov, Origin and physiological roles of inflammation, Nature 454 (2008) 428–435.

[68] O. Gornik, G. Lauc, Glycosylation of serum proteins in inflammatory diseases, Dis. Markers 25 (2008) 267–278.

[69] C. Váradi, S. Mittermayr, Á. Szekrényes, J. Kádas, L. Takacs, I. Kurucz, A. Guttman, Analysis of haptoglobin N-glycome alterations in inflammatory and malignant lung diseases by capillary electrophoresis, Electrophoresis 34 (2013) 2287–2294.

[70] Y. Rombouts, H.S. Jónasdóttir, A.L. Hipgrave Ederveen, K.R. Reiding, B.C. Jansen, J. Freysdottir, I. Hardardottir, A. Ioan-Facsinay, M. Giera, M. Wuhrer, Acute phase inflammation is characterized by rapid changes in plasma/peritoneal fluid N-glycosylation in mice, Glycoconj. J. 33 (2016) 457–470.

[71] S. Albrecht, L. Unwin, M. Muniyappa, P.M. Rudd, Glycosylation as a marker for inflammatory arthritis, Cancer Biomarkers 14 (2014) 17–28.

[72] N. Rosenthal, S. Brown, The mouse ascending: perspectives for human-disease models, Nat. Cell Biol. 9 (2007) 993–999.

[73] J.R. Whiteaker, C. Lin, J. Kennedy, L. Hou, M. Trute, I. Sokal, P. Yan, R.M. Schoenherr, L. Zhao, U.J. Voytovich, K.S. Kelly-Spratt, A. Krasnoselsky, P.R. Gafken, J.M. Hogan, L.A. Jones, P. Wang, L. Amon, L.A. Chodosh, P.S. Nelson, M.W. McIntosh, C.J. Kemp, A.G. Paulovich, A targeted proteomics–based pipeline for verification of biomarkers in plasma, Nat Biotechnol 29 (2011) 625–634.

[74] C. Von Toerne, M. Kahle, A. Schäfer, R. Ispiryan, M. Blindert, M. Hrabe De Angelis, S. Neschen, M. Ueffing, S.M. Hauck, Apoe, Mbl2, and Psp plasma protein levels correlate with diabetic phenotype in NZO mice - An optimized rapid workflow for SRM-based quantification,

J. Proteome Res. 12 (2013) 1331–1343.

[75] 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) R113.

[76] C.A. Notley, J.J. Inglis, S. Alzabin, F.E. McCann, K.E. McNamee, R.O. Williams, Blockade of tumor necrosis factor in collagen-induced arthritis reveals a novel immunoregulatory pathway for Th1 and Th17 cells, J. Exp. Med. 205 (2008) 2491–2497.

[77] D.H. Dube, C.R. Bertozzi, Glycans in cancer and inflammation-potential for therapeutics and diagnostics, Nat. Rev. Drug Discov. 4 (2005) 477–488.

[78] F.K. Lin, S. Suggs, C.H. Lin, J.K. Browne, R. Smalling, J.C. Egrie, K.K. Chen, G.M. Fox, F. Martin, Z. Stabinsky, Cloning and expression of the human erythropoietin gene, Proc. Natl. Acad. Sci. U. S. A. 82 (1985) 7580–7584.

[79] J.D. Cody, E.M. 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).

[80] 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) 5.

[81] 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.

[82] 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.

[83] E. Balaguer, U. Demelbauer, M. Pelzing, V. Sanz-Nebot, J. Barbosa, C. Neusüß, Glycoform characterization of erythropoietin combining glycan and intact protein analysis by capillary

electrophoresis – electrospray – time-of-flight mass spectrometry, Electrophoresis 27 (2006) 2638–2650.

[84] 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.

[85] 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.

[86] J. Hirabayashi, Chapter 20. Lectin-Based Glycomics: How and When Was the Technology Born?, in: J. Hirabayashi (Ed.), Lectins. Methods and Protocols, Springer New York, 2014, pp. 225–242.

[87] 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.

[88] Š. Belický, J. Katrlík, J. Tkáč, Glycan and lectin biosensors, Essays Biochem. 60 (2016) 37–47.

[89] Z. Yang, L.E. Harris, D.E. Palmer-Toy, W.S. Hancock, Multilectin affinity chromatography for characterization of multiple glycoprotein biomarker candidates in serum from breast cancer patients, Clin. Chem. 52 (2006) 1897–1905.

[90] S. Fanayan, M. Hincapie, W.S. Hancock, Using lectins to harvest the plasma/serum glycoproteome, Electrophoresis 33 (2012) 1746–1754.

[91] M. Madera, Y. Mechref, M. V. Novotny, Combining lectin microcolumns with high-resolution separation techniques for enrichment of glycoproteins and glycopeptides, Anal. Chem. 77 (2005) 4081–4090.

[92] 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.

[93] K. Engholm-Keller, M.R. Larsen, Titanium dioxide as chemo-affinity chromatographic sorbent of biomolecular compounds - Applications in acidic modification-specific proteomics, J. Proteomics 75 (2011) 317–328.

[94] 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 TiO2 chromatography, Glycoconj. J. 29 (2012) 433–443.

[95] M.R. Larsen, S.S. Jensen, L.A. Jakobsen, N.H.H. Heegaard, Exploring the sialiome using titanium dioxide chromatography and mass spectrometry, Mol. Cell. proteomics 6 (2007) 1778–1787.

[96] A.D. Roddick-Lanzilotta, A.J. McQuillan, An in situ Infrared Spectroscopic Study of Glutamic Acid and of Aspartic Acid Adsorbed on TiO(2): Implications for the Biocompatibility of Titanium, J. Colloid Interface Sci. 227 (2000) 48–54.

[97] D.A. Abi-Ghanem, L.R. Berghman, Chapter 5. Immunoaffinity chromatography: A review, in: Dr. Sameh Magdeldin (Ed.), Affinity chromatography, InTech, 2012, pp. 91–106.

[98] M. Uhlén, Affinity as a tool in life science, Biotechniques 44 (2008) 649–654.

[99] D.S. Hage, Survey of recent advances in analytical applications of immunoaffinity chromatography, J Chromatogr B Biomed Sci Appl 715 (1998) 3–28.

[100] M. Lesney, Sticking with affinity chromatography, Today's Chem. Work (2003) 38–40.

[101] S. Sarantopoulos, Chapter 3. Antibody-Antigen Interactions, in: E.A. Greenfield (Ed.), Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press, 2014.

[102] J. Fitzgerald, P. Leonard, E. Darcy, R. O'Kennedy, Chapter 3. Immunoaffinity Chromatography, in: D. Walls, S.T. Loughran (Eds.), Protein Chromatography. Methods and Protocols, Humana Press, 2011, pp. 35–59.

[103] M. Wuhrer, Glycomics using mass spectrometry, Glycoconj. J. 30 (2013) 11–22.

[104] A. Banazadeh, L. Veillon, K.M. Wooding, M. Zabet-moghaddam, Y. Mechref, Recent advances in mass spectrometric analysis of glycoproteins, Electrophoresis 38 (2017) 162–189.

[105] M. Thaysen-Andersen, N.H. Packer, Advances in LC-MS/MS-based glycoproteomics: Getting closer to system-wide site-specific mapping of the N- and O-glycoproteome, Biochim. Biophys. Acta - Proteins Proteomics 1844 (2014) 1437–1452.

[106] C. Junot, F. Fenaille, B. Colsch, F. Bécher, High resolution mass spectrometry based techniques at the crossroads of metabolic pathways, Mass Spectrom. Rev. 33 (2014) 471–500.

[107] W. Morelle, K. Canis, F. Chirat, V. Faid, J.C. Michalski, The use of mass spectrometry for the proteomic analysis of glycosylation, Proteomics 6 (2006) 3993–4015.

[108] H. Geyer, R. Geyer, Strategies for analysis of glycoprotein glycosylation, Biochim. Biophys. Acta - Proteins Proteomics 1764 (2006) 1853–1869.

[109] 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.

[110] 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.

[111] 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.

[112] Y. Zhang, A.R. Sinaiko, G.L. Nelsestuen, Chapter 10. Glycoproteins and Glycosylation: Apolipoprotein C3 Glycoforms by Top-Down MALDI-TOF Mass Spectrometry, in: D. Josic, D.C. Hixson (Eds.), Liver Proteomics. Methods and Protocols, Humana Press, 2012, pp. 141–150.

[113] J. Nilsson, Liquid chromatography-tandem mass spectrometry-based fragmentation analysis of glycopeptides, Glycoconj. J. 33 (2016) 261–272.

[114] 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.

[115] D.J. Harvey, Derivatization of carbohydrates for analysis by chromatography; electrophoresis and mass spectrometry, J. Chromatogr. B Anal. Technol. Biomed. Life Sci. 879 (2011) 1196–1225.

[116] M.J. Kailemia, L.R. Ruhaak, C.B. Lebrilla, I.J. Amster, Oligosaccharide Analysis By Mass Spectrometry: A Review Of Recent Developments, Anal. Chem. 86 (2014) 196–212.

[117] E.J. Finehout, K.H. Lee, An introduction to mass spectrometry applications in biological research, Biochem. Mol. Biol. Educ. 32 (2004) 93–100.

[118] E. de Hoffmann, V. Stroobant, Chapter 1. Ion Sources, in: E. de Hoffmann, V. Stroobant (Eds.), Mass Spectrometry. Principles and Applications, Wiley, 2007, pp. 15–84.

[119] E. de Hoffmann, V. Stroobant, Chapter 2. Mass Analysers, in: E. de Hoffmann, V. Stroobant (Eds.), Mass Spectrometry. Principles and Applications, Wiley, 2007, pp. 85–174.

[120] E. de Hoffmann, V. Stroobant, Chapter 3. Detectors and Computers, in: E. de Hoffmann, V. Stroobant (Eds.), Mass Spectrometry. Principles and Applications, Wiley, 2007, pp. 175–188.

[121] R.D. Smith, J.A. Loo, C.G. Edmonds, C.J. Barinaga, H.R. Udseth, New developments in biochemical mass spectrometry: electrospray ionization, Anal. Chem. 62 (1990) 882–899.

[122] M. Dole, L.L. Mack, R.L. Hines, R.C. Mobley, L.D. Ferguson, M.B. Alice, Molecular Beams of Macroions, J. Chem. Phys. 49 (1968) 2240–2249.

[123] L.L. Mack, P. Kralik, A. Rheude, M. Dole, Molecular Beams Of Macroions II, J. Chem. Phys. 52 (1970) 4977–4986.

[124] P. Kebarle, U.H. Verkerk, Electrospray: from ions in solution to ions in the gas phase, what we know now, Mass Spectrom. Rev. 28 (2009) 898–917.

[125] M. Wilm, Principles of Electrospray Ionization, Mol. Cell. proteomics 10 (2011) M111.009407.

[126] M.S. Wilm, M. Mann, Electrospray and Taylor-Cone theory, Dole's beam of macromolecules at last?, Int. J. Mass Spectrom. Ion Process. 136 (1994) 167–180.

[127] M. Wilm, M. Mann, Analytical Properties of the Nanoelectrospray Ion Source, Anal. Chem 68 (1996) 1–8.

[128] E. van Duijn, Current limitations in native mass spectrometry based structural biology, J. Am. Soc. Mass Spectrom. 21 (2010) 971–978.

[129] K. Lorenzen, E. van Duijn, Native Mass Spectrometry as a Tool in Structural Biology, in: J.E. Coligan, B.M. Dunn, D.W. Speicher, P.T. Wingfield (Eds.), Current Protocols in Protein Science, Wiley, 2010.

[130] R.J. Cotter, Time-of-flight mass spectrometry for the structural analysis of biological molecules, Anal. Chem. 64 (1992) 1027–1039.

[131] A.W.T. Bristow, Accurate mass measurement for the determination of elemental formula - A tutorial, Mass Spectrom. Rev. 25 (2006) 99–111.

[132] J.H.J. Dawson, M. Guilhaus, Orthogonal-acceleration time-of-flight mass spectrometer, Rapid Commun. mass Spectrom. 3 (1989) 155–159.

[133] M. Guilhaus, D. Selby, V. Mlynski, Orthogonal acceleration time-of-flight mass spectrometry, Mass Spectrom. Rev. 19 (2000) 65–107.

[134] J. Coles, M. Guilhaus, Orthogonal acceleration — a new direction for time-of-flight mass spectrometry: Fast, sensitive mass analysis for continuous ion sources, Trends Anal. Chem. 12 (1993) 203–213.

[135] J. Zaia, Mass spectrometry and glycomics, OMICS 14 (2010) 401–418.

[136] M. Wuhrer, M.I. Catalina, A.M. Deelder, C.H. Hokke, Glycoproteomics based on tandem mass spectrometry of glycopeptides, J. Chromatogr. B Anal. Technol. Biomed. Life Sci. 849 (2007) 115–128.

[137] C.M. Woo, A. Felix, L. Zhang, J.E. Elias, C.R. Bertozzi, Isotope-targeted glycoproteomics (IsoTaG) analysis of sialylated N- and O-glycopeptides on an Orbitrap Fusion Tribrid using azido and alkynyl sugars, Anal. Bioanal. Chem. 409 (2017) 579–588.

[138] A. Halim, M.C. Carlsson, C.B. Madsen, S. Brand, S.R. Møller, C.E. Olsen, S.Y. Vakhrushev, J.

Brimnes, P.A. Wurtzen, H. Ipsen, B.L. Petersen, H.H. Wandall, Glycoproteomic Analysis of Seven Major Allergenic Proteins Reveals Novel Post-translational Modifications, Mol. Cell. proteomics 14 (2015) 191–204.

[139] Y. Hu, Y. Mechref, Comparing MALDI-MS, RP-LC-MALDI-MS and RP-LC-ESI-MS glycomic profiles of permethylated N-glycans derived from model glycoproteins and human blood serum, Electrophoresis 33 (2012) 1768–1777.

[140] 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.

[141] P. Both, A.P. Green, C.J. Gray, R. Šardzík, J. Voglmeir, C. Fontana, M. Austeri, M. Rejzek, D. Richardson, R.A. Field, G. Widmalm, S.L. Flitsch, C.E. Eyers, Discrimination of epimeric glycans and glycopeptides using IM-MS and its potential for carbohydrate sequencing, Nat Chem 6 (2014) 65–74.

[142] D.J. Harvey, C.A. Scarff, M. Edgeworth, M. Crispin, C.N. Scanlan, F. Sobott, S. Allman, K. Baruah, L. Pritchard, J.H. Scrivens, Travelling wave ion mobility and negative ion fragmentation for the structural determination of N-linked glycans, Electrophoresis 34 (2013) 2368–2378.

[143] 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.

[144] M. Mancera-Arteu, E. Giménez, J. Barbosa, V. Sanz-Nebot, 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, Anal. Chim. Acta 940 (2016) 92–103.

[145] 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.

[146] 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.

[147] 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.

[148] E. V. Da Costa, A.S.P. Moreira, F.M. Nunes, M.A. Coimbra, D. V. Evtuguin, M.R.M. Domingues, Differentiation of isomeric pentose disaccharides by electrospray ionization tandem mass spectrometry and discriminant analysis, Rapid Commun. Mass Spectrom. 26 (2012) 2897–2904.

[149] 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.

[150] 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.

[151] 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.

[152] D. Isailovic, R.T. Kurulugama, M.D. Plasencia, S.T. Stokes, Z. Kyselova, R. Goldman, Y. Mechref, M. V. Novotny, D.E. Clemmer, Profiling of Human Serum Glycans Associated with Liver Cancer and Cirrhosis by IMS–MS, J. Proteome Res. 7 (2008) 1109–1117.

[153] M. Sarbu, F. Zhu, J. Peter-Katalinic, D.E. Clemmer, A.D. Zamfir, Application of ion mobility tandem mass spectrometry to compositional and structural analysis of glycopeptides extracted from the urine of a patient diagnosed with Schindler disease, Rapid Commun. Mass Spectrom. 29 (2015) 1929–1937.

[154] M.M. Maurer, G.C. Donohoe, S.J. Valentine, Advances in ion mobility-mass spectrometry instrumentation and techniques for characterizing structural heterogeneity, Analyst 140 (2015)

6782–6798.

[155] J. Hofmann, H.S. Hahm, P.H. Seeberger, K. Pagel, Identification of carbohydrate anomers using ion mobility-mass spectrometry, Nature 526 (2015) 241–244.

[156] B.T. Ruotolo, J.L.P. Benesch, A.M. Sandercock, S.-J. Hyung, C. V Robinson, Ion mobility-mass spectrometry analysis of large protein complexes, Nat. Protoc. 3 (2008) 1139–1152.

[157] A. Konijnenberg, J.F. van Dyck, L.L. Kailing, F. Sobott, Extending native mass spectrometry approaches to integral membrane proteins, Biol. Chem. 396 (2015) 991–1002.

[158] M.F. Bush, Z. Hall, K. Giles, J. Hoyes, C. V. Robinson, B.T. Ruotolo, Collision cross sections of proteins and their complexes: A calibration framework and database for gas-phase structural biology, Anal. Chem. 82 (2010) 9557–9565.

[159] F. Lanucara, S.W. Holman, C.J. Gray, C.E. Eyers, The power of ion mobility-mass spectrometry for structural characterization and the study of conformational dynamics, Nat. Chem. 6 (2014) 281–294.

[160] A.B. Kanu, P. Dwivedi, M. Tam, L. Matz, H.H. Jr. Hill, Ion mobility–mass spectrometry, J. mass Spectrom. 43 (2008) 1–22.

[161] 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.

[162] A.A. Shvartsburg, R.D. Smith, Fundamentals of traveling wave ion mobility spectrometry, Anal. Chem. 80 (2008) 9689–9699.

[163] J.P. Williams, M. Kipping, J.P.C. Vissers, Traveling Wave Ion Mobility Mass Spectrometry.

[164] V.R. Meyer, Practical High-Performance Liquid Chromatography, 5th ed., Wiley, 2010.

[165] C. Horváth, Ed., High-Performance Liquid Chromatography. Advances and Perspectives, Elsevier Inc., 1988.

[166] H. Colin, G. Guiochon, M. Martin, Chapter 1. Liquid Chromatographic Equipment, in: H. Engelhardt (Ed.), Practice of High Performance Liquid Chromatography. Applications, Equipment and Quantitative Analysis, Springer-Verlag Berlin Heidelberg, 1986, pp. 1–64.

[167] 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 (2014) 321–333.

[168] Y. Mechref, M. V Novotny, Structural investigations of glycoconjugates at high sensitivity, Chem. Rev. 102 (2002) 321–369.

[169] M. V. Novotny, Y. Mechref, New hyphenated methodologies in high-sensitivity glycoprotein analysis, J. Sep. Sci. 28 (2005) 1956–1968.

[170] K.A. Rubinson, J.F. Rubinson, Eds., Chapter 14. Liquid Chromatography, Contemporary Instrumental Analysis, Prentice-Hall, 2000, pp. 628–671.

[171] J. Šesták, D. Moravcová, V. Kahle, Instrument platforms for nano liquid chromatography, J. Chromatogr. A 1421 (2015) 2–17.

[172] E. Giménez, M. Gay, M. Vilaseca, Automatic and rapid identification of glycopeptides by nano-UPLC-LTQ-FT-MS and proteomic search engine, J. Proteomics 152 (2017) 236–242.

[173] 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.

[174] B. Wang, Y. Tsybovsky, K. Palczewski, M.R. Chance, Reliable Determination of Site-specific in vivo Protein N- Glycosylation based on Collision-Induced MS/MS and Chromatographic Retention Time Benlian, J Am Soc Mass Spectrom 25 (2014) 729–741.

[175] 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.

[176] W.R. Alley Jr, M.V. Novotny, Structural Glycomic Analyses at High Sensitivity: A Decade of Progress, Annu. Rev. Anal. Chem. 6 (2013) 237–265.

[177] M. Wuhrer, A.R. de Boer, A.M. Deelder, Structural glycomics using hydrophilic interaction

chromatography (HILIC) with mass spectrometry, Mass Spectrom. Rev. 28 (2009) 192–206.

[178] G. Zauner, A.M. Deelder, M. Wuhrer, Recent advances in hydrophilic interaction liquid chromatography (HILIC) for structural glycomics, Electrophoresis 32 (2011) 3456–3466.

[179] A. Helander, A. Husa, J.O. Jeppsson, Improved HPLC Method for Carbohydrate-deficient Transferrin in Serum, Clin. Chem. 49 (2003) 1881–1890.

[180] U.K. Laemmli, Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4, Nature 227 (1970) 680–685.

[181] Q. Shi, G. Jackowski, Chapter 1. One-dimensional polyacrylamide gel electrophoresis, in: B.D. Hames (Ed.), Gel Elektrophoresis of Proteins: A Practical Approach, Oxford University Press, 1998, pp. 1–52.

[182] S. Magdeldin, S. Enany, Y. Yoshida, B. Xu, Y. Zhang, Z. Zureena, I. Lokamani, E. Yaoita, T. Yamamoto, Basics and recent advances of two dimensional- polyacrylamide gel electrophoresis, Clin. Proteomics 11 (2014) 16.

[183] A. Görg, W. Weiss, M.J. Dunn, Current two-dimensional electrophoresis technology for proteomics, Proteomics 4 (2004) 3665–3685.

[184] A. Rogowska-Wrzesinska, M.C. Le Bihan, M. Thaysen-Andersen, P. Roepstorff, 2D gels still have a niche in proteomics, J. Proteomics 88 (2013) 4–13.

[185] D.R.M. Graham, M.J. Mitsak, S.T. Elliott, D. Chen, S.A. Whelan, G.W. Hart, J.E. van Eyk, Two-Dimensional Gel Based Approaches for the Assessment of N-Linked and O-GlcNAc Glycosylation in Human and Simian Immunodeficiency Viruses, Proteomics 8 (2008) 4919–4930.

[186] E. Llop, R.G. Gallego, V. Belalcazar, G.J. Gerwig, J.P. Kamerling, J. Segura, J.A. Pascual, Evaluation of protein N-glycosylation in 2-DE: Erythropoietin as a study case, Proteomics 7 (2007) 4278–4291.

[187] F.E. Ahmed, The role of capillary electrophoresis-mass spectrometry to proteome analysis and biomarker discovery, J. Chromatogr. B Anal. Technol. Biomed. Life Sci. 877 (2009)

1963–1981.

[188] H. Whatley, Chapter 2. Basic principles and modes of capillary electrophoresis, in: J. Petersen, A.A. Mohammad (Eds.), Clinical and Forensic Applications of Capillary Electrophoresis, Humana Press, 2001, pp. 21–58.

[189] H.H. Lauer, G.P. Rozing, Chapter 2. Modes of operation, High performance capillary electrophoresis, Agilent Technologies, 2014, pp. 33–56.

[190] European Pharmacopoeia, 4 (2002) 1123–1148.

[191] J. Zaia, Capillary electrophoresis-mass spectrometry of carbohydrates, Methods Mol. Biol. 984 (2013) 13–25.

[192] Y. Mechref, M. V. Novotny, Glycomic analysis by capillary electrophoresis-mass spectrometry, Mass Spectrom. Rev. 28 (2009) 207–222.

[193] S. Amon, A.D. Zamfir, A. Rizzi, Glycosylation analysis of glycoproteins and proteoglycans using capillary electrophoresis-mass spectrometry strategies., Electrophoresis 29 (2008) 2485–2507.

[194] D.F. Cortes, J.L. Kabulski, A.C. Lazar, I.M. Lazar, Recent advances in the MS analysis of glycoproteins: Capillary and microfluidic workflows, Electrophoresis 32 (2011) 14–29.

[195] R.D. Smith, C.J. Barinaga, H.R. Udseth, Improved Electrospray Ionization Interface for Capillary Zone-Mass Spectrometry, Anal. Chem. 60 (1988) 1948–1952.

[196] M. Herrero, E. Ibañez, A. Cifuentes, Capillary electrophoresis-electrospray-mass spectrometry in peptide analysis and peptidomics, Electrophoresis 29 (2008) 2148–2160.

[197] W. Zhang, T. Hankemeier, R. Ramautar, Next-generation capillary electrophoresis-mass spectrometry approaches in metabolomics, Curr. Opin. Biotechnol. 43 (2017) 1–7.

[198] H.J. Issaq, G.M. Janini, K.C. Chan, T.D. Veenstra, Sheathless electrospray ionization interfaces for capillary electrophoresis-mass spectrometric detection: Advantages and limitations, J. Chromatogr. A 1053 (2004) 37–42.

[199] A.D. Zamfir, Recent advances in sheathless interfacing of capillary electrophoresis and

electrospray ionization mass spectrometry, J. Chromatogr. A 1159 (2007) 2–13.

[200] M. Moini, Simplifying CE-MS operation. 2. Interfacing low-flow separation techniques to mass spectrometry using a porous tip, Anal. Chem. 79 (2007) 4241–4246.

[201] V. Kašička, Recent developments in CE and CEC of peptides (2009-2011), Electrophoresis 33 (2012) 48–73.

[202] P.G. Righetti, R. Sebastiano, A. Citterio, Capillary electrophoresis and isoelectric focusing in peptide and protein analysis, Proteomics 13 (2013) 325–340.

[203] S. Mittermayr, M. Olajos, T. Chovan, G.K. Bonn, A. Guttman, Mobility modeling of peptides in capillary electrophoresis, Trends Anal. Chem. 27 (2008) 407–417.

[204] M. Sugimoto, A. Hirayama, M. Robert, S. Abe, T. Soga, M. Tomita, Prediction of metabolite identity from accurate mass, migration time prediction and isotopic pattern information in CE-TOFMS data, Electrophoresis 31 (2010) 2311–2318.

[205] S. Català-Clariana, F. Benavente, E. Giménez, J. Barbosa, V. Sanz-Nebot, Identification of bioactive peptides in hypoallergenic infant milk formulas by CE-TOF-MS assisted by semiempirical model of electromigration behavior, Electrophoresis 34 (2013) 1886–1894.

[206] E.C. Rickard, M.M. Strohl, R.G. Nielsen, Correlation of electrophoretic mobilities from capillary electrophoresis with physicochemical properties of proteins and peptides, Anal. Biochem. 197 (1991) 197–207.

[207] N.J. Adamson, E.C. Reynolds, Rules relating electrophoretic mobility, charge and molecular size of peptides and proteins, J. Chromatogr. B 699 (1997) 133–147.

[208] A. Cifuentes, H. Poppe, Behavior of peptides in capillary electrophoresis: Effect of peptide charge, mass and structure, Electrophoresis 18 (1997) 2362–2376.

[209] N. Hashii, A. Harazono, 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.

[210] J.E. Jackson, A Users' Guide to Principal Components, John Wiley & Sons, Inc., 1991.

[211] S. Wold, K. Esbensen, P. Geladi, Principal component analysis, Chemom. Intell. Lab. Syst 2 (1987) 37–52.

[212] I.T. Jolliffe, Principal component analysis, Springer-Verlag New York, 2002.

[213] T. Rajalahti, O.M. Kvalheim, Multivariate data analysis in pharmaceutics: A tutorial review, Int. J. Pharm. 417 (2011) 280–290.

[214] K.L. Sainani, Introduction to principal components analysis, PM R 6 (2014) 275–278.

[215] I.T. Jolliffe, J. Cadima, Principal component analysis : a review and recent developments, Philos. Trans. R. Soc. A 374 (2016) 20150202.

[216] 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, J. Sci. Stat. Comput. 5 (1984) 735–743.

[217] P. Geladi, B.R. Kowalski, Partial least-squares regression—a tutorial, Anal. Chim. Acta 185 (1986) 1–17.

[218] S. Wold, M. Sjöström, L. Eriksson, PLS-regression: a basic tool of chemometrics, Chemom. Intell. Lab. Syst. 58 (2001) 109–130.

[219] R.G. Brereton, G.R. Lloyd, Partial least squares discriminant analysis: Taking the magic away, J. Chemom. 28 (2014) 213–225.

[220] 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.

[221] C.C. Chen, W.C. Su, B.Y. Huang, Y.J. Chen, H.C. Tai, R.P. Obena, Interaction modes and approaches to glycopeptide and glycoprotein enrichment, Analyst 139 (2014) 688–704.

[222] Y.H. Ahn, J.Y. Kim, J.S. Yoo, Quantitative mass spectrometric analysis of glycoproteins combined with enrichment methods, Mass Spectrom. Rev. 34 (2015) 148–165.

[223] Z. Zhu, H. Desaire, Carbohydrates on Proteins: Site-Specific Glycosylation Analysis by Mass Spectrometry, Annu. Rev. Anal. Chem. 8 (2015) 463–483.

[224] R. Chen, X. Jiang, D. Sun, G. Han, F. Wang, M. Ye, L. Wang, H. Zou, Glycoproteomics analysis of human liver tissue by combination of multiple enzyme digestion and hydrazide chemistry, J. Proteome Res. 8 (2009) 651–661.

[225] Y. Wang, M. Liu, L. Xie, C. Fang, H. Xiong, H. Lu, Highly efficient enrichment method for glycopeptide analyses: Using specific and nonspecific nanoparticles synergistically, Anal. Chem. 86 (2014) 2057–2064.

[226] N. Sun, C. Deng, Y. Li, X. Zhang, Highly selective enrichment of N-linked glycan by carbon-functionalized ordered graphene/mesoporous silica composites, Anal. Chem. 86 (2014) 2246–2250.

[227] J. Zheng, Y. Xiao, L. Wang, Z. Lin, H. Yang, L. Zhang, G. Chen, Click synthesis of glucose-functionalized hydrophilic magnetic mesoporous nanoparticles for highly selective enrichment of glycopeptides and glycans, J. Chromatogr. A 1358 (2014) 29–38.

[228] Q. Cao, C. Ma, H. Bai, X. Li, H. Yan, Y. Zhao, W. Ying, X. Qian, Multivalent hydrazide-functionalized magnetic nanoparticles for glycopeptide enrichment and identification, Analyst 139 (2014) 603–609.

[229] C. Lei, K. Qian, O. Noonan, A. Nouwens, C. Yu, Applications of nanomaterials in mass spectrometry analysis, Nanoscale 5 (2013) 12033–12042.

[230] C. Wang, M. Gao, P. Zhang, X. Zhang, Efficient proteolysis of glycoprotein using a hydrophilic immobilized enzyme reactor coupled with MALDI-QIT-TOF-MS detection and μHPLC analysis, Chromatographia 77 (2014) 413–418.

[231] J. Ma, L. Zhang, Z. Liang, Y. Shan, Y. Zhang, Immobilized enzyme reactors in proteomics, Trends Anal. Chem. 30 (2011) 691–702.

[232] B.H. Clowers, E.D. Dodds, R.R. Seipert, C.B. Lebrilla, Site determination of protein glycosylation based on digestion with immobilized nonspecific proteases and fourier transform ion cyclotron resonance mass spectrometry, J. Proteome Res. 6 (2007) 4032–4040.

[233] Y. Mechref, Y. Hu, J.L. Desantos-Garcia, A. Hussein, H. Tang, Quantitative glycomics

strategies, Mol. Cell. proteomics 12 (2013) 874–884.

[234] B.L. Parker, M. Thaysen-Andersen, N. Solis, N.E. Scott, M.R. Larsen, M.E. Graham, N.H. Packer, S.J. Cordwell, Site-specific glycan-peptide analysis for determination of N-glycoproteome heterogeneity, J. Proteome Res. 12 (2013) 5791–5800.

[235] C.C. Nwosu, R.R. Seipert, J.S. Strum, S.S. Hua, H.J. An, A.M. Zivkovic, B.J. German, C.B. Lebrilla, Simultaneous and Extensive Site-specific N- and O- Glycosylation Analysis in Protein Mixtures, J. Proteome Res. 10 (2011) 2612–2624.

[236] L.R. Ruhaak, A.M. Deelder, M. Wuhrer, Oligosaccharide analysis by graphitized carbon liquid chromatography–mass spectrometry, Anal. Bioanal. Chem. 394 (2009) 163–174.

[237] G.O. Staples, M.J. Bowman, C.E. Costello, A.M. Hitchcock, J.M. Lau, N. Leymarie, C. Miller, H. Naimy, X. Shi, J. Zaia, A Chip-based Amide-HILIC LC/MS Platform for Glycosaminoglycan Glycomics Profiling, Proteomics 9 (2009) 686–695.

[238] H. Malerod, M. Rogeberg, N. Tanaka, T. Greibrokk, E. Lundanes, Large volume injection of aqueous peptide samples on a monolithic silica based zwitterionic-hydrophilic interaction liquid chromatography system for characterization of posttranslational modifications, J. Chromatogr. A 1317 (2013) 129–137.

[239] D. Wang, M. Hincapie, T. Rejtar, B.L. Karger, Ultrasensitive Characterization of Site-Specific Glycosylation of Affinity-Purified Haptoglobin from Lung Cancer Patient Plasma Using 10 µm i.d. Porous Layer Open Tubular Liquid Chromatography- Linear Ion Trap Collision-Induced Dissociation/Electron Transfer Dissociation Mass Spectrometry, Anal. Chem. 83 (2011) 2029–2037.

[240] I. Mitra, W.R. Alley Jr., J.A. Goetz, J.A. Vasseur, M. V. Novotny, S.C. Jacobson, Comparative Profiling of N-Glycans Isolated from Serum Samples of Ovarian Cancer Patients and Analyzed by Microchip Electrophoresis, J. Proteome Res. 12 (2013) 4490–4496.

[241] Y. Mechref, Analysis of glycans derived from glycoconjugates by capillary electrophoresis-mass spectrometry, Electrophoresis 32 (2011) 3467–3481.

[242] E.A. Redman, J.S. Mellors, J.A. Starkey, J.M. Ramsey, Characterization of Intact Antibody Drug Conjugate Variants Using Microfluidic Capillary Electrophoresis-Mass Spectrometry, Anal. Chem. 88 (2016) 2220–2226.

[243] I. Mitra, C.M. Snyder, X. Zhou, M.I. Campos, W.R. Alley Jr., M. V. Novotny, S.C. Jacobson, Structural Characterization of Serum N-Glycans by Methylamidation, Fluorescent Labeling, and Analysis by Microchip Electrophoresis, Anal. Chem. 88 (2016) 8965–8971.

[244] C. Schwedler, M. Kaup, S. Weiz, M. Hoppe, E.I. Braicu, J. Sehouli, B. Hoppe, R. Tauber, M. Berger, V. Blanchard, Identification of 34 N-glycan isomers in human serum by capillary electrophoresis coupled with laser-induced fluorescence allows improving glycan biomarker discovery, Anal. Bioanal. Chem. 406 (2014) 7185–7193.

[245] R.G. Jayo, M. Thaysen-Andersen, P.W. Lindenburg, R. Haselberg, T. Hankemeier, R. Ramautar, D.D.Y. Chen, Simple capillary electrophoresis-mass spectrometry method for complex glycan analysis using a flow-through microvial interface, Anal. Chem. 86 (2014) 6479–6486.

[246] Y.W. Zhang, M.Z. Zhao, J.X. Liu, Y.L. Zhou, X.X. Zhang, Double-layer poly(vinyl alcohol)-coated capillary for highly sensitive and stable capillary electrophoresis and capillary electrophoresis with mass spectrometry glycan analysis, J. Sep. Sci. 38 (2015) 475–482.

[247] M. Pabst, F. Altmann, Glycan analysis by modern instrumental methods, Proteomics 11 (2011) 631–643.

[248] N. Leymarie, J. Zaia, Effective use of mass spectrometry for glycan and glycopeptide structural analysis, Anal. Chem. 84 (2012) 3040–3048.

[249] M. Sarbu, R.M. Ghiulai, A.D. Zamfir, Recent developments and applications of electron transfer dissociation mass spectrometry in proteomics, Amino Acids 46 (2014) 1625–1634.

[250] R. Goldman, M. Sanda, Targeted methods for quantitative analysis of protein glycosylation, Proteomics - Clin. Appl. 9 (2015) 17–32.

[251] 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 α1-3 fucosylation of α-1-acid glycoprotein (AGP) in pancreatic cancer, J. Proteomics 132 (2016) 144–154.

[252] M. Hedlund, E. Ng, A. Varki, N.M. Varki, α2-6-linked sialic acids on N-glycans modulate carcinoma differentiation in vivo, Cancer Res. 68 (2008) 388–394.

[253] A. Sarrats, R. Saldova, E. Pla, E. Fort, D.J. Harvey, W.B. Struwe, R. de Llorens, P.M. Rudd, R. Peracaula, Glycosylation of liver acute-phase proteins in pancreatic cancer and chronic pancreatitis, Proteomics-Clinical Appl. 4 (2010) 432–448.

[254] 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.

[255] L. Hajba, E. Csanky, A. Guttman, Liquid phase separation methods for N-glycosylation analysis of glycoproteins of biomedical and biopharmaceutical interest. A critical review, Anal. Chim. Acta 943 (2016) 8–16.

[256] D. Kolarich, M. Windwarder, K. Alagesan, F. Altmann, Isomer-Specific Analysis of Released N-Glycans by LC-ESI MS/MS with Porous Graphitized Carbon, in: A. Castilho (Ed.), Glyco-Engineering: Methods and Protocols, Springer New York, 2015, pp. 427–435.

[257] K. Stavenhagen, D. Kolarich, M. Wuhrer, Clinical Glycomics Employing Graphitized Carbon Liquid Chromatography–Mass Spectrometry, Chromatographia 78 (2015) 307–320.

[258] A.D. Catherman, K.R. Durbin, D.R. Ahlf, B.P. Early, R.T. Fellers, J.C. Tran, P.M. Thomas, N.L. Kelleher, Large-scale top-down proteomics of the human proteome: membrane proteins, mitochondria, and senescence, Mol. Cell. Proteomics 12 (2013) 3465–3473.

[259] A.D. Catherman, O.S. Skinner, N.L. Kelleher, Top Down proteomics: Facts and perspectives, Biochem. Biophys. Res. Commun. 445 (2014) 683–693.

[260] K. Scheffler, Chapter 31. Top-Down Proteomics by Means of Orbitrap Mass Spectrometry, in: D. Martins-de-Souza (Ed.), Shotgun Proteomics. Methods and Protocols,

Humana Press, 2014, pp. 465–487.

[261] K.P. Gopaul, M.A. Crook, Sialic acid: A novel marker of cardiovascular disease?, Clin. Biochem. 39 (2006) 667–681.

[262] Y. Kizuka, S. Kitazume, N. Taniguchi, N -glycan and Alzheimer's disease, Biochim. Biophys. Acta (2017). DOI: 10.1016/j.bbagen.2017.04.012.

[263] J.J. Heinisch, R. Brandt, Signaling pathways and posttranslational modifications of tau in Alzheimer's disease: the humanization of yeast cells, Microb. Cell 3 (2016) 135–146.

[264] W.Y. Wani, J.C. Chatham, V. Darley-Usmar, L.L. McMahon, J. Zhang, O-GlcNAcylation and neurodegeneration, Brain Res. Bull. (2016). DOI: 10.1016/j.brainresbull.2016.08.002.

[265] Y. Zhu, X. Shan, S.A. Yuzwa, D.J. Vocadlo, The emerging link between O-GlcNAc and Alzheimer disease, J. Biol. Chem. 289 (2014) 34472–34481.

[266] S.R. Stowell, T. Ju, R.D. Cummings, Protein glycosylation in cancer, Annu Rev Pathol 10 (2015) 473–510.

[267] V. Padler-Karavani, Aiming at the sweet side of cancer: Aberrant glycosylation as possible target for personalized-medicine, Cancer Lett. 352 (2014) 102–112.

[268] R. Peracaula, S. Barrabés, A. Sarrats, P.M. Rudd, R. De Llorens, Altered glycosylation in tumours focused to cancer diagnosis, Dis. Markers 25 (2008) 207–218.

[269] J. Munkley, I.G. Mills, D.J. Elliott, The role of glycans in the development and progression of prostate cancer, Nat Rev Urol 13 (2016) 324–333.

[270] B.N. Vajaria, P.S. Patel, Glycosylation: a hallmark of cancer?, Glycoconj. J. 34 (2017) 147–156.

[271] A. Mehta, H. Herrera, T. Block, Glycosylation and Liver Cancer, Adv. Cancer Res. 126 (2015) 257–279.

[272] R. Saldova, L. Royle, C.M. Radcliffe, U.M. Abd hamid, R. Evans, J.N. Arnold, R.E. Banks, R. Hutson, D.J. Harvey, R. Antrobus, S.M. Petrescu, R.A. Dwek, P.M. Rudd, Ovarian cancer is associated with changes in glycosylation in both acute-phase proteins and IgG, Glycobiology 17

(2007) 1344–1356.

[273] A. Kirwan, M. Utratna, M.E. O'Dwyer, L. Joshi, M. Kilcoyne, Glycosylation-Based Serum Biomarkers for Cancer Diagnostics and Prognostics, Biomed Res. Int. (2015). DOI: 10.1155/2015/490531.

[274] S. Gilgunn, P.J. Conroy, R. Saldova, P.M. Rudd, R.J. O'Kennedy, Aberrant PSA glycosylation—a sweet predictor of prostate cancer, Nat. Rev. Urol. 10 (2013) 99–107.

[275] F. Cartault, P. Munier, M.L. Jacquemont, J. Vellayoudom, B. Doray, C. Payet, H. Randrianaivo, J.M. Laville, A. Munnich, V. Cormier-Daire, Expanding the clinical spectrum of B4GALT7 deficiency: homozygous p.R270C mutation with founder effect causes Larsen of Reunion Island syndrome, Eur. J. Hum. Genet. 23 (2015) 49–53.

[276] H.H. Freeze, E.A. Eklund, B.G. Ng, M.C. Patterson, Neurological aspects of human glycosylation disorders, Annu. Rev. Anal. Chem. 38 (2015) 105–125.

[277] Pierce Trypsin Protease , MS Grade. Instructions, Thermo Scientific.

[278] A.J. Link, J. LaBaer, Solution protein digest, Cold Spring Harb. Protoc. 6 (2011) 202–207.

[279] R.L. Gundry, M.Y. White, C.I. Murray, L.A. Kane, Q. Fu, B.A. Stanley, J.E. Van Eyk, Preparation of Proteins and Peptides for Mass Spectrometry Analysis in a Bottom-Up Proteomics Workflow, Curr. Protoc. Mol. Biol. (2009) 88:VI:10.25:10.25.1–10.25.23.

[280] G.W. Slysz, D.C. Schriemer, On-column digestion of proteins in aqueous-organic solvents, Rapid Commun. Mass Spectrom. 17 (2003) 1044–1050.

[281] S.J. Bark, N. Muster, J.R. Yates III, G. Siuzdak, High-temperature protein mass mapping using a thermophilic protease, J. Am. Chem. Soc. 123 (2001) 1774–1775.

[282] D.J. Janecki, J.P. Reilly, Denaturation of metalloproteins with EDTA to facilitate enzymatic digestion and mass fingerprinting, Rapid Commun. Mass Spectrom. 19 (2005) 1268–1272.

[283] F. Benavente, B. Andón, E. Giménez, J. Barbosa, V. Sanz-Nebot, Modeling the migration behavior of rabbit liver apothioneins in capillary electrophoresis, Electrophoresis 29 (2008) 2790–2800.

[284] R. Ullmer, A.M. Rizzi, Use of a novel ionic liquid matrix for MALDI-MS analysis of glycopeptides and glycans out of total tryptic digests, J. mass Spectrom. 44 (2009) 1596–1603.

[285] Z.M. Segu, L.A. Hammad, Y. Mechref, Rapid and efficient glycoprotein identification through microwave-assisted enzymatic digestion, Rapid Commun. mass Spectrom. 24 (2010) 3461–3468.

[286] E.I. Chen, D. Cociorva, J.L. Norris, J.R. Yates III, Optimization of Mass Spectrometry Compatible Surfactants for Shotgun Proteomics, J. Proteome Res. 6 (2007) 2529–2538.

[287] E.I. Chen, D. McClatchy, K.P. Sung, J.R. Yates III, Comparisons of mass spectrometry compatible surfactants for global analysis of the mammalian brain proteome, Anal. Chem. 80 (2008) 8694–8701.

[288] Y.Q. Yu, M. Gilar, P.J. Lee, E.S.P. Bouvier, J.C. Gebler, Enzyme-Friendly, Mass Spectrometry-Compatible Surfactant for In-Solution Enzymatic Digestion of Proteins, Anal. Chem. 75 (2003) 6023–6028.

[289] C. Pop, C. Mogosan, F. Loghin, Evaluation of Rapigest Efficacy for the Digestion of Proteins From Cell Cultures and Heart Tissue, Clujul Med. 87 (2014) 258–262.

[290] J.L. Proc, M. a Kuzyk, D.B. Hardie, J. Yang, D.S. Smith, A.M. Jackson, C.E. Parker, C.H. Borchers, A Quantitative Study of the Effects of Chaotropic Agents, Surfactants, and Solvents on the Digestion Efficiency of Human Plasma Proteins by Trypsin, J. Proteome Res. 9 (2010) 5422–5437.

[291] V. Sanz-Nebot, P. González, I. Toro, A. Ribes, J. Barbosa, Characterization of human transferrin glycoforms by capillary electrophoresis and electrospray ionization mass spectrometry, J. Chromatogr. B 798 (2003) 1–7.

[292] V. Sanz-Nebot, E. Balaguer, F. Benavente, C. Neusüß, J. Barbosa, Characterization of transferrin glycoforms in human serum by CE-UV and CE-ESI-MS, Electrophoresis 28 (2007) 1949–1957.

[293] S. Medina-Casanellas, F. Benavente, J. Barbosa, V. Sanz-Nebot, Preparation and

evaluation of an immunoaffinity sorbent for the analysis of opioid peptides by on-line immunoaffinity solid-phase extraction capillary electrophoresis–mass spectrometry, Anal. Chim. Acta 789 (2013) 91–99.

[294] M. Barker, W. Rayens, Partial least squares for discrimination, J. Chemom. 17 (2003) 166–173.

[295] A. Helandera, J. Jaekenc, G. Matthijsd, G. Eggertsena, Asymptomatic phosphomannose isomerase deficiency (MPI-CDG) initially mistaken for excessive alcohol consumption, Clin. Chim. Acta 431 (2014) 15–18.

[296] J. Postigo, M. Iglesias, D. Cerezo-Wallis, A. Rosal-Vela, S. García-Rodríguez, M. Zubiaur, J. Sancho, R. Merino, J. Merino, Mice deficient in CD38 develop an attenuated form of collagen type II-induced arthritis, PLoS One 7 (2012). DOI: 10.1371/journal.pone.0033534.

[297] M.I. Kamboh, R.E. Ferrell, Human transferrin polymorphism, Hum. Hered. 37 (1987) 65–81.

[298] E. Marklová, Z. Albahri, Transferrin D protein variants in the diagnosis of congenital disorders of glycosylation (CDG), J. Clin. Lab. Anal. 23 (2009) 77–81.

[299] A. Zühlsdorf, J.H. Park, Y. Wada, S. Rust, J. Reunert, I. DuChesne, M. Grüneberg, T. Marquardt, Transferrin variants: pitfalls in the diagnostics of Congenital disorders of glycosylation, Clin. Biochem. 48 (2015) 11–13.

[300] J.H. Park, A. Zühlsdorf, Y. Wada, C. Roll, S. Rust, I. Du Chesne, M. Grüneberg, J. Reunert, T. Marquardt, The novel transferrin E592A variant impairs the diagnostics of congenital disorders of glycosylation, Clin. Chim. Acta 436 (2014) 135–139.

[301] A. Helander, G. Eriksson, H. Stibler, J.O. Jeppsson, Interference of transferrin isoform types with carbohydrate-deficient transferrin quantification in the identification of alcohol abuse, Clin. Chem. 47 (2001) 1225–1233.

[302] A. Grahn, P. Bengtson, E. Eklund, J. Asin-Cayuela, A novel mutation on the transferrin gene abolishes one N-glycosylation site and alters the pattern of transferrin isoforms,

mimicking that observed after excessive alcohol consumption, Clin. Biochem. 49 (2016) 511–513.

[303] A. Sillero, J.M. Ribeiro, Isoelectric points of proteins: theoretical determination, Anal. Biochem. 179 (1989) 319–325.

[304] K.I. Roy, C.A. Lucy, Influence of methanol as a buffer additive on the mobilities of organic cations in capillary electrophoresis, Electrophoresis 24 (2003) 370–379.

[305] R. Lee, A.S. Ptolemy, L. Niewczas, P. Britz-McKibbin, Integrative metabolomics for characterizing unknown low-abundance metabolites by capillary electrophoresis-mass spectrometry with computer simulations, Anal. Chem 79 (2007) 403–415.

[306] M. Guttman, K.K. Lee, Site-Specific Mapping of Sialic Acid Linkage Isomers by Ion Mobility Spectrometry, Anal. Chem. 88 (2016) 5212–5217.

[307] Corina Demo Interactive, Mol. Networks GmbH Altamira, LLC . https://www.mn-com/online_demos/corina_demo_int.

[308] C. Larriba, C.J. Hogan Jr., Free molecular collision cross section calculation methods for nanoparticles and complex ions with energy accommodation, J. Comput. Phys. 251 (2013) 344–363.

[309] C. Larriba, C.J. Hogan Jr., Ion mobilities in diatomic gases: Measurement versus prediction with non-specular scattering models, J. Phys. Chem. A 117 (2013) 3887–3901.

[310] The University of Georgia. Complex Carbohydrate Research Center (CCRC), www.glycam.org.