



UNIVERSITAT DE
BARCELONA

Detección y caracterización de péptidos, proteínas y glicoproteínas como biomarcadores en fluidos biológicos. Aplicación al diagnóstico y seguimiento de diversas patologías

Laura Pont Villanueva

ADVERTIMENT. La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX (www.tdx.cat) i a través del Dipòsit Digital de la UB (diposit.ub.edu) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX ni al Dipòsit Digital de la UB. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX o al Dipòsit Digital de la UB (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

ADVERTENCIA. La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR (www.tdx.cat) y a través del Repositorio Digital de la UB (diposit.ub.edu) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR o al Repositorio Digital de la UB. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR o al Repositorio Digital de la UB (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

WARNING. On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX (www.tdx.cat) service and by the UB Digital Repository (diposit.ub.edu) has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized nor its spreading and availability from a site foreign to the TDX service or to the UB Digital Repository. Introducing its content in a window or frame foreign to the TDX service or to the UB Digital Repository is not authorized (framing). Those rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.



UNIVERSITAT^{DE}
BARCELONA

FACULTAD DE QUÍMICA

DEPARTAMENTO DE INGENIERÍA QUÍMICA Y QUÍMICA ANALÍTICA

Programa de Doctorado:

Química Analítica del Medi Ambient i la Pol·lució

**Detección y caracterización de péptidos, proteínas y glicoproteínas
como biomarcadores en fluidos biológicos. Aplicación al diagnóstico y
seguimiento de diversas patologías**

Memoria presentada por

Laura Pont Villanueva

para optar al grado de Doctora por la Universidad de Barcelona

Bajo la dirección de

Dra. Victoria Sanz Nebot

Dr. Fernando J. Benavente Moreno

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 el Dr. Fernando J. Benavente Moreno, profesor agregado del mismo Departamento,

HACEN CONSTAR,

Que la presente memoria titulada: “**Detección y caracterización de péptidos, proteínas y glicoproteínas como biomarcadores en fluidos biológicos. Aplicación al diagnóstico y seguimiento de diversas patologías**”, ha sido realizada bajo su dirección por la Sra. **Laura Pont Villanueva** y que todos los resultados presentados son fruto de las experiencias realizadas por la citada doctoranda.

Y para que así conste, expiden el presente certificado.

Barcelona, Junio 2017

Dra. Victoria Sanz Nebot

Dr. Fernando J. Benavente Moreno

“The art and science of asking questions is the source of all knowledge”

Thomas Berger

Agradecimientos

Después de mucho esfuerzo y dedicación, ha llegado el final de esta bonita etapa de mi vida y no me gustaría acabar sin agradecer a todas aquellas personas que, de alguna manera, han contribuido a la realización de este trabajo.

En primer lugar, quería agradecer todo el trabajo realizado por mis directores de tesis, la Dra. Victoria Sanz Nebot y el Dr. Fernando J. Benavente Moreno. Vicki, gracias por tu dedicación, por tus ánimos y por el esfuerzo puesto en la realización de este trabajo. Gracias a ti Fernando, por tus ideas día a día, por tu apoyo incondicional cuando las cosas no salían del todo bien, por decirme que es necesario tener fe en lo que haces, por tus horas de dedicación y porque sin ti este trabajo no sería el mismo. También me gustaría agradecer al Dr. José Barbosa Torralbo, por aceptarme en el grupo y por los consejos y orientación recibidos a lo largo del doctorado. Gracias también a la Dra. Estela Giménez López, por ayudarme siempre que he tenido alguna duda y por aconsejarme en todo lo posible. A todos vosotros, gracias por transmitirme vuestros conocimientos y vuestro entusiasmo por el mundo de la investigación.

En especial, me gustaría agradecer a mis compañeros de laboratorio, con quienes he compartido todo tipo de experiencias y muchas horas de trabajo. A ti Albert, me cuesta resumir en unas pocas líneas todo lo que me gustaría decirte. Gracias no sólo por ser un gran compañero de trabajo, sino por haber sido mi mejor amigo durante todo este tiempo, por los buenos momentos vividos y por pasar juntos el estrés de estas últimas semanas. Me quedo con un sinfín de recuerdos y experiencias positivas, y espero que podamos seguir compartiendo todo lo bueno, y no tan bueno, que nos depare el futuro. Montse, gracias por haber revolucionado el laboratorio, por tu energía positiva, por haber conseguido formar un grupo tan dispar y, en especial, por ser como eres. Y como ya te dije una vez, ¡no dejes nunca de sonreír! A ti Roger, gracias por aportar un poco de serenidad y cordura al grupo, por tu apoyo durante todo este tiempo y por demostrarme día a día que eres un gran compañero y amigo. Irena, gracias por tu apoyo incondicional, por tu constancia, por tus ganas de aprender y por mirarme de esa manera tan especial que sólo tú y yo sabemos. Y a mis chicas preferidas de máster, Lorena y Laura, gracias

por el apoyo y por todas las risas y buenos momentos vividos, tanto en el laboratorio como en nuestras divertidas excursiones. Todos juntos formamos un gran equipo, y espero que esta amistad que hemos forjado se mantenga viva durante mucho tiempo.

También me gustaría agradecer el compañerismo y buen ambiente que han creado todas las personas que han formado parte de este grupo de investigación: Silvia, Lorena, Marcos, Marta, Fran, Leo, Mulu, Marina, Amélie, Fatma... Gracias por vuestros consejos y por hacerme sentir apoyada en todo momento. Nunca olvidaré vuestro paso por el laboratorio. My dear Kader, I would like to thank you for your support, kindness and friendliness. I hope that someday we will be able to meet again.

I would also like to thank Dr. Oleg Mayboroda and Dr. Manfred Wuhrer for giving me the opportunity of working in their laboratory and for their kindness and support during my stay in Leiden. Thanks also to Guinevere, Sarah, Alexander, Shivani and Elena for your hospitality and for sharing with me a great and enjoyable experience.

Tampoco podría olvidarme de dar las gracias a mis compañeros y amigos, Gerard, Elena, Sergio, Meri, Ana y Stanis, con los que he compartido muchas risas y los que han hecho que los momentos difíciles fueran menos duros.

Y finalmente, me gustaría agradecer a mis padres, a mis amigos y a Víctor, por haberme escuchado en todo momento, por compartir mi alegría cuando las cosas salían bien, y mi mal humor cuando la situación se complicaba, por estar siempre ahí y por formar parte de mi vida.

En general, quisiera agradecer a todas aquellas personas que, de alguna manera, han contribuido a la realización de este trabajo, ayudándome profesionalmente o compartiendo a mi lado ilusiones y esperanzas. A todos vosotros,

Gracias

Abbreviations and acronyms

A

%A	Relative abundance (%)
ACN	Acetonitrile
AD	Alzheimer's disease
AIMS	Aspiration ion mobility spectrometry
ALS	Amyotrophic lateral sclerosis/Alternating least squares

Amino acids (aa)

Amino acid	Code (3 letters)	Code (1 letter)
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

A_{norm}	Normalized electrophoretic/chromatographic peak area
A_{p}	Electrophoretic/chromatographic peak area

B

BGE	Background electrolyte
BPE	Base peak electropherogram
BSA	Bovine serum albumin

C

C	Constant region
C_8	Octylsilane-bonded silica
C_{18}	Octadecylsilane-bonded silica
C_{18} -SPE-CE-MS	On-line C_{18} solid-phase capillary electrophoresis mass spectrometry

CA	Cellulose acetate
CAG	Cytosine-adenine-guanine
CapLC	Capillary liquid chromatography
CapLC-MS	Capillary liquid chromatography mass spectrometry
CDI	Carbonyldiimidazole
CE	Capillary electrophoresis
CEA	Carcinoembryonic antigen
CEC	Capillary electrochromatography
CE-ESI-MS	Capillary electrophoresis electrospray ionization mass spectrometry
CE-MS	Capillary electrophoresis mass spectrometry
CE-MS/MS	Capillary electrophoresis tandem mass spectrometry
CFG	Consortium for Functional Glycomics
CGE	Capillary gel electrophoresis
CID	Collision induced dissociation
CIEF	Capillary isoelectric focusing
CNSA	Central nervous system selective amyloidosis
CRC	Colorectal cancer
CSF	Cerebrospinal fluid
CZE	Capillary zone electrophoresis

D

DI	Dimer
DMS	Differential-mobility spectrometry
DNA	Deoxyribonucleic acid
DTIMS	Drift time ion mobility spectrometry
DTT	Dithiothreitol
Dyn A	Dynorphin A (1-7)

E

EIC	Extracted ion chromatogram
EIE	Extracted ion electropherogram
EIM	Extracted ion mobility profile
ELISA	Enzyme linked immunosorbent assay
End 1	Endomorphin 1
EOF	Electroosmotic flow
E _r	Mass error
ESI	Electrospray ionization
ESI-MS	Electrospray ionization mass spectrometry

F

Fab'	Antigen binding fragment
------	--------------------------

Fab'-IA-SPE-CE-MS	On-line immunoaffinity solid-phase extraction capillary electrophoresis mass spectrometry with Fab' antibody fragments
FAC	Familial amyloidotic cardiomyopathy
FAIMS	Field asymmetric waveform ion mobility spectrometry
FAP	Familial amyloidotic polyneuropathy
FAP-I	Familial amyloidotic polyneuropathy type I
Fc	Fragment crystallizable region
FTICR	Fourier transform ion cyclotron resonance mass analyzer

G

GC	Gas chromatography
GC-MS	Gas chromatography mass spectrometry

H

H	Heavy chain
HAc	Acetic acid
HD	Huntington's disease
HFor	Formic acid
HMDB	Human Metabolome Database
HSPS	High sensitivity porous sprayer
HTT	Huntingtin gene

I

IA	Immunoaffinity
IAA	Iodoacetamide
IACE	Immunoaffinity capillary electrophoresis
IA-SPE	Immunoaffinity solid-phase extraction
IA-SPE-CE	On-line immunoaffinity solid-phase extraction capillary electrophoresis
IA-SPE-CE-MS	On-line immunoaffinity solid-phase extraction capillary electrophoresis mass spectrometry
id	Inner diameter
IEF	Isoelectric focusing
Ig	Immunoglobulin
IgG	Immunoglobulin G
IM	Ion mobility
IMAC	Immobilized metal ion affinity chromatography
IM-MS	Ion mobility mass spectrometry
IMS	Ion mobility spectrometry
IP	Immunoprecipitation

iPrOH	Propan-2-ol
IT	Ion trap mass analyzer
IT-MS	Ion trap mass spectrometry
ITP	Isotachophoresis

K

KEGG	Kyoto Encyclopedia of Genes and Genomes
------	---

L

L	Light chain
LC	Liquid chromatography
LC-ESI-MS	Liquid chromatography electrospray ionization mass spectrometry
LC-MS	Liquid chromatography mass spectrometry
LC-MS/MS	Liquid chromatography tandem mass spectrometry
LE	Leading electrolyte
LIF	Laser induced fluorescence
LOD	Limit of detection
L_T	Total length
LV	Latent variable

M

MALDI	Matrix assisted laser desorption ionization
MALDI-TOF-MS	Matrix assisted laser desorption ionization time-of-flight mass spectrometry
MBs	Magnetic beads
MCR-ALS	Multivariate curve resolution alternating least squares
2-MEA	2-mercaptoethylamine
MEKC	Micellar electrokinetic chromatography
MeOH	Methanol
Met	Met-enkephalin
M_{exp}	Experimental molecular mass
MO	Monomer

Monosaccharides

Monosaccharide	Code (1 letter)
Galactose	H
Mannose	H
N-acetylgalactosamine	N
N-acetylglucosamine	N
Fucose	F
Sialic acid	S

M_r	Relative molecular mass
ΔM_r	Relative molecular mass variation
MRI	Magnetic resonance imaging
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
M_{theo}	Theoretical molecular mass
m/z	Mass-to-charge ratio

N

n	Number of replicates
nanoESI	Nano electrospray ionization
nanoESI-IM-MS	Nano electrospray ionization ion mobility mass spectrometry
nanoLC	Nano liquid chromatography
NH ₄ Ac	Ammonium acetate
NHS	N-Hydroxysuccinimide

O

oa	Orthogonal acceleration
oa-TOF	Orthogonal acceleration time-of-flight mass analyzer
OCT	Octamer
od	Outer diameter

P

PBS	Phosphate buffered saline
PC	Principal component
PCA	Principal component analysis
PD	Parkinson's disease
PEG	Polyethylene glycol
PES	Polyethersulfone
pI	Isoelectric point
PLS	Partial least squares
PLS-DA	Partial least squares discriminant analysis
PolyQ	Polyglutamine
ppm	Part-per-million
PTM	Post-translational modification

Q

q	Electric charge
Q	Quadrupole mass analyzer
QqQ	Triple quadrupole mass analyzer

R

r	Ion radius
RNA	Ribonucleic acid
RP-LC	Reversed phase liquid chromatography
%RSD	Relative standard deviation (%)

S

SIMPLISMA	Interactive self modeling mixture analysis
SiOH	Silanol groups
S/N	Signal-to-noise ratio
SOD-I	Superoxide dismutase I
SPE	Solid-phase extraction
SPE-CE	On-line solid-phase extraction capillary electrophoresis
SPE-CE-MS	On-line solid-phase extraction capillary electrophoresis mass spectrometry
SSA	Senile systemic amyloidosis
SSMCC	Sulfosuccinimidyl-4-(N-maleimidomethyl)-cyclohexane-1 carboxylate
SVD	Singular value decomposition

T

TE	Tetramer/Terminating electrolyte
Tf	Transferrin
TIC	Total ion chromatogram
TIE	Total ion electropherogram
tITP	Transient isotachopheresis
t_m	Migration time
TOF	Time-of-flight mass analyzer
TOF-MS	Time-of-flight mass spectrometry
t_r	Retention time
TRI	Trimer
TTR	Transthyretin
TTR(Met30)	Transthyretin with the mutation Met30
TWIMS	Traveling-wave ion mobility spectrometry

U

UHPLC	Ultra-high pressure liquid chromatography
UV	Ultraviolet spectrophotometry
UV-Vis	Ultraviolet-visible

V

V	Variable region
VIP	Variable importance in the projection

W

wt	Wild-type
----	-----------

Symbols

Ω	Collision cross-section
Ω/z	Collision cross-section-to-charge ratio
μ_e	Electrophoretic mobility
η	Viscosity

Index

ABSTRACT	1
AIMS	9
CHAPTER 1. INTRODUCTION	15
1.1. OMICS SCIENCES	17
1.1.1. PROTEOMICS	18
1.1.2. METABOLOMICS	21
1.1.3. TARGETED AND UNTARGETED ANALYSIS	21
1.2. BIOMARKER DISCOVERY	23
1.2.1. AMYLOIDOTIC NEURODEGENERATIVE DISEASES	23
1.2.1.1. Transthyretin (TTR). Hereditary TTR amyloidoses	25
1.2.1.2. Huntington's disease	28
1.2.2. CANCER. COLORECTAL CANCER AND LIVER METASTATIC DISEASE	30
1.2.2.1. Glycosylation	30
1.2.2.2. Human carcinoembryonic antigen	34
1.3. PURIFICATION TECHNIQUES	37
1.3.1. ANTIBODIES	37
1.3.2. CONVENTIONAL IP IN SOLUTION AND IP WITH MAGNETIC BEADS	39
1.4. ANALYTICAL TECHNIQUES	42
1.4.1. CAPILLARY ELECTROPHORESIS (CE)	42
1.4.1.1. Separation principles in CE	43
1.4.2. LIQUID CHROMATOGRAPHY (LC)	45
1.4.3. MASS SPECTROMETRY (MS)	47
1.4.3.1. Ionization techniques. Electrospray ionization (ESI)	47
1.4.3.2. Mass analyzers	53
1.4.4. ION MOBILITY MASS SPECTROMETRY (IM-MS)	55
1.5. ON-LINE PRECONCENTRATION IN CE	59
1.5.1. SPE-CE WITH CONVENTIONAL CHROMATOGRAPHIC SORBENTS	61
1.5.2. SPE-CE WITH IMMUNOAFFINITY SORBENTS (IA-SPE-CE)	62
1.6. DATA ANALYSIS. CHEMOMETRIC METHODS	68
1.6.1. MCR-ALS	68
1.6.2. PCA	71
1.6.3. PLS-DA	71
CHAPTER 2. TARGETED ANALYSIS OF PROTEIN BIOMARKERS. TOP-DOWN PROTEOMICS	75
CHAPTER 3. TARGETED ANALYSIS OF PROTEIN BIOMARKERS. BOTTOM-UP PROTEOMICS	127
CHAPTER 4. UNTARGETED ANALYSIS FOR METABOLOMICS STUDIES	201
CAPÍTULO 5. RESULTADOS Y DISCUSIÓN	227

5.1. ANÁLISIS DIRIGIDO DE BIOMARCADORES PROTEICOS. LA TRANSTIRETINA EN LA POLINEUROPATÍA AMILOIDÓTICA FAMILIAR TIPO I	229
5.1.1. ANÁLISIS DE PATRONES DE TTR POR CE-MS Y CAPLC-MS	230
5.1.2. ANÁLISIS DE TTR EN MUESTRAS DE SUERO HUMANO POR CE-MS Y CAPLC-MS	240
5.1.2.1. IP convencional en solución	240
5.1.2.2. IP con partículas magnéticas	245
5.1.3. ESTUDIO DE LA ESTRUCTURA NATIVA DE LA TTR POR IM-MS	250
5.1.3.1. Análisis de patrones de TTR por IM-MS	251
5.1.3.2. Optimización de un método de IP <i>off-line</i> en condiciones no desnaturalizantes	255
5.1.3.3. Análisis de TTR en muestras de suero humano por IM-MS	257
5.1.4. ANÁLISIS DE TTR POR IA-SPE-CE-MS	263
5.1.4.1. Construcción del preconcentrador	264
5.1.4.2. Análisis de patrones de TTR por IA-SPE-CE-MS	267
5.1.4.3. Análisis de TTR en muestras de suero humano por IA-SPE-CE-MS	274
5.2. ANÁLISIS DIRIGIDO DE BIOMARCADORES PROTEICOS. ANÁLISIS <i>BOTTOM-UP</i> DEL ANTÍGENO CARCINOEMBRIÓNARIO HUMANO	280
5.2.1. ANÁLISIS DE LOS <i>N</i> -GLICOPÉPTIDOS DE CEA MEDIANTE <i>SHEATHLESS</i> CE-MS/MS	281
5.2.1.1. Interpretación de los electroferogramas y los espectros de MS/MS	282
5.2.1.2. Identificación de los <i>N</i> -glicopéptidos de CEA	285
5.2.1.3. Modelos semiempíricos de predicción de la migración electroforética	292
5.2.2. ANÁLISIS MULTIVARIANTE	293
5.2.2.1. PCA	294
5.2.2.2. PLS-DA	295
5.3. ANÁLISIS NO DIRIGIDO PARA LA IDENTIFICACIÓN DE BIOMARCADORES METABÓLICOS (<i>UNTARGETED ANALYSIS</i>). APLICACIÓN A LA ENFERMEDAD DEL HUNTINGTON	301
5.3.1. OPTIMIZACIÓN DE UN TRATAMIENTO DE MUESTRA DE PLASMA HUMANO PREVIO AL ANÁLISIS POR C ₁₈ -SPE-CE-MS	302
5.3.2. ANÁLISIS DE COMPUESTOS DE BAJA MASA MOLECULAR EN PLASMA DE RATONES WT Y HD POR C ₁₈ -SPE-CE-MS	309
5.3.3. HERRAMIENTAS QUIMIOMÉTRICAS PARA LA DETECCIÓN E IDENTIFICACIÓN DE BIOMARCADORES METABÓLICOS DEL HUNTINGTON	312
5.3.3.1. MCR-ALS	312
5.3.3.2. Análisis multivariante. PCA y PLS-DA	314
5.3.4. IDENTIFICACIÓN DE BIOMARCADORES METABÓLICOS Y RUTAS METABÓLICAS ASOCIADAS AL HUNTINGTON	316
CONCLUSIONS	329
REFERENCES	335

Abstract

In recent years, an increased emphasis has been placed on the detection, characterization and quantification of proteins, peptides and other metabolites that may be able to act as disease biomarkers in biological fluids. This is particularly the case of biomarkers for amyloidotic neurodegenerative diseases, in which some organ functions are destroyed as a result of deposits of normally soluble proteins that form stable insoluble amyloid fibrils. The mechanism involving amyloid formation remains relatively unknown. Therefore, clarification of the mechanisms underlying aggregation of these biomarker proteins is crucial to further our understanding of the onset of amyloidotic neurodegenerative diseases. Similarly, biomarker discovery in cancer is also a challenge that currently requires a great deal of effort. Even though an increasing number of potentially curative treatments including surgical procedures, chemotherapy, and molecular targeting therapies have been developed for several types of cancer, in most cases, early diagnosis, disease mechanism elucidation and clinical prognosis still remain uncertain. In this regard, cancer biomarkers are being investigated for their potential in clinical use and their value in the development of future treatments.

In this thesis, we have addressed the detection and characterization of amyloidotic neurodegenerative and cancer biomarkers from different proteomics and metabolomics perspectives, depending on the analytical and the sample preparation strategies employed: (i) **targeted** or **untargeted** strategies and (ii) **top-down** or **bottom-up** strategies (only in the case of proteomics analysis).

Research progress on purification methods, as well as the introduction of highly sensitive and selective analytical methods, in particular high-performance separation techniques, such as liquid chromatography (LC) and capillary electrophoresis (CE), coupled on-line with mass spectrometry (LC-MS and CE-MS, respectively), are crucial for the reliable identification and characterization of disease biomarkers in complex biological samples. Specifically, CE-MS is widely used for the highly efficient separation and characterization of a wide range of analytes, from low molecular mass compounds (e.g. metabolites) to large biomolecules (e.g. proteins and

glycoproteins). In addition to the high resolution capabilities, CE has many other advantages. For instance, the separation mechanism based on electrophoretic mobility, which is related to the ion charge-to-radius ratio, is complementary to the separation mechanism based on hydrophobicity of reversed phase liquid chromatography (RP-LC). In the present thesis, CE-MS has been extensively applied for the study of amyloidotic neurodegenerative diseases, in particular, familial amyloidotic polyneuropathy type I (FAP-I) and Huntington's disease (HD), as well as for the identification of colorectal cancer (CRC) and liver metastatic biomarkers.

Transthyretin (TTR) is a homotetrameric protein known to misfold and aggregate causing different types of amyloidosis. FAP-I, which is the most common hereditary systemic amyloidosis, is associated with a TTR variant that presents a single amino acid substitution of valine for methionine at position 30 of the monomeric sequence (Met30). In this thesis, we have used a **targeted top-down** strategy for the analysis of TTR, hence allowing the detection of intact TTR proteoforms. In this regard, we first developed a novel CE-MS method for the detection and characterization of normal and variant monomeric proteoforms of TTR. In this case, we applied an off-line conventional immunoprecipitation (IP) in solution under denaturing conditions of serum samples from healthy controls and FAP-I patients. Later, a novel off-line sample pretreatment based on IP using magnetic beads (MBs) was also proposed in order to improve protein recovery, reproducibility and minimize unspecific binding of other proteins. The IP method based on MBs allowed detecting monomeric proteoforms found at lower concentrations that were not possible to analyze by conventional IP in solution. In addition, the performance of capillary liquid chromatography mass spectrometry (CapLC-MS) for the analysis of serum TTR was compared with CE-MS. Both techniques gave similar results and allowed detecting the mutant TTR(Met30) proteoforms in FAP-I patients and, in general, an increased amount of oxidized TTR proteoforms.

Unfortunately, despite the usefulness of the proposed methodologies to screen for individuals with suspected TTR amyloidosis, CE-MS and CapLC-MS allowed only analyzing TTR under

denaturing conditions, and no information was obtained about the native tetrameric TTR or other oligomeric intermediates that may be involved in triggering protein aggregation. To overcome this major drawback, we developed a novel ion mobility mass spectrometry (IM-MS) method for the characterization of serum TTR after off-line IP under non-denaturing conditions. The proposed IM-MS methodology allowed the separation and characterization of tetrameric, trimeric and dimeric TTR gas ions due to their differential drift times. In addition, a comparison between tetramer (TE) and dimer (DI) ratios in serum samples from healthy controls and FAP-I patients demonstrated the presence of a higher abundance and stability of the TE in FAP-I patients.

On-line solid-phase extraction capillary electrophoresis mass spectrometry (SPE-CE-MS) was also evaluated as a powerful approach to improve the performance of CE-MS for the analysis of TTR. The most typical configuration for SPE-CE includes a microcartridge placed near the inlet of the separation capillary, which contains an appropriate packed sorbent to selectively retain the target analytes. In this way, it is possible to introduce large volumes of sample and the retained analytes are eluted in a smaller volume of an appropriate solution, which results in sample clean-up and concentration enhancement with minimum sample handling. The major drawback of the typical chromatographic sorbents used in SPE (e.g. C₈, C₁₈, polymeric with hydrophilic-lipophilic balance, etc.) is their limited selectivity, which hinders the analysis of complex samples such as biological fluids. In this regard, immunoaffinity (IA) sorbents are an interesting alternative with improved selectivity, which may also provide excellent extraction efficiency if the immunoreactivity and orientation of the antibody are optimum. In this thesis, we developed a method for purification, separation and characterization of TTR from serum samples by IA-SPE-CE-MS using an IA sorbent prepared by covalent attachment of antibody fragments (antigen binding fragment (Fab')) against TTR to succinimidyl silica particles. Under optimized conditions with TTR standard solutions, repeatability and reproducibility were acceptable, microcartridges lifetime was good (>20 analyses in consecutive days), limits of detection (LODs) were around 0.5 µg/mL (50-fold lower than by CE-MS, ~25 µg/mL) and

different TTR conformers were detected (folded and unfolded). In addition, the proposed methodology was applied to the analysis of serum samples from healthy controls and FAP-I patients to demonstrate its potential for the diagnosis of FAP-I.

HD was also subject of study in this thesis. It is an inherited neurodegenerative disorder, which is characterized by progressive motor and cognitive disturbances. HD is caused by an expansion of the cytosine-adenine-guanine (CAG) repeat in the exon 1 of the huntingtin gene (HTT), which encodes a stretch of glutamines in the huntingtin protein that make it prone to form amyloid fibrils. However, the mechanisms involving aggregation onset or progression are not fully understand yet. Therefore, studies targeting metabolism as an alternative to the huntingtin protein are regarded as necessary. Despite some potential low molecular mass biomarkers have been previously described in the brain of HD mice, attempts to find such metabolites in blood plasma have hitherto been unsuccessful. In the present thesis, we have used an **untargeted metabolomics** strategy for the identification of HD biomarkers in mice plasma. In this regard, different plasma pretreatments were evaluated to increase the recovery of low molecular mass compounds prior to SPE-CE-MS (with a C₁₈ sorbent). The optimized sample pretreatment was then applied to plasma samples from wild-type (wt) and HD mice of different ages (8, 12 and 30 weeks; early, middle and late disease stage, respectively). The proposed C₁₈-SPE-CE-MS methodology demonstrated to be suitable to ensure a reliable metabolite profiling of plasma samples. In addition, advanced chemometric data analysis tools were also investigated for data exploration and classification of the complex and large data sets generated in the study. The combination of multivariate curve resolution alternating least squares (MCR-ALS) with other chemometric tools, such as partial least squares discriminant analysis (PLS-DA), allowed the comprehensive analysis of the C₁₈-SPE-CE-MS metabolomics data, resolving electrophoretic peaks and mass spectra of a large number of low molecular mass compounds. The list of potential metabolites useful to distinguish between control and HD plasma samples were tentatively identified using on-line databases search, and the most affected metabolic pathways were discussed. In particular, the intracellular signaling pathway was found to be the most

altered pathway in HD, agreeing with dysfunction of specific neurons and changed expression of neurotransmitters.

CRC and liver metastatic disease were also studied in the present thesis. Specifically, human carcinoembryonic antigen (CEA) is a highly *N*-glycosylated protein (60% m/m of glycans and 28 *N*-linked glycosylation sites) found in normal human colonic epithelial cells, as well as in tumor forming CRC cell lines. Despite the demonstrated usefulness of CEA as a biomarker for monitoring CRC and liver metastatic disease patients' therapy, elevated CEA levels can also be detected in other neoplasms and in inflammatory conditions. Therefore, improvement in the selectivity of CEA detection remains a challenge for clinical tumor diagnosis. In this case, we have used a **targeted bottom-up** strategy for the analysis of the *N*-glycopeptides of CEA, hence allowing an indirect characterization of the intact protein glycoforms through the glycopeptide fragments derived from proteolytic digestion. CEA samples purified from human colon carcinoma and liver metastatic disease were digested and the enzymatic hydrolysates subsequently analyzed by sheathless capillary electrophoresis tandem mass spectrometry (sheathless CE-MS/MS). The information obtained using a combination of specific enzymes, such as trypsin and endoproteinase Glu-C, and non-specific enzymes, such as pronase, allowed the detection of most of the potential *N*-glycosylation sites (20 out of 28), as well as their degree of occupancy and their site-specific dominant *N*-glycopeptide type (219 different *N*-glycopeptide glycoforms were identified). Furthermore, a targeted multivariate data analysis approach using PLS-DA was also applied to easily differentiate the complex and characteristic glycopeptide maps of CEA samples purified from human colon carcinoma and liver metastatic disease, observing that samples obtained from liver metastases were, in general, more fucosylated and sialylated.

Aims

The present thesis is focused on the separation, detection and characterization of proteins, glycoproteins and metabolites in biological fluids that may be able to act as disease biomarkers. Specifically, it addresses the analysis of proteins such as transthyretin (TTR), an oligomeric protein related to familial amyloidotic polyneuropathy type-I (FAP-I), and human carcinoembryonic antigen (CEA), a tumor marker for colorectal cancer (CRC) and liver metastatic disease. The analysis of metabolite biomarkers is investigated in Huntington's disease (HD), another inherited amyloidotic neurodegenerative disease. In order to purify the target analytes, different strategies are developed for serum and plasma pretreatment. Furthermore, different microscale separation techniques, such as capillary electrophoresis (CE), capillary liquid chromatography (CapLC) and ion mobility spectrometry (IMS), all of them hyphenated to mass spectrometry (CE-MS, CapLC-MS and IM-MS, respectively) are evaluated for the analysis of different proteins and metabolites of interest following targeted or untargeted analytical strategies.

In order to achieve these main objectives, the work plan has been the following:

- In FAP-I:

- Optimization of off-line immunoprecipitation (IP) procedures (conventional IP in solution and IP with magnetic beads (MBs)) in order to purify TTR from human serum under denaturing and non-denaturing conditions.
- Development, optimization and validation of targeted top-down methodologies based on CE-MS, CapLC-MS and IM-MS for the analysis of TTR in serum samples from healthy controls and FAP-I patients. Characterization of TTR isoforms and post-translational modifications (PTMs) and TTR oligomers in order to distinguish between healthy controls and FAP-I patients.

- Preparation of an immunoaffinity (IA) sorbent using antibody fragments (antigen binding fragment (Fab')) for the purification, preconcentration and characterization of human serum TTR by on-line immunoaffinity solid-phase extraction capillary electrophoresis mass spectrometry (IA-SPE-CE-MS).

- In CRC and liver metastatic disease:

- Optimization of digestion procedures with specific proteases (i.e. trypsin and endoproteinase Glu-C) and non-specific proteases (i.e. pronase) in order to enzymatically digest CEA samples purified from human colon carcinoma and liver metastatic disease.
- Development of a targeted bottom-up strategy based on sheathless capillary electrophoresis tandem mass spectrometry (sheathless CE-MS/MS) for the analysis of CEA digests. Identification and characterization of CEA *N*-glycopeptide glycoforms.
- Application of chemometric data analysis tools, such as principal component analysis (PCA) and partial least squares discriminant analysis (PLS-DA), for differentiation of CEA samples purified from human colon carcinoma and liver metastatic disease.

- In HD:

- Optimization of a human sample pretreatment for the analysis of low molecular mass compounds before solid-phase extraction coupled on-line to capillary electrophoresis mass spectrometry with a C₁₈ sorbent (C₁₈-SPE-CE-MS).
- Development of an untargeted metabolomics strategy based on C₁₈-SPE-CE-MS for the identification of HD metabolite biomarkers in mice plasma and for the tentative identification of metabolic pathways involved in HD early diagnosis and progression.

- Application of advanced chemometric data analysis tools, such as multivariate curve resolution alternating least squares (MCR-ALS) and PLS-DA, for a comprehensive analysis of the C₁₈-SPE-CE-MS metabolomics data, as well as for differentiation between wild-type (wt) and HD mice of different ages (8, 12 and 30 weeks; early, middle and late disease stage, respectively).

Chapter 1. Introduction

1.1. Omics sciences

The term omics refers to any analytical study that provides collective information on a biological system. Conventional omics sciences include genomics, transcriptomics, proteomics and metabolomics (**Figure 1.1**), and new omics are constantly being added, such as lipidomics or glycomics [1,2]. Each omics science is crucial for understanding a certain biological system and complements the information provided by the other approaches. Furthermore, omics sciences can be applied not only for the greater knowledge of physiological processes, but also for screening, diagnosis and prognosis of disease pathogenesis [3–5].

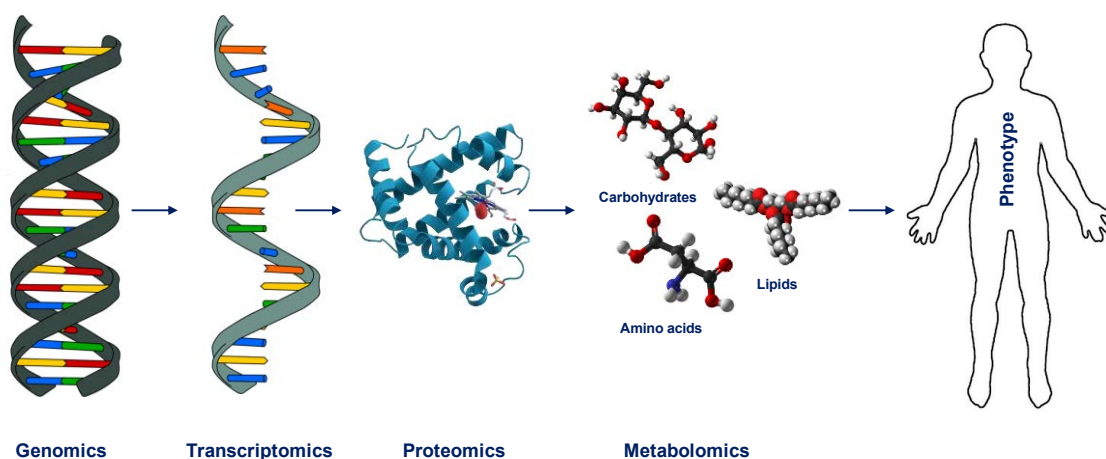


Figure 1.1. Conventional omics sciences.

The first omics sciences to come to life were genomics and transcriptomics. Genomics involves the study of the structure, function and expression of all the genes in an organism. It is specifically aimed to understand the structure of the genome, including gene mapping and deoxyribonucleic acid (DNA) sequencing, as well as to explore the molecular mechanisms or the interplay of genetic and environmental factors in an organism [6,7]. In contrast, transcriptomics involves the study of transcriptomes, i.e., the complete set of ribonucleic acid (RNA) transcripts produced by the genome at any one time. It is focused on how transcript

patterns are affected by development, disease, or environmental factors such as hormones and drugs, among others [8,9].

Together with the great advancement in analytical sciences and new instrumentation, there has been a considerable explosion of new omics sciences in the last two decades, being proteomics and metabolomics two of the most important ones, mainly due to their ability to provide integrative information on biological functions and define the phenotypes of biological systems in response to genetic or environmental changes.

1.1.1. Proteomics

Proteomics involves the study of proteoforms (isoforms, post-translational modifications (PTMs) and multiproteoform complexes), including their identification, characterization and quantification in a variety of contexts [10,11]. The proteomics nomenclature followed throughout this thesis is summarized in **Figure 1.2** [11,12].


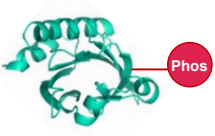

PROTEOFORMS			
	Isoform	Post-translational modification (PTM)	Multiproteoform complex
Diagram			
Description	Specific base primary amino acid sequence (includes genetic variations and point mutations)	Specific molecular form of a protein produced from a specific gene (includes phosphorylation, glycosylation, etc.)	A protein complex comprising specific proteoforms

Figure 1.2. Nomenclature accepted in proteomics: proteoform, isoform, post-translational modification (PTM) and multiproteoform complex. Abbreviations: **Phos**, phosphorylation.

Specifically, proteomics offers complementary information to genomics and transcriptomics essential for molecular level understanding of complex biological processes. Although changes in protein abundance and function can be attributed to altered gene expression, multiple post-transcriptional, co-translational and post-translational mechanisms can also affect [13,14]. In particular, PTMs greatly contribute to the much larger diversity of proteins than genes. Protein PTMs increase the functional diversity of the proteome by the covalent addition of functional groups or proteins, proteolytic cleavage of regulatory subunits or degradation of entire proteins. These modifications basically include phosphorylation, glycosylation, ubiquitination, nitrosylation, methylation, acetylation, lipidation and proteolysis, and influence almost all aspects of normal cell biology and pathogenesis [13,14]. Therefore, identifying and understanding PTMs is critical in the study of cell biology and disease treatment and prevention. In this regard, mass spectrometry (MS) and hyphenated separation techniques (i.e. liquid chromatography mass spectrometry (LC-MS) and capillary electrophoresis mass spectrometry (CE-MS)) have emerged as the most important tools to identify and characterize proteins and their PTMs with high throughput and on a large scale [15,16]. However, the goal of profiling proteins represents an analytical challenge, mainly due to the extreme complexity typically present in proteomics samples and the wide range of relative abundances of the different proteins (i.e. dynamic range), which are also structurally very complex.

There are two main approaches in proteomics analysis differentiated by the sample preparation methods (**Figure 1.3**) [10,17]. The most common approach is **bottom-up proteomics**, in which protein samples are enzymatically digested (most frequently with trypsin) into peptides prior to MS analysis, hence providing an indirect analysis of the intact proteins through peptides derived from proteolytic digestion. In a typical bottom-up experiment, the peptide mixture is fractionated, purified and subjected to LC or CE tandem mass spectrometry (LC-MS/MS or CE-MS/MS) analysis. Peptide identification is achieved afterwards by comparing the MS/MS spectra derived from peptide fragmentation with theoretical MS/MS spectra generated from *in silico* digestion, which are stored in a protein database [18–20]. However, despite peptides are

easier to handle because they are readily solubilized and separated prior to MS, several disadvantages to this approach become clear. A peptide or even several peptides may not be specific to an individual protein. In addition, large regions of the protein may not be identified, which can leave behind important information regarding PTMs. In this regard, **top-down proteomics** seeks to overcome these major drawbacks by analyzing the intact protein and, if possible, its fragment ion masses by MS/MS. This approach theoretically allows for 100% sequence coverage and full characterization of PTMs and isoforms [11,21–24]. However, the difficulty in the ionization of large molecules such as intact proteins has caused top-down proteomics to lag behind bottom-up proteomics in terms of proteome coverage, sensitivity and throughput. Anyway, top-down proteomics is regarded as a powerful alternative to digestion-based approaches and has been benefited from the recent advances in high-performance separation techniques, MS instrumentation and customized bioinformatic tools.

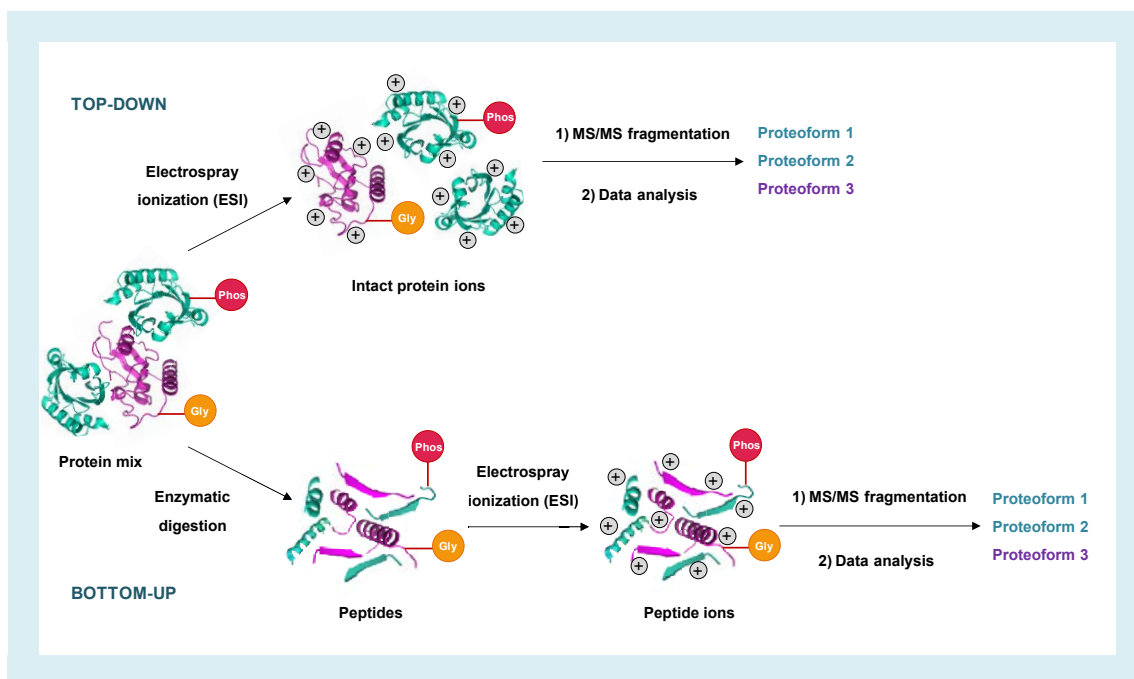


Figure 1.3. Top-down and bottom-up proteomics based on MS/MS analysis. Abbreviations: **Gly**, glycosylation; **Phos**, phosphorylation.

1.1.2. Metabolomics

Unlike proteomics analysis, which is commonly complicated by regulatory processes such as PTMs, metabolite profiles provide direct signatures of cellular activity and are, therefore, easier to correlate with the phenotype. In this regard, metabolomics has rapidly become a profiling technique of choice for biomarker elucidation and molecular diagnostics [1,25,26]. Metabolites are typically low molecular mass compounds (relative molecular mass (M_r)<1,500) and play key roles in biology as signaling molecules, energy sources and metabolic intermediates. The metabolome contains the smallest domain (~5,000 metabolites), but metabolites show broad variations in physicochemical properties (such as polarity, acidity, solubility, etc.), hence making difficult the simultaneous analysis of a wide range of them [27]. Recently, MS-based metabolomics has become the most popular metabolomics strategy. Specifically, metabolomics studies have been performed using various hyphenated analytical techniques such as gas chromatography mass spectrometry (GC-MS) [28,29], LC-MS [30,31] and CE-MS [32,33]. GC-MS is usually the preferred technique for small and volatile metabolites, including steroids and fatty acids, and chemical derivatization is often necessary to increase their volatility. LC-MS is also commonly used as it allows the detection of a wider range of metabolites, regardless of their hydrophilic or hydrophobic nature. Finally, CE-MS can also be used for metabolomics studies and, as it presents a separation mechanism based on the molecular charge-to-radius ratio, it is the most appropriate for the analysis of charged metabolites.

1.1.3. Targeted and untargeted analysis

Both proteomics and metabolomics can be approached using **targeted** or **untargeted** analysis [34]. This thesis will present studies based on both analytical strategies for the analysis of protein and metabolite biomarkers in biological fluids.

Targeted analysis provides information for a predefined list of proteins or metabolites, and the selected analytical method can be established using commercially available standards or chemically synthesized compounds [35–37]. Targeted approaches offer better sensitivity than untargeted approaches, as well as reliable results. In this regard, targeted strategies are especially preferable for the analysis of proteins and metabolites found at low concentrations, which often play important roles in biological systems [38,39]. Different techniques can be applied for targeted analysis depending on the physicochemical properties of the proteins or metabolites being investigated.

Untargeted analysis or global profiling measures simultaneously as many proteins or metabolites as possible in order to obtain a comprehensive profile or fingerprint of a biological sample. Untargeted approaches can implicate previously unrecognized proteins or metabolites and, therefore, they are a powerful strategy to elucidate novel biomarkers and gain insight into disease pathogenesis [40,41]. However, the most difficult tasks in untargeted approaches are the simultaneous analysis of a broad range of compounds with different properties and concentrations, as well as the identification of unknowns. Furthermore, it is practically impossible to detect all the proteins or metabolites present in a biological sample with a single analysis. In this regard, multiple analytical platforms such as LC-MS and CE-MS have been simultaneously used for global proteome and metabolome profiling [42–45]. The data acquired by untargeted analysis consists of a list of the mass-to-charge ratios (m/z) of the detected proteins, peptides or metabolites and their retention or migration times, for LC-MS and CE-MS, respectively. To identify the unknown compounds, the M_r provided by an accurate mass and high-resolution mass spectrometer can be used to find matches in proteomics and metabolomics databases, such as the Human Protein Reference Database [46] or the Human Metabolome Database (HMDB) [47], respectively. The identified proteins and metabolites can be subjected afterwards to biological pathway analysis to find relationships with specific disorders or physiological characteristics using web databases such as the Kyoto Encyclopedia of Genes and Genomes (KEGG) [48], the MetaCyc and the BioCyc [49,50].

1.2. Biomarker discovery

It is widely accepted that a biomarker is “a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes or pharmacologic responses to a therapeutic intervention” [51,52]. Over the past decades, there has been a growing interest in applying proteomics and metabolomics strategies to search for novel biomarkers related to neurodegenerative diseases [53–55] and cancer [56–58]. The goal of current research is to identify biomarkers that could allow a non-invasive and cost-effective diagnosis, as well as to recognize the best prognostic panel and to define reliable predictive biomarkers for the available treatments. This thesis in particular will present the use of proteomics and metabolomics strategies for the identification of protein and metabolite biomarkers related to amyloidotic neurodegenerative diseases, colorectal cancer (CRC) and liver metastatic disease.

1.2.1. Amyloidotic neurodegenerative diseases

Neurodegenerative diseases, such as Alzheimer’s disease (AD), Parkinson’s disease (PD), Huntington’s disease (HD), amyotrophic lateral sclerosis (ALS) and familial amyloidotic polyneuropathy (FAP), are disorders caused by the deterioration of certain nerve cells [59,60]. There are no specific cures for these diseases because the neurons of the central nervous system cannot regenerate after cell death or damage. Nowadays, diagnosis of amyloidotic neurodegenerative diseases is primarily made on clinical grounds including neuropsychological testing and brain imaging, which often lead to misdiagnosis or late diagnosis. Therefore, there is an urgent need to develop a reliable approach for early diagnosis of amyloidotic neurodegenerative disorders, as well as to find new therapeutical targets or to monitor responses of the patients to new therapies. Over the past decades, many investigations have revealed the existence of a common pathogenic mechanism associated with the above mentioned

amyloidotic neurodegenerative disorders, i.e., the aggregation of misfolded proteins and their deposition as insoluble amyloid aggregates in different regions of the brain, thus leading to amyloidosis of the central nervous system. Although the detailed mechanism for the formation of amyloid fibrils remains relatively unclear, the initiating event seems to be protein misfolding, which results in the formation of aggregation-prone structures that oligomerize and grow by an autocatalytic mechanism [59]. There are several proteins associated to amyloid processes leading to high social impact amyloidotic neurodegenerative diseases [61]. **Table 1.1** shows the relationship between different biomarker proteins and the most important amyloidotic neurodegenerative diseases, as well as a few differential features.

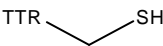
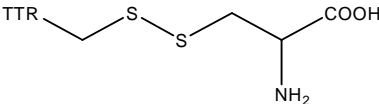
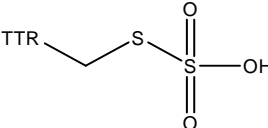
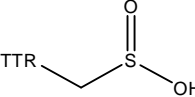
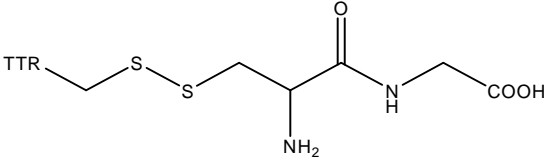
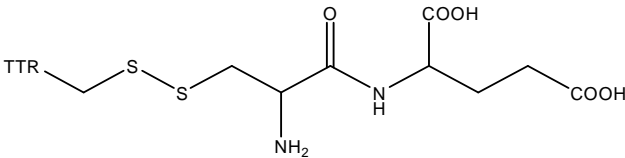
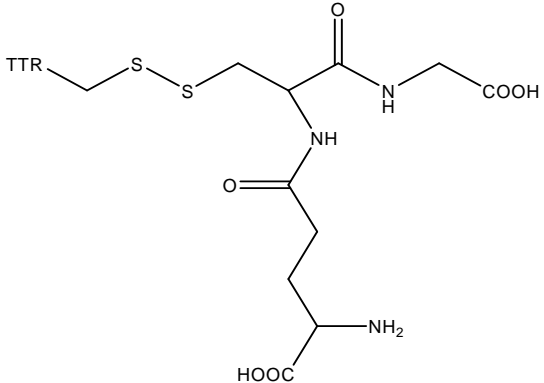
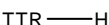
Table 1.1. Biomarker proteins and amyloidotic neurodegenerative diseases. Abbreviations: **CJD**, Creutzfeldt-Jakob disease; **GSSD**, Gerstmann-Straussler-Scheinker disease; **FFI**, fatal familial insomnia.

Biomarker proteins and precursors				
Protein	Precursor	Distribution	Type	Syndrome
Aβ	Amyloid β -protein precursor	Localized	Acquired	Sporadic Alzheimer's disease
		Localized	Hereditary	Hereditary cerebral amyloid angiopathy
PrP	Prion protein	Localized	Acquired	Sporadic (iatrogenic) CJD
		Localized	Hereditary	Familial CJD, GSSD, FFI
ABri	Amyloid Bri protein precursor	Localized or systemic	Hereditary	Familial dementia
Cys	Cystatin C	Systemic	Hereditary	Hereditary cerebral amyloid angiopathy
TTR	Transthyretin	Systemic	Hereditary	Hereditary TTR amyloidosis. Familial amyloidotic polyneuropathy
		Systemic	Acquired	Senile systemic amyloidosis
Gel	Gelsolin	Systemic	Hereditary	Hereditary amyloidosis
HTT	Huntingtin	Localized	Hereditary	Huntington's disease
SNCA	α -Synuclein	Localized	Acquired	Parkinson's disease
SOD-I	Superoxide dismutase I	Localized	Acquired	Amyotrophic lateral sclerosis

1.2.1.1. Transthyretin (TTR). Hereditary TTR amyloidoses

Native transthyretin (TTR) is a homotetramer composed of four identical, non-covalently associated subunits. Each monomer consists of 127 amino acid residues ($M_r \sim 14,000$). TTR is predominantly expressed in the liver and in the choroid plexus of the brain. It binds to thyroid hormones and transports thyroxine and retinol (associated with the retinol-binding protein). This protein is highly abundant in serum (200-400 μg TTR/mL) and cerebrospinal fluid (CSF) (10-40 μg TTR/mL) [62,63]. Today, about 100 point mutations are known in the TTR gene. Several of the resulting genetic TTR variants are harmless, but many of them play crucial roles in different types of hereditary TTR amyloidoses [64]. In addition to mutations at the protein sequence level (namely isoforms), this protein exists in several proteoforms with different PTMs. Specifically, each monomer contains a cysteine residue at position 10 (Cys10), which explains most of the microheterogeneity of TTR. The most common PTMs related to Cys10 are S-cysteinylation (TTR-Cys), S-sulfonation (TTR-Sulfonated), S-glycinylcysteinylation (TTR-CysGly), S-glutamylcysteinylation (TTR-CysGlu) and S-glutathionylation (TTR-Glutathione) [65]. Furthermore, TTR Cys10 may additionally be oxidized to glycine ((10) C-G) or cysteine sulfinic acid (TTR-Sulfinic) (see **Table 1.2**). These Cys10 proteoforms are believed to reflect the redox balance of the *in vivo* environment, which may suggest that TTR participates in the defense against oxidative stress through thiol conjugation. In addition, it is thought that PTMs at Cys10 could play an important role in the onset and pathological processes of TTR related amyloidoses, although their implications are still relatively unknown [66,67]. Apart from the Cys10 residue, TTR features additional sites for PTMs, such as phosphorylation and dehydroxylation (TTR-Phosphorylated and TTR-Dehydroxylated, respectively). However, the role of these specific PTMs in the mechanisms underlying TTR amyloidoses has not been reported in the literature.

Table 1.2. TTR modifications on Cys10. Chemical structure, theoretical M_r and theoretical calculated M_r variation (ΔM_r) taking as reference Free-TTR.

Modifications on Cys10			
Modification	Chemical structure	Theoretical M_r	Theoretical ΔM_r
Free-TTR (none)		13,761.2640	0
TTR-Cys		13,880.4022	+119.1382
TTR-Sulfonated		13,841.3283	+80.0643
TTR-Sulfinic		13,793.2628	+31.9988
TTR-CysGly		13,937.4590	+176.1950
TTR-CysGlu		14,009.5218	+248.2578
TTR-Glutathione		14,066.5732	+305.3092
(10) C-G		13,715.1713	-46.0927

There are several TTR amyloidoses namely familial amyloidotic polyneuropathy (FAP), familial amyloidotic cardiomyopathy (FAC), central nervous system selective amyloidosis (CNSA) and senile systemic amyloidosis (SSA). FAP, FAC and CNSA are hereditary [68], while SSA is related to the deposition of wild-type TTR and affects the myocardium of around 25% of the population older than 80 years [69]. The present thesis is focused on FAP and a deeper insight into this disease is going to be addressed.

FAP is an inherited severe systemic amyloidosis caused by a mutated TTR, and it is characterized by amyloid deposition mainly in the peripheral nervous system and the heart. The disease was first described by Andrade in 1952 [70], in the Portuguese city of Porto. However, other endemic areas have been described since then, mainly Sweden, Japan and Mallorca Island [71,72]. Among FAP, FAP type I Met30 (FAP-I, single amino acid substitution of valine for methionine at position 30 of the TTR monomeric sequence) is the most common [73,74]. The disease is characterized by a progressive painful peripheral neuropathy, although symptoms involving the gastrointestinal tract, heart, kidney, ocular tissues and autonomic nervous system are also present [73,74]. FAP-I is a fatal disease and liver transplantation is the most promising therapy today [75,76].

The mechanism by which TTR leads into fibril formation and aggregation remains still unknown. It is hypothesized that amyloid fibril formation might be triggered by tetramer dissociation into non-native monomers with low conformational stability [77]. These partially folded monomers would be highly prone to aggregate and form amyloid fibrils [68]. **Figure 1.4** shows a schematic representation of the proposed mechanism for TTR fibril formation.

Nowadays, the analysis of normal and variant proteoforms of TTR is based on clinical genetic testing, including targeted mutation analysis, sequence analysis of exons, and deletion/duplication analysis [73]. In the last years, a special effort has been placed on the analysis of TTR as a biomarker for FAP-I in biological fluids (specifically in serum and CSF) to

confirm diagnosis and clarify the aggregation mechanism. In particular, this thesis will describe new strategies for the analysis of TTR in human serum from FAP-I patients.

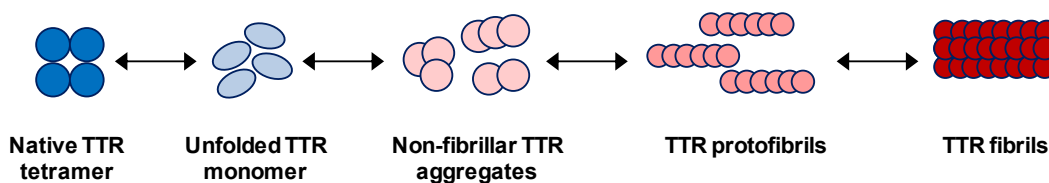


Figure 1.4. Schematic representation of the hypothetical pathway for fibril formation.

1.2.1.2. Huntington's disease

Huntington's disease (HD) is an inherited neurodegenerative disorder, which is characterized by progressive motor and cognitive disturbances [78,79]. HD is caused by an expansion of the cytosine-adenine-guanine (CAG) repeat in the exon 1 of the huntingtin gene (HTT). The presence of the CAG repeat in the HTT gene encodes an expanded polyglutamine (polyQ) domain in the N-terminal portion of the huntingtin protein, which induces conformational changes that result in inclusion formation [80]. Repeat lengths in the range between 36 and 39 are referred to as "reduced penetrance" alleles and may or may not produce a clinical HD syndrome during an individual's natural life [81]. In the expanded range (39 repeats and over), the length of the CAG repeat predicts about 50-70% of the age at clinical onset, with larger repeat lengths tending to result in earlier onset [82].

Although the HTT gene is expressed as the huntingtin protein in many tissues, HD pathology has primarily been located to the basal ganglia and to the neocortex. The pathology involves atrophy and dysfunction of specific neurons, loss of several types of receptors and changed expression of neurotransmitters and key proteins [78,79]. HD is a fatal disease, and the median interval between clinical diagnosis and death is typically given as 15–20 years [83].

By use of predictive genetic testing, it is possible to identify individuals who carry the HTT gene defect before the onset of symptoms, providing a unique window of opportunity for intervention aimed at preventing or delaying the disease [78,84]. However, without robust and practical measures of disease progression, the efficacy of therapeutic interventions in this premanifest HD cannot be readily assessed.

Neuroimaging and biochemical biomarkers are being investigated for their potential in clinical use and their value in the development of future treatments [78,84]. Modern neuroimaging techniques such as magnetic resonance imaging (MRI) enable obtaining high quality images of brain structure and function, while advanced image analysis techniques allow robust measurement of differences between subjects and change within individuals. As well as being relatively non-invasive, its clear relevance to neuropathology makes neuroimaging an appealing source of biomarkers [85,86]. However, sometimes they are not informative enough. Molecular biomarkers that can be quantified in biofluids such as blood or urine are appealing because of the minimal requirement for patient involvement and the availability of reliable and sensitive analytical methods. Approaches to the discovery of potential biomarkers have ranged from hypothesis-targeted approaches investigating pathways known to be altered in HD [87,88], to the application of untargeted strategies to find differences between healthy controls and HD patients [89,90].

The ideal peripheral blood biomarker of HD would be a direct product of neuronal dysfunction (normally absent in blood) that leaked across the blood-brain barrier and became detectable in blood. However, attempts to find such biomarkers have hitherto been unsuccessful. Most of the potential biomarkers studied to date are largely produced by peripheral tissues as a result of effects of the expressed mutant huntingtin. This thesis will focus on the identification of novel plasma biomarker candidates that may be directly related to brain-striatal effects involved in HD progression.

1.2.2. Cancer. Colorectal cancer and liver metastatic disease

Cancer is a leading cause of death, with estimations of mortality around 8 million worldwide in 2012 [91]. In particular, colorectal cancer (CRC) is the second most commonly diagnosed cancer among females and third among males worldwide [92,93]. Even though an increasing number of potentially curative treatments including surgical procedures, chemotherapy, and molecular targeting therapies have been developed for CRC, the clinical prognosis still remains unsatisfactory. Furthermore, in patients with CRC, the liver is the most commonly involved site of distant metastasis, with over half of all patients with CRC experiencing liver metastasis during their disease course [94,95]. At present, the only potentially curative treatment option for colorectal cancer liver metastasis is hepatic surgical resection, with a reported overall survival probability after resection of 25%–58% [94].

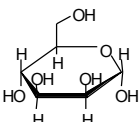

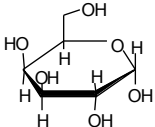

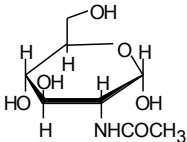

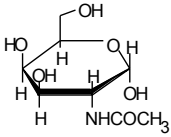

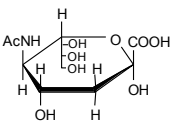

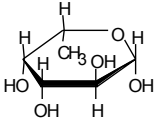

In the past few years, glycomics has been at the forefront of revolutionizing biological and medical sciences, holding out the promise of both fully understanding and effectively treating human diseases. Recent research in the glycomics field has given insight into the biological significance of the *N*-glycome in human health and disease. Special emphasis has been placed on exploring the connection between altered *N*-glycosylation and different diseases, particularly cancer [96–98]. Therefore, specific glycoforms of certain glycoproteins may serve as biomarkers for either the early diagnosis, prognosis or the evaluation of therapeutic efficacy for treatment diseases. Glycomics becomes, therefore, an emerging field that can make unique contributions to the discovery of cancer biomarkers.

1.2.2.1. Glycosylation

Glycosylation is defined as the enzymatic process that produces glycosidic linkages of carbohydrates (i.e. sugars) to other carbohydrates, proteins or lipids [99,100]. In particular, protein glycosylation is referred to the covalent attachment of carbohydrates, also known as

glycans, to the polypeptide chain of a protein [100,101]. Glycans can adopt several structures, from a simple monosaccharide unit to a branched polysaccharide. These glycan chains are composed of well-known monosaccharide units covalently attached to each other by glycosidic bonds [100,101]. **Table 1.3** shows the most common monosaccharide units that can be found in glycan structures, with the standardized symbol nomenclature proposed by the Consortium for Functional Glycomics (CFG) [102].

Table 1.3. Monosaccharide units found in glycan structures. Symbol nomenclature proposed by the Consortium for Functional Glycomics (CFG) [102].

Monosaccharide units			
Name	Structure	Nomenclature	Symbol
Mannose		H	
Galactose		H	
N-Acetylglucosamine		N	
N-Acetylgalactosamine		N	
N-Acetylneuraminic acid (sialic acid)		S	
Fucose		F	

Depending on the glycan attachment sites in the polypeptide chain, protein glycosylation can be divided into *O*- and *N*-linkage type [103]. *O*-linked glycosylation consists of attaching the glycans to a serine (Ser) or a threonine (Thr) amino acid in the polypeptide chain of a protein [103]. In *N*-linked glycosylation, glycans are attached to asparagines (Asn) in the sequence containing Asn-X-Ser/Thr, where X corresponds to any amino acid except proline (Pro) [103]. All *N*-glycans share a common structure, commonly known as core (**Figure 1.5**). However, several monosaccharide units can be attached to this core fragment, thus giving rise to the three main types of *N*-glycans: high mannose, complex and hybrid [104,105]. **Figure 1.5** shows the different types of *N*-glycans (following the symbols recommended by the CFG [102]).

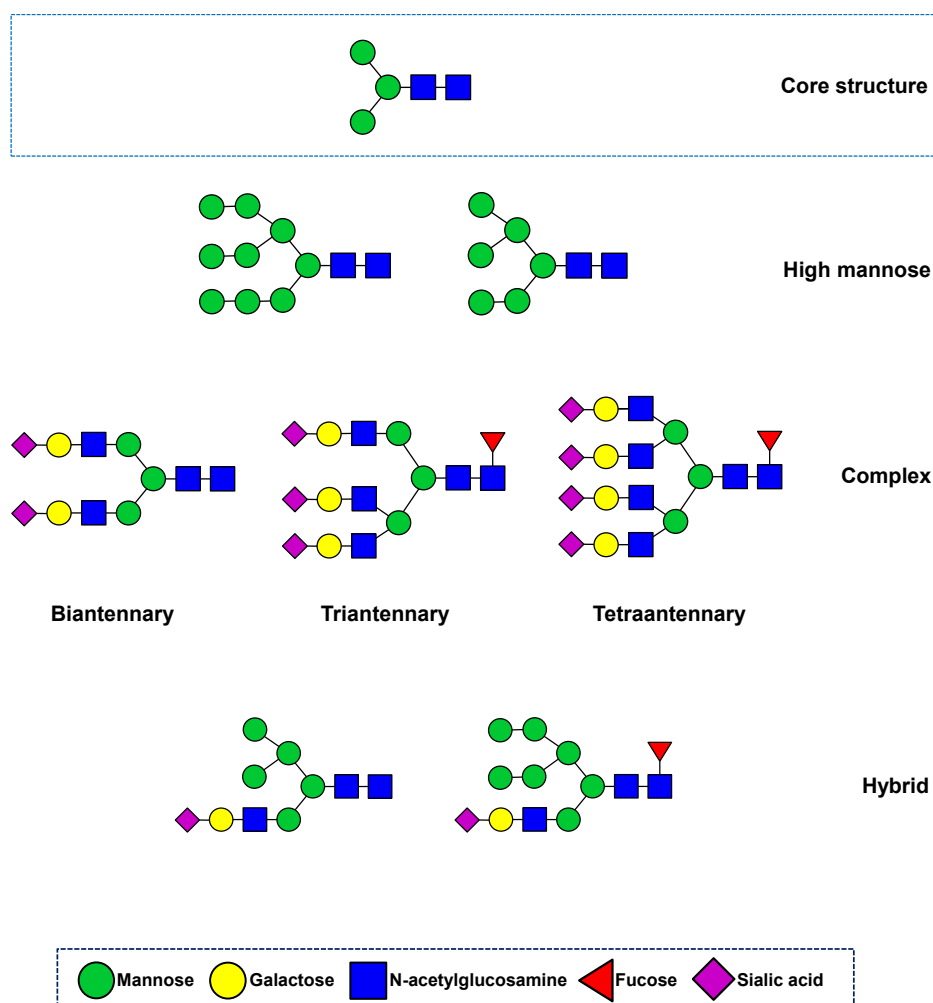


Figure 1.5. Types of *N*-glycans following the symbols nomenclature recommended by the Consortium for Functional Glycomics (CFG) [102].

- **High mannose glycans:** only mannose residues are attached to the core structure of the glycan.
- **Complex glycans:** N-acetylglucosamine residues are bound to the mannoses of the core, resulting in further branched structures, commonly denominated antennae (bi-, tri- and tetraantennary structures). These antennae are formed by N-acetyllactosamine residues (LacNAc: N-acetylglucosamine plus galactose) that may end with a sialic acid residue.
- **Hybrid glycans:** can be considered a combination of the other two types of *N*-glycans.

Different techniques are currently used for the analysis of protein glycosylation, including matrix assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF-MS) [106], or electrospray ionization mass spectrometry (ESI-MS), coupled on-line or not, with LC [107] and CE [108]. Unfortunately, despite the usefulness of these techniques for analyzing a wide range of glycoproteins, characterization of protein glycoforms becomes increasingly difficult when analyzing highly glycosylated proteins. In this regard, three approaches can be used for the study of protein glycosylation. In the top-down approach, the intact glycoprotein is analyzed, which is more direct and considerably faster than the other two alternatives, as barely any sample treatment is necessary. However, targeting the intact glycoprotein can be quite challenging due to different causes: the difficulty in the ionization of large molecules such as glycoproteins, and the limited resolving power of high-performance separation techniques and mass spectrometers to resolve complex mixtures of glycoforms. Moreover, information of glycosylation sites and their degree of occupancy cannot be obtained [109].

The other two approaches involve the use of enzymes or specific reagents to obtain glycans or glycopeptides of lower molecular mass. In one approach, glycans are enzymatically (with glycosidases) or chemically cleaved from the glycoprotein and subjected to purification prior analysis. However, the utility of this method is limited when more than one glycosylation site is present in the glycoprotein, because it fails to correlate glycan composition with the different

attachment sites [103]. This information can be obtained by following a typical bottom-up proteomics approach, which consists of analyzing glycopeptides after subjecting the glycoprotein to enzymatic proteolysis [110,111]. Proteases are enzymes that catabolize the hydrolysis of peptide bonds between amino acids in the polypeptide chain of a protein. There are several proteases, each one with their own efficiency and specificity, being trypsin widely employed in a multitude of studies [112,113]. Using glycopeptide analysis, the glycan composition can be correlated to the attachment site, due to the presence of the peptide portion on each glycopeptide. However, working with glycopeptides instead of glycans can be rather difficult, mainly due to the higher molecular mass of the glycopeptides, which often show worse ionization efficiencies, and the limited resolved power of the analytical techniques.

In conclusion, the three approaches are complementary as they have their own advantages and drawbacks, and must be taken into account in order to obtain comprehensive information to put together the glycoprotein structural puzzle.

1.2.2.2. Human carcinoembryonic antigen

Human carcinoembryonic antigen (CEA) is a highly *N*-glycosylated protein (60% m/m of glycans) found in normal human colonic epithelial cells, as well as in tumor forming and colonic adenocarcinogenic cell lines [114,115]. CEA was discovered almost 50 years ago, in 1965, and it still remains the only tumor marker of recognized efficacy in monitoring CRC patients' therapy [93,116]. For patients with liver metastatic disease, changes in CEA concentration during treatment can also indicate response to therapy [93]. Despite the reported benefits of using CEA as a biomarker for recurrence and response to treatment, there are some significant limitations. Although CEA was first considered specific for CRC, elevated CEA levels can also be observed in other neoplasms, e.g. gastric and pancreatic cancers, or in inflammatory conditions [93]. Additionally, CEA does not allow differentiation between benign

and malignant tumors [93]. Therefore, improvement in the selectivity of CEA detection remains a scientific challenge.

CEA ($M_r \sim 180,000$) contains 28 potential *N*-linked glycosylation sites [117]. **Figure 1.6** shows the amino acid sequence of CEA and its potential *N*-glycosylation sites marked in red. The carbohydrate fraction includes mannoses, galactoses, N-acetylglucosamines, N-acetylgalactosamines, fucoses and sialic acids (see **Table 1.3**).



Figure 1.6. Amino acid sequence of human carcinoembryonic antigen (CEA) and its potential 28 *N*-linked glycosylation sites marked in red.

The glycan composition of CEA displays a considerably heterogeneity in the sugar content, but detailed carbohydrate structures of the 28 *N*-glycosylation sites have so far not been reported. One earlier study (1983) identified the structures of CEA carbohydrates and proposed that it

contains approximately 40 different glycan compositions [118]. In a recent study (2015), 61 CEA glycans were determined in tumor tissues, and high mannose and complex-type carbohydrate chains were identified [117].

In contrast to these studies based on targeting glycans, this thesis has been focused on the glycopeptide site-specific analysis of CEA and the modifications of the *N*-glycosylation patterns induced during CRC and liver metastatic disease.

1.3. Purification techniques

In recent years, an increased emphasis has been placed on the separation, detection, characterization and quantification of normal and variant proteoforms of different proteins as predictive indicators of an ongoing disease state [119,120]. In this regard, research progress on protein isolation and purification has been crucial for an increase in the accuracy and reproducibility of protein profiling to identify PTMs and mutations at the protein sequence level [121,122]. Nowadays, affinity chromatography, size exclusion chromatography and ion exchange chromatography are very suitable techniques for the purification of a wide range of proteins from biological fluids and tissues [123–125]. However, these classical chromatographic procedures are often labor-intensive and time-consuming. In this regard, some authors have described several alternatives including liquid-liquid extraction, dialysis, centrifugal filtration and immunoprecipitation (IP) methods. Most of them have limited selectivity and provide low recoveries, with the exception of IP techniques, which take advantage of the high selectivity of the antibodies to isolate the antigenic target proteins out of complex samples. These IP methods can also be combined with the rest of purification techniques to get improved results.

1.3.1. Antibodies

Antibodies, also known as immunoglobulins (Ig), are glycosylated proteins exclusively synthesized by B cells whose main function is to defend the organism against external agents (referred to as antigens). There are different types of Ig, being IgG the most common one (approximately 70-75% of all the Ig in human serum are IgG). IgG share a common Y-shaped structure that consists of four polypeptide chains (**Figure 1.7**): two identical heavy (H) chains ($M_r \sim 50,000$) and two identical light (L) chains ($M_r \sim 25,000$). These four chains are joined together by disulfide bonds and non-covalent interactions, and can be divided into constant (C) and variable (V) domains. The C domains are located in the carboxyl terminal portion of the

antibody, whereas the V regions are located in the amino terminal portion [126,127]. There are also two distinguishable sides of the antibody: the upper arms, commonly referred to as the fragment antibody binding (Fab') region; and the lower part of the molecule, known as fragment crystallizable (Fc) region. The Fab' region can present several characteristic amino acid sequences depending on the target antigen that binds to the antibody. The Fc region only varies in its amino acid sequence if we consider a different Ig than IgG.

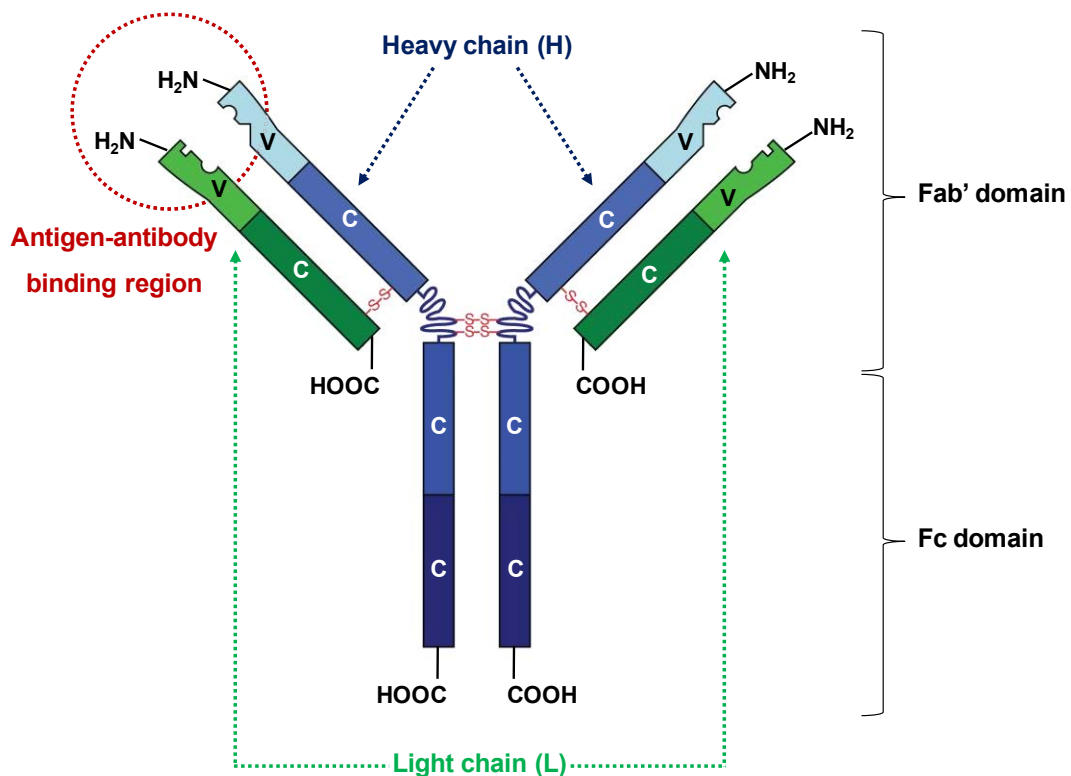


Figure 1.7. Structure of an immunoglobulin G (IgG).

Antibodies bind to a specific region of the antigen known as antigenic determinant or epitope. Depending on how the antibodies have been generated, they can be divided in two types, monoclonal or polyclonal. Monoclonal antibodies are those produced by a single line of clones of the same B cell and only recognize one epitope on an antigen [128]. In contrast, polyclonal antibodies are generated by multiple clones of different B cells, and recognize various epitopes on an antigen. The decision regarding whether to use a polyclonal or monoclonal antibody

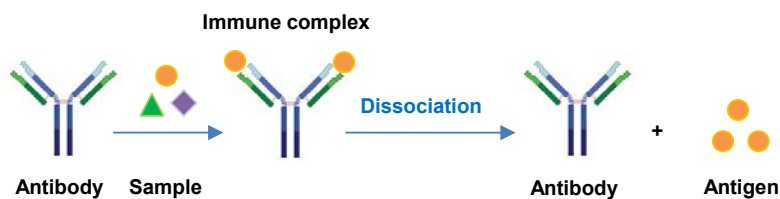
depends on the intended use and the availability. Polyclonal antibodies can be generated much more rapidly, are less expensive and require less technical skills. In contrast, monoclonal antibodies are more selective, a useful characteristic while working with complex samples. However, this high selectivity may represent a drawback in some cases, e.g. when focused on the analysis of a family of structurally related compounds. For instance, some changes in the structure of an epitope, such as genetic polymorphism, glycosylation or denaturation, may lead to poor or no recognition by the antibody. In contrast, these changes do not affect polyclonal antibodies, since they are heterogeneous and recognize a set of antigenic epitopes [129].

1.3.2. Conventional IP in solution and IP with magnetic beads

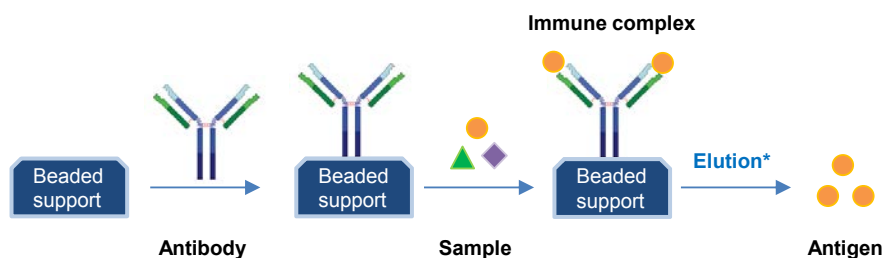
Two different off-line IP approaches will be discussed in this thesis: conventional IP in solution and IP with magnetic beads (MBs).

Conventional IP in solution consists of incubating a solution of the antibody with the sample containing the target protein (**Figure 1.8 A**). After centrifugation of the mixture, the supernatant is removed, hence allowing the collection of the immune complex (i.e. antigen-antibody complex) as a solid residue or precipitate. If necessary, this complex is later dissociated using an appropriate solution, in general, lowering the pH value, increasing the ionic strength and/or adding organic solvents. Conventional IP methods in solution require an optimum proportion between the molecules of the antigen and the antibody, commonly known as equivalence zone, in order to achieve maximum immunoprecipitation. The time course and recoveries vary dramatically depending on the affinity of the antigen, the size and number of epitopes and the medium conditions (e.g. the reaction temperature, ionic strength, pH), among other factors [130].

A) Conventional IP in solution



B) Pre-immobilized antibody approach



C) Free antibody approach

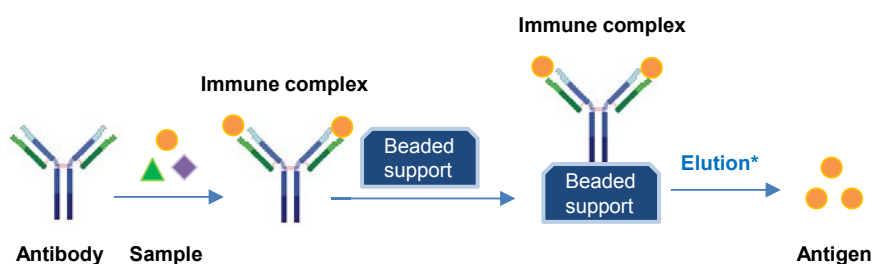


Figure 1.8. Schematic representation of (A) conventional IP in solution, (B) pre-immobilized antibody approach and (C) free antibody approach. *If the antibody is not covalently coupled to the MBs, it will be eluted with the antigen.

In recent years, the emergence and development of micro and nano MBs for analytical purposes have led to an evolution of the conventional IP methods in solution. Nowadays, modern **IP techniques with MBs** are widely applied to improve protein recovery, reproducibility and minimize unspecific binding of other proteins [131,132]. In general, antibodies are immobilized by strategies that bind the antibody to the beaded support through covalent bonds or non-

covalent interactions. Immobilized Protein A or Protein G crosslinked agarose beads, which bind the antibody through non-covalent interactions, are easier to handle and cheaper than silica beads. The decision to use Protein A or Protein G depends upon the source and class of the antibody used and the overall cost. Immobilized Protein A is generally less expensive than Protein G, however, Protein G binds a wider variety of antibody classes [133]. Additionally, orientation of the antibodies on the solid support is of great concern because the performance of the IP depends on the availability of the binding sites for the antigen-antibody interaction. Specifically, Protein A or Protein G crosslinked agarose MBs allow an optimum orientation of the antibody to the MBs through its Fc region. There are two different IP approaches depending on when the immobilization of the antibody to the beaded support takes place, either before or after antigen-antibody recognition [133]. In the first case, an antibody against a specific protein is pre-immobilized onto a beaded support and then incubated with the sample containing the target protein (**Figure 1.8 B**). The immobilized immune complex is then collected from the sample, eluted from the support and subsequently analyzed. In the second case, the free antibody forms the immune complex in solution, which is then pulled down by the insoluble beaded support (**Figure 1.8 C**).

In particular, this thesis will present the use of off-line conventional IP in solution and IP with Protein A crosslinked agarose MBs in order to purify TTR from human serum samples.

1.4. Analytical techniques

Proteomics and metabolomics have dramatically increased the demand of sensitive and selective methods for the analysis of biological samples. In this context, mass spectrometry (MS) plays a key role due to its selectivity and potential with respect to the identification of unknown compounds. Additionally, due to the complexity of most biological samples, the use of high-performance separation techniques is essential prior MS analysis. Capillary electrophoresis (CE) and liquid chromatography (LC) have gained much interest due to their high separation efficiency, sensitivity, reproducibility and possibility of on-line coupling and full automation. These features ensure high throughput, sensitive and reproducible qualitative and quantitative analysis and reliable identification by capillary electrophoresis mass spectrometry (CE-MS) and liquid chromatography mass spectrometry (LC-MS).

1.4.1. Capillary electrophoresis (CE)

Since its introduction in the ninetieth century, CE has become an increasingly important technique for the highly efficient separation and characterization of a wide range of molecules, from low molecular mass compounds (e.g. metabolites) to large biomolecules (e.g. proteins and glycoproteins) [33,134,135]. In addition to high resolution capabilities, CE has many other desirable characteristics, including a separation mechanism based on electrophoretic mobility (which is related to the ion charge-to-radius ratio), low consumption of sample and reagents, reduced analysis times, instrumental simplicity and full automation. Furthermore, if necessary, CE analyses can be carried out under mild conditions without the need of organic solvents or very high salt concentrations. This enables the study of proteins without causing conformational changes or protein degradation during analysis. Many PTMs like glycosylation or phosphorylation can potentially change the electrophoretic mobility of a protein, and the resulting proteoforms can therefore be separated with CE. The most common modes of

operation in CE are capillary zone electrophoresis (CZE, in general referred to as CE), capillary gel electrophoresis (CGE), micellar electrokinetic chromatography (MEKC), capillary electrochromatography (CEC), capillary isoelectric focusing (CIEF) and isotachopheresis (ITP) [136,137]. In particular, this thesis will present methods based on CZE (from now on it will be referred to as simply CE) for the analysis of a wide range of molecules, from low molecular mass compounds to proteins and glycoproteins.

1.4.1.1. Separation principles in CE

CE is a microscale analytical technique where charged species are separated on the basis of their electrophoretic mobility when an electric field is applied in a fused silica capillary of small internal diameter (id) (in general, between 25-75 μm), filled with an appropriate background electrolyte (BGE) [137]. Sample solution (typically between 1 and 20 nL) is then introduced at the inlet end of the capillary on the opposite side to the detector (which is the anode in normal polarity mode). Application of a high voltage (typically between 10 and 30 kV) causes electrophoretic and electroosmotic movements across the capillary, hence resulting in the movement along the capillary of the ionic species in the sample, which migrate through the detector at different times. The electrophoretic mobility (μ_e) of a target solute ion is largely governed by its size and charge, as deduced by the following equation [137]:

$$\mu_e = \frac{q}{6\pi\eta r} \quad \text{Eq. 1}$$

where q is the total ion charge, η is the viscosity of the medium (BGE) and r is the solvated ion radius, which is related to its size.

A basic fundament of CE separation is the so-called electroosmotic flow (EOF). When fused silica capillaries are used, the inner walls of the capillary contain silanol groups (SiOH) that ionize to the anionic form at pH above 2. Therefore, using an aqueous BGE at a specific pH

value, the inner walls always present a certain amount of negative charges that favor 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 at the end of the capillary in normal polarity mode), hence generating a cathodic EOF [137–139].

The generation of the EOF provides several advantages. Since the electro-driven force of the flow is uniformly distributed along the capillary, the velocity is nearly uniform throughout. This flat flow velocity profile is beneficial since it does not contribute to the broadening of analyte zones. This is in contrast to the flow velocity profile generated by pressure in LC, which yields a laminar flow with a parabolic flow velocity profile. Another great advantage of the EOF is that it causes nearly all species to be dragged towards the outlet end in a fused silica capillary and normal polarity (cathodic EOF and cathode at the outlet end). Thus cations, neutrals, and most anions can be separated in a single run since they all migrate in the same direction (**Figure 1.9**) [137–139].

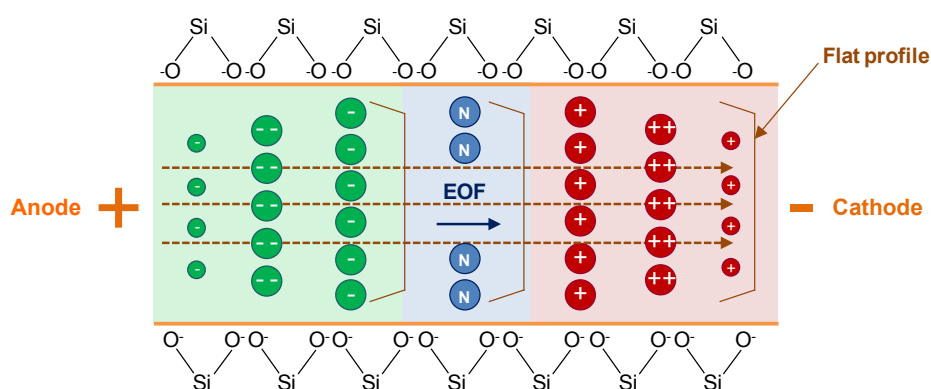


Figure 1.9. Schematic representation of the migration behavior of cations, neutrals and anions in a fused silica capillary in normal polarity (cathode at the outlet end, close to the detector).

As it is shown in **Figure 1.9**, cations migrate the fastest since the electrophoretic attraction towards the cathode and the EOF are in the same direction, neutrals migrate at the velocity of

the EOF but are not separated from each other, and anions migrate slower since they are attracted to the anode but are still pushed by the EOF towards the cathode if the EOF mobility is higher than their own electrophoretic mobility.

1.4.2. Liquid chromatography (LC)

Since its introduction in the ninetieth century, LC has become an increasingly important technique for the highly efficient separation and characterization of a number of molecules, including targeted studies in proteomics and metabolomics [31,140]. Nowadays, LC is considered the separation analytical technique par excellence and, in the last decades, it has experienced a worldwide expansion with thousands of applications.

Separations in LC rely on pumping a pressurized liquid solvent (known as mobile phase) containing the sample mixture through a column filled with a solid adsorbent material (known as stationary phase). Under optimum conditions, each component in the sample interacts slightly different with the stationary phase, thus leading to the separation of the components as they flow through the column [141]. The time at which a specific compound elutes (referred to as retention time) is governed by several factors including the type of stationary phase, the mobile phase used or the physicochemical properties of the compound (e.g. polarity, size, charge, among others).

Nowadays, there are several LC separation modes depending on the column and type of interaction between the analyte and the stationary phase (e.g. reversed phase, normal phase, hydrophilic interaction, size exclusion, ion exchange and affinity LC) [142]. However, reversed phase liquid chromatography (RP-LC) has been the most extensively used mode for the analysis of a wide range of molecules, from low molecular mass compounds to large biomolecules such as proteins and glycoproteins. Typical reversed phase stationary phases are based on silica matrices functionalized by a high density of alkyl groups (e.g. C₂, C₈ and C₁₈). In particular, the

most common columns for proteomics and metabolomics studies are C₁₈ and C₈, which have several advantages, such as their commercial availability in different sizes, their relative low cost compared to other type of columns and the fact that typical acidic hydroorganic volatile mobile phases can be used (in general, acetonitrile (ACN)-water or methanol (MeOH)-water mixtures with 0.1% (v/v) formic acid (HFor)), which favor the ionization by MS [143].

As pointed out above, apart from the type of stationary phase within the chromatographic column, several factors may influence the efficiency and resolution of the chromatographic separation in RP-LC (e.g. column dimensions, stationary phase particle diameter and porosity, type of mobile phase, presence of additives in the mobile phase). New significant advances in instrumentation and column technology have given rise to different types of LC where column dimensions, particle sizes, sample handling, pressure ranges and flow rates are considerably modified to meet with the new challenges and demands of users in different fields, including omics sciences. In this regard, ultra-high pressure liquid chromatography (UHPLC), capillary liquid chromatography (CapLC) and nano liquid chromatography (nanoLC) have been extensively used for biomarker discovery in proteomics and metabolomics studies [144–146].

In particular, CapLC is used in this thesis for the analysis of intact proteins. In CapLC, the internal column diameter is greatly reduced when compared to conventional LC, hence allowing low flow rates (2–4 $\mu\text{L}/\text{min}$) and the use of minute sample volumes (0.5 μL), with the consequent reduction of the limits of detection (LODs). This advantage is due to the reduced dilution of the chromatographic peak during analysis and the improved ionization efficiency when columns with smaller internal diameter are used. Furthermore, this technique provides many other advantages including good separation efficiency, short analysis time, more robustness than nanoLC (0.1 $\mu\text{L}/\text{min}$ < flow rate < 0.5 $\mu\text{L}/\text{min}$) or microchip technologies, instrumental simplicity and full automation [145,147,148].

1.4.3. Mass spectrometry (MS)

In the last decade, MS detection for CE and LC has been extensively used, proving its suitability for the detection and characterization of metabolites and large biomolecules at the attomole and femtomole level [42,149]. MS is a more universal detector than ultraviolet-visible (UV-Vis), laser induced fluorescence (LIF) or electrochemical detectors, as it provides a second informative dimension (i.e. molecular mass) and, therefore, enables the characterization of comigrating or coeluting molecules in CE and LC, respectively. Additionally, structural elucidation and unequivocal identification can be carried out by interpreting the fragmentation patterns obtained by tandem MS (MS/MS or MSⁿ) [150,151].

A mass spectrometer is a complex and sophisticated analytical instrument that consists of several parts, being the most important ones the ionization source, the analyzer and the detector [152]. Briefly, in the simplest design, the ionization source allows molecules in the liquid phase to be transferred directly into ions in the gas phase. Then, in the analyzer, ions are separated on the basis of their mass-to-charge ratios (m/z). The mechanism behind this separation is highly dependent on the type of analyzer, which could also affect the sensitivity and resolution of the mass spectra. Finally, the detector records either the charge induced or the current produced when an ion passes by or hits a surface.

1.4.3.1. Ionization techniques. Electrospray ionization (ESI)

Since the revolutionary introduction of electrospray ionization (ESI) in the late sixties [153], the use of ESI-MS hyphenated techniques, such as LC-MS, or, to a lesser extent, CE-MS, has given rise to numerous publications and advances in omics sciences [31,33]. The impact of ESI in modern chemistry was acknowledged in 2002 when John B. Fenn received the Nobel Prize in Chemistry.

ESI is considered a soft ionization technique as relatively little energy is imparted to the analyte, hence, barely any in-source fragmentation occurs during the ionization process. The formation of gas-phase ions from ions or neutral molecules in solution by ESI can be explained in three different steps considering a typical ESI interface for LC-MS [152,154,155] (**Figure 1.10**).

A) 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. As a result, charged droplets are formed at the end of the capillary. Spray formation and stability are quite often assisted with an auxiliary gas, being N₂ the most commonly used.

B) The solvent is evaporated, usually using an inert gas that flows in the opposite direction, being N₂ the most used one. The charged droplets are desolvated until they reach the so-called Rayleigh limit, i.e., the state at which the surface tension that holds them together is equal to the Coulombic repulsion between the charges on their surface.

C) In the Rayleigh limit, new, smaller and more stable droplets are formed from the initial charged droplet, which is often referred to as fission. Several successive rounds of desolvation and droplet fission occur to form the analyte gas ions.

The principal advantage of ESI is that multiply charged ions are formed, and both small and large molecules can be measured using a mass analyzer with a limited range of m/z . Although ESI ionization is considered a soft ionization technique, special care has to be taken regarding the potential applied to form the spray when analyzing labile molecules. A middle ground has to be found between applying too much voltage, which may cause in-source fragmentation, or too low, which will considerably decrease the sensitivity. 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 important ones.

Furthermore, the elevated vacuum inside the mass spectrometer or the acidic pH of the solution may promote disruption of certain oligomeric proteins, such the native tetrameric form of TTR.

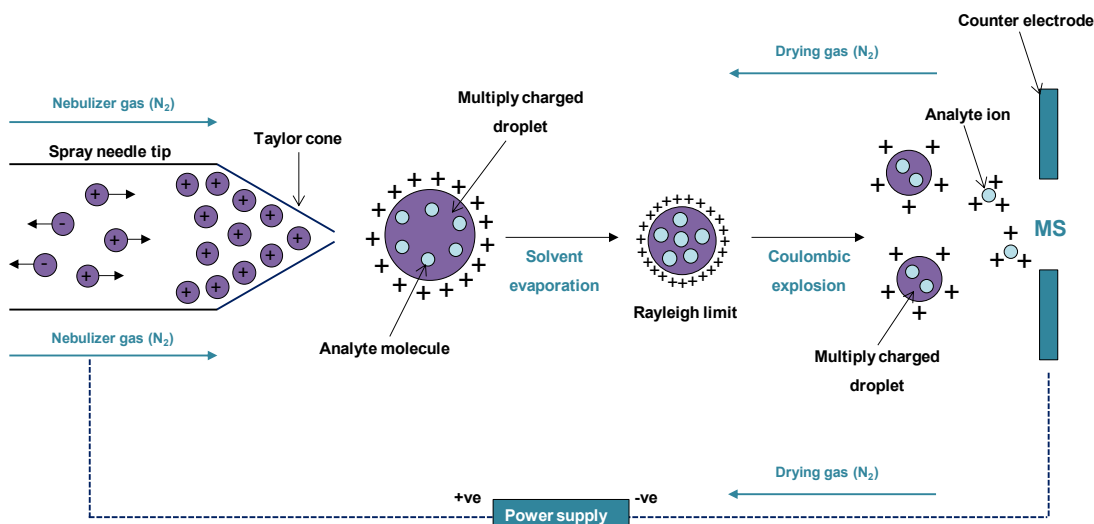


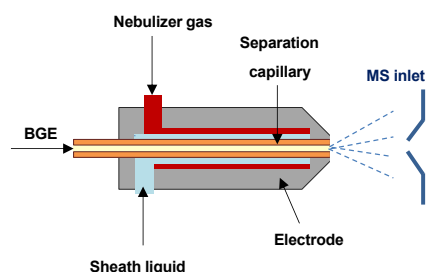
Figure 1.10. Basic principles of electrospray ionization (ESI) considering a typical ESI interface for LC-MS.

In the last few years, the scientific community has applied a new variant of ESI called nano electrospray (nanoESI), which was already introduced by Wilm and Mann in the late nineties [156]. This interface is ideally suited for the ionization of analytes in low flow rate systems (between 0.1 and 0.5 $\mu\text{L}/\text{min}$), such as nanoLC or CE. In nanoESI, the initial diameter of the formed droplets is considerably reduced, which allows faster ionization, increased sensitivity and less adduct formation.

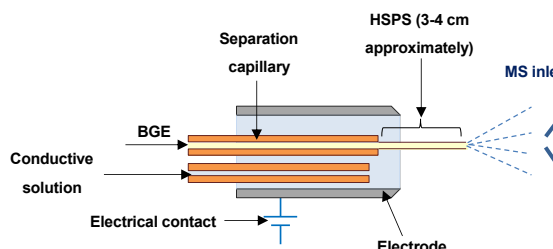
In contrast to the relative simplicity of the on-line coupling of LC to MS through ESI interfaces (LC-ESI-MS or LC-MS), the on-line coupling of CE (CE-ESI-MS or CE-MS) is not so straightforward. Compared to LC-MS, two additional key requirements have to be considered when working with CE-MS: the CE separation circuit needs to be electrically closed, and very low BGE dependent flows through the separation capillary need to be handled. In order to

accomplish the above mentioned requirements, three different CE-MS interfaces have been described: sheath flow, sheathless and liquid-junction interfaces (**Figure 1.11**) [136,138,157].

A) Sheath flow CE-MS interface



B) Sheathless CE-MS interface (HSPS)



C) Liquid-junction CE-MS interface

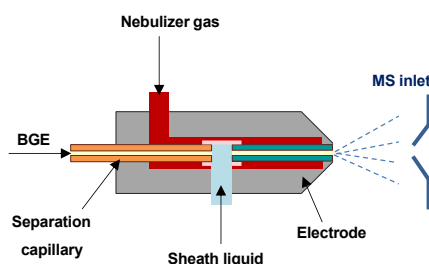


Figure 1.11. ESI interfaces for CE-MS: (A) sheath flow, (B) sheathless and (C) liquid-junction.

Sheath flow interface (Figure 1.11 A)

Firstly developed by Smith et al. [158], this robust design offers a great flexibility for BGE selection and, therefore, it can be used in a wide range of applications. Additionally, the supply of a coaxial sheath liquid enhances the reproducibility on delivering a stable spray and ionization [136,157,159]. The main drawback is that analytes are diluted in the outlet end of the separation capillary. In this regard, some authors have pointed out that this could decrease the sensitivity, while others did not find significant differences with similar microflow sheathless interfaces [160]. The first commercial sheath flow interface with nebulizer gas (N_2) assistance was developed in 1995 by Agilent Technologies, Inc. (Waldbronn, Germany). Since then,

several attempts have been made to improve the sensitivity of the sheath flow interface [161,162]. In particular, two different designs were commercialized (**Figure 1.12**).

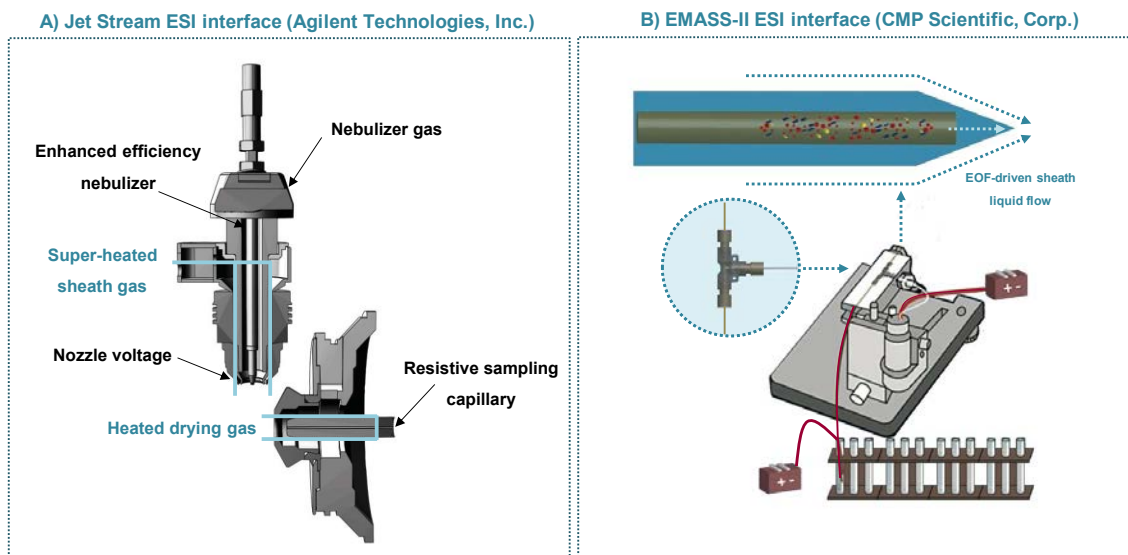


Figure 1.12. Sheath flow CE-MS interfaces: (A) Jet Stream ESI interface (Agilent Technologies, Inc.) and (B) EMASS-II ESI interface (CMP Scientific, Corp.).

The first commercial coaxial sheath flow interface adapted to Jet Stream technology was developed in 2011 by Agilent Technologies, Inc. to enhance the sensitivity by improving the desolvation and spatial focusing of the ions (**Figure 1.12 A**). A super-heated sheath gas (N₂) confines the nebulizer spray to improve ion drying and concentrate the ions in a thermal confinement zone. In this case, the gain in sensitivity was at least equivalent or only up to two-fold for the tested analytes [163]. CMP Scientific, Corp. (New York, United States) developed in 2014 a novel commercial coaxial sheath flow interface, namely EMASS II, which incorporates an EOF-driven sheath liquid electrospray emitter technology. In this interface, the separation capillary terminus makes contact with the electrospray emitter inside, forming a small dead volume to avoid loss of separation. Sheath liquid solution is introduced through a tee junction at a flow driven by a borosilicate glass surface EOF, hence minimizing dilution of the CE effluent in order to maximize sensitivity (**Figure 1.12 B**). Compared with the optimal

commercial sheath flow interface, the manufacturer declares that this design results in 50-100 fold or higher increase in MS signal. Nowadays, the original commercial sheath flow interface with nebulizer gas (N₂) assistance developed by Agilent Technologies, Inc. is the most used worldwide, mainly due to its robustness, appropriate sensitivity and versatility in BGE selection [159]. This is the interface that has been used in most of the CE-MS studies of this thesis.

Sheathless interface (Figure 1.11 B)

With sheathless interfaces no sheath liquid is necessary and, if the initial droplets formed during the electrospray process are smaller, the process leads to a more efficient ionization when compared to the sheath flow interface [157,164]. This effect is not significant with microflow sheathless interfaces [160], but optimized nanoflow sheathless interfaces may offer higher sensitivities by CE-MS. However, they have not been extensively used for CE-MS, most likely due to their limited robustness and lack of commercial availability [164,165]. Recently, some interesting new designs of CE-MS nanoflow interfaces have been described, and one of them commercialized, but at a very high price. The commercial high sensitivity porous sprayer (HSPS) nanoflow sheathless CE-MS interface (Applied Biosystems/MDS Sciex, Framingham, MA), which is based on the design of Moini [166], have gained in the last years quite some attention for several applications [135]. In this design, the last 3-4 cm of a fused silica capillary (30 µm id x 150 µm external diameter (od)) are etched with hydrofluoric acid producing a porous wall that is conductive when in contact with a conductive solution. The porous capillary outlet protrudes from a stainless steel ESI needle filled with a static electrolyte, hence allowing electrical contact and electrospray formation at the capillary tip. Haselberg et al. [167] showed that with this nanoflow sheathless CE-MS interface, LODs for four model proteins were improved by a factor of 50 to 140 compared to sheath flow CE-MS, hence leading to subnanomolar LODs. Likewise, Medina-Casanellas et al. [168] demonstrated the good performance of this nanoflow sheathless CE-MS interface, in combination with on-line solid-

phase extraction CE (SPE-CE), for the analysis of opioid peptides at the picomolar level. In this thesis, this interface has been used to analyze CEA samples.

Liquid-junction interface (Figure 1.11 C)

Firstly developed by Henion and co-workers [169], the liquid-junction interface provides the electrical connection for closing the CE circuit via a liquid reservoir. An ESI emitter capillary is positioned opposite to the end of the separation capillary with a gap of 10-25 μm between them, in such a way that the liquid and analytes from the separation capillary pass to the emitter tube and are sprayed afterwards. The main advantage of this setup is that CE and ESI-MS can be operated independently with any mass spectrometer due to the partial electrical and physical disconnection of the CE separation from the ESI emitter [136,157]. In contrast, the main disadvantages have been the difficulty of positioning the transfer capillary in a reproducible way and the potential loss of separation efficiency due to the dead volumes in passage through the ESI emitter. Although several new developments improving the original design have been reported, some of the original difficulties (e.g. peak broadening and loss of separation efficiency) frequently remain, and the construction of the junction between the CE capillary and the ESI emitter continues being a difficult task [136,157].

1.4.3.2. Mass analyzers

The mass analyzer is the part of the mass spectrometer in which ions are separated based on their m/z values. Nowadays, five main analyzers are widely used, namely, quadrupole (Q), ion trap (IT), time-of-flight (TOF), Orbitrap and Fourier transform ion cyclotron resonance (FTICR). These analyzers vary in terms of size, price, resolution, scanning range, scanning speed, dynamic range and the ability to perform MS/MS experiments. Several high-end instruments combine the advantages of more than one analyzer to form complex hybrid instruments with increased sensitivity, resolution and MS/MS capabilities (e.g. Q-TOF, triple

quadrupole (QqQ)). The following sections will give a brief overview of the mass analyzers used throughout this thesis: TOF and Q-TOF analyzers, which are widely applied in proteomics and metabolomics, because of the ease of operation, simplicity, high mass resolution, high mass accuracy, good scanning speed and wide dynamic and scanning ranges.

Time-of-flight mass analyzer (TOF)

This ion separation methodology is one of the simplest and, although it was first described in the middle of the twentieth century [170], it was not rediscovered until the early nineties [171]. TOF simply relies on the free flight of the ionized molecules in a tube of 1-2 m length, before reaching the detector. The time (**t**) taken for an ion to traverse this tube is dependent on its *m/z* ratio, as deduced by the following equation [172,173]:

$$m/z = kt^2 \qquad \text{Eq. 2}$$

where **k** is the calibration factor. In this regard, ions with greater charge and lower mass will cross the tube before those with higher mass and lower charge.

The most recent and important advancement in TOF technology was the development of the orthogonal acceleration TOF (oa-TOF) in the late eighties, which considerably further improved mass accuracy and resolving power [173–175]. In brief, the main difference introduced with the oa-TOF is the use of a separated direction (for the TOF analysis), orthogonal to the continuous ion-beam of the ion source. This distribution provides several advantages, including better efficiency in gating ions from an external continuous source (e.g., ESI), reduction of velocity and spatial dispersion and, consequently, improved mass accuracy and resolving power [174,175]. As a result of the rediscovery and the recognition of the potential of oa-TOF, several commercial instruments are now available for a range of ionization methods amongst which ESI oa-TOF has been one of the areas of greatest activity.

Q-TOF mass analyzer

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 by performing MS/MS experiments [176]. MS/MS, in a very generic description, is a process in which an ion (formed in an ion source) is mass-selected in the first stage of the analysis, fragmented and, then, the charged products from the fragmentation are analyzed in the second stage of the analysis. The type, quantity and quality of the obtained data can vary greatly depending upon the type of analyzer used in the first and second stages of the analysis, and the type of fragmentation performed between them.

The Q-TOF mass analyzer was first described in 1996 [177] as a means of combining the scanning capabilities of a quadrupole (Q) and the resolving power of a TOF mass analyzer. The Q-TOF can provide high-quality, one-stage MS and MS/MS spectra. It is composed of two quadrupoles (Q1 and q2) linked to a TOF analyzer that is geometrically aligned in an orthogonal configuration with respect to the quadrupoles (referred to as Q-TOF orthogonal mass spectrometry). Q1 quadrupole is the ion filter portion of the instrument used during MS/MS analysis, while q2 is the collision cell where fragmentation occurs (typically, low-energy collision induced dissociation, CID). It is worth mentioning that Q-TOF instruments usually include an additional quadrupole (q0) as an ion focusing device to provide collisional cooling of the ions, hence improving the quality of the ion beam [178].

1.4.4. Ion mobility mass spectrometry (IM-MS)

Although a major contributor to the rapid advances in macromolecule characterization can be traced to advances in hyphenated MS techniques (such as LC-MS and CE-MS), these techniques are unable to distinguish between ions with identical m/z ratios, thus hindering the

interpretation of mass spectra from coeluting or comigrating mixtures of isomers, conformers or oligomers that are usually highly overlapped.

Ion mobility spectrometry (IMS) has enabled the separation of gas ions on the basis of their shape, charge and size, hence resolving ions that would be otherwise indistinguishable solely by their m/z ratios in conventional MS [179–181]. With the development of ESI as an ion source for IMS [182], applications have expanded from those limited to vapor-phase samples with volatile analytes to include aqueous samples containing non-volatile analytes. Furthermore, when coupled with MS, IM-MS becomes a powerful analytical tool for investigating the molecular structure of analytes in complex samples, such as those found in proteomics [183,184] and metabolomics [185] research.

Specifically, IMS measures the time (namely drift time) taken for an ion to cross a region containing a background inert gas (usually N_2 or He) at a controlled pressure under the influence of a weak electric field (see **Figure 1.13**).

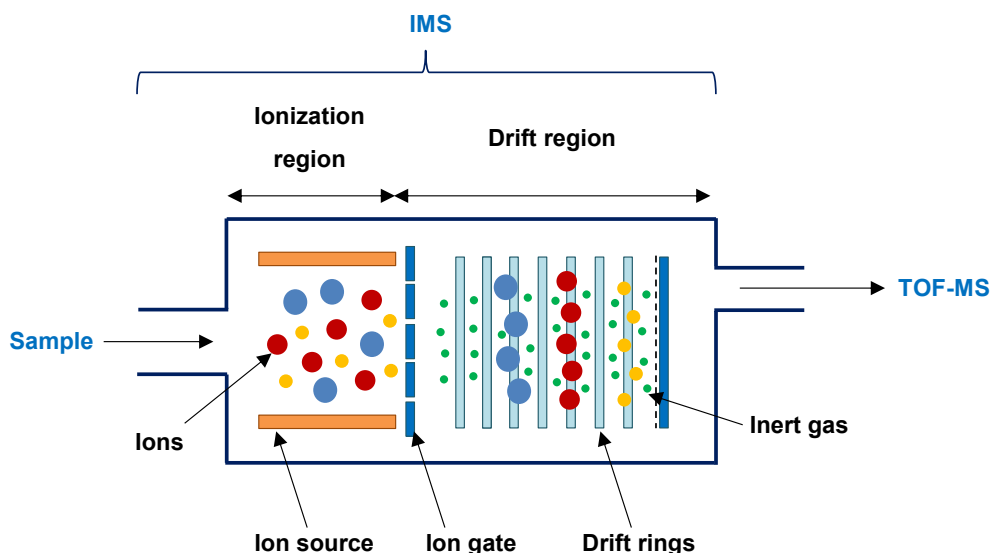


Figure 1.13. Schematic representation of an IM-MS instrument.

Drift time depends on ion-gas collisions; hence ions are separated on the basis of their ion-neutral collision cross-section (Ω), which is related to the overall shape and topology of the ion

[186]. Therefore, small ions cross first as a result of their smaller Ω . 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. Consequently, IMS is often described as being proportional to collision cross-section-to-charge ratio (Ω/z), and, when coupled on-line with MS, it provides three-dimensional analytical information, i.e., drift time, m/z and abundance, hence allowing reliable analyte identification [179–181,187,188].

Nowadays, there are several IMS modes, such as drift time ion mobility spectrometry (DTIMS), aspiration ion mobility spectrometry (AIMS), differential-mobility spectrometry (DMS), which is also called field asymmetric waveform ion mobility spectrometry (FAIMS), and traveling-wave ion mobility spectrometry (TWIMS) [189]. DTIMS provides the highest IMS resolving power and is the only IMS method that can directly measure Ω . AIMS is a low resolution mobility separation method but can monitor ions in a continuous manner. DMS/FAIMS offers continuous-ion monitoring capability as well as orthogonal ion mobility separation in which high-separation selectivity can be achieved. Finally, this thesis will present studies with an IM-MS instrument based on TWIMS, a novel method of IMS that has recently been developed and commercially introduced by Waters Corporation (Milford, United States) (see **Figure 1.14**).

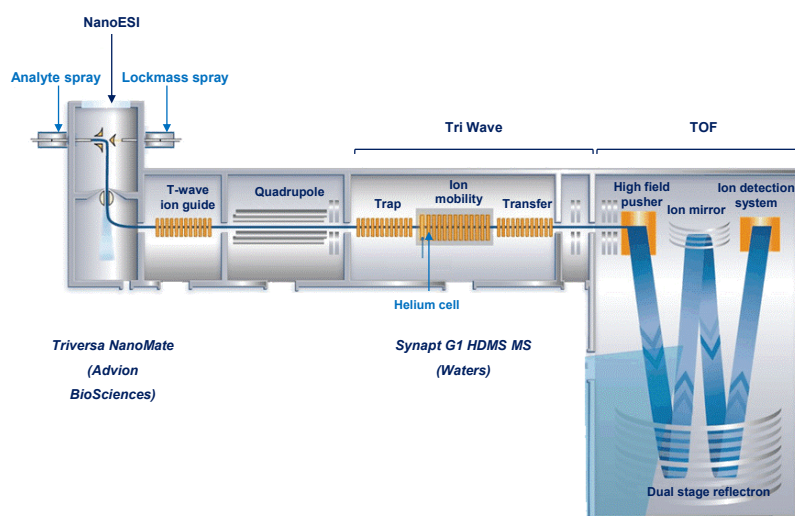


Figure 1.14. Schematic representation of a TWIMS instrument from Waters Corporation.

In TWIMS, ions are separated based on their mobility when they are propelled due to a sequence of symmetric potential waves continually propagating through a cell [189–191]. One of the main advantages of TWIMS is that it disperses ion mixtures, fact that allows the simultaneous measure of multiple species. This, in conjunction with the high sensitivity obtained when TWIMS is coupled to certain analyzers in MS, such as TOF or Q-TOF, has made this technique an appealing option for structural analysis and isomer separation [189–191].

1.5. On-line preconcentration in CE

As introduced before, CE separations are very efficient and require only small amounts of reagents, solvents and samples. However, when compared to LC, concentration sensitivity is generally rather low, one of the reasons that explain its limited use for routine chemical analysis or bioanalytical studies. The low loadability (i.e. nL injection volumes) of CE is the main cause for this poor concentration sensitivity. In this regard, in addition to the use of improved detectors, the concentration sensitivity of CE can be enhanced by electrophoretic or chromatographic preconcentration [192–198]. Electrophoretic preconcentration based on stacking or isotachopheresis (ITP) has been widely used, in some cases, with preconcentration factors higher than 10,000 fold [192–194]. However, these methods are, in general, analyte and matrix dependent, and show limited reproducibility. Chromatographic preconcentration techniques using on-line solid-phase extraction capillary electrophoresis (SPE-CE) show a more general applicability and can be also very efficient yielding high concentration factors and better reproducibility [194,197,199–201]. Moreover, SPE-CE can be used for sample clean-up and purification, with minimum sample handling. In the typical on-line SPE-CE configuration applied in the studies of this thesis (**Figure 1.15**), a microcartridge placed near the inlet of the separation capillary contains a sorbent, which selectively retains the target analyte, hence enabling the introduction of large volumes of sample (~50-100 μ L). After washing to remove non-retained molecules, the retained analyte is eluted in a small volume of an appropriate solution (~25-50 nL), resulting in sample clean-up and concentration enhancement before electrophoretic separation and detection [197,201].

Despite the usefulness of SPE-CE for the analysis of a wide range of molecules, many drawbacks need to be addressed before it can be routinely used to analyze low-abundance analytes in complex samples. For instance, it is difficult to find sorbents for optimum performance in SPE-CE, due to the reduced dimensions of the microcartridges and the fact that the extraction is undertaken on-line with a voltage-driven separation. In SPE, selectivity is

mainly governed by the specific affinity of the sorbent to the target analyte, while sorbent capacity also depends on the number of active sites available for the extraction, which is related to the sorbent particle diameter and porosity [202]. Reversed phase, ionic exchange, size-exclusion, molecular imprinted polymers, antibodies, aptamers or lectin-based sorbents that are commonly developed for off-line SPE or affinity chromatography can be applied to SPE-CE [197,200,201,203].

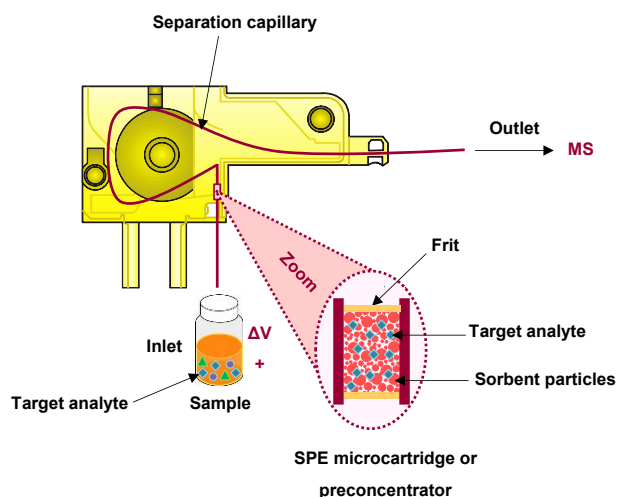


Figure 1.15. On-line SPE-CE setup used in this thesis.

The main disadvantage of the above mentioned design for SPE-CE is the fact that the sample is introduced in the direction of the separation, from the inlet end to the outlet end, hence making possible the adsorption of the matrix compounds in the inner walls of the separation capillary. In order to overcome this major drawback, some new designs have been developed, where the sample is introduced through a transport tube or passage, orthogonal to the separation (cruciform and staggered designs) (**Figure 1.16**). The new cruciform and staggered designs introduce a number of advantages when compared with the original design, including elimination of back-pressure, control of fluid flow by valves, no carry-over of substances into the separation capillary and extended useful life of the separation capillaries [197,198].

However, they are more difficult to construct than the typical SPE microcartridges used in the studies of this thesis (**Figure 1.15**).

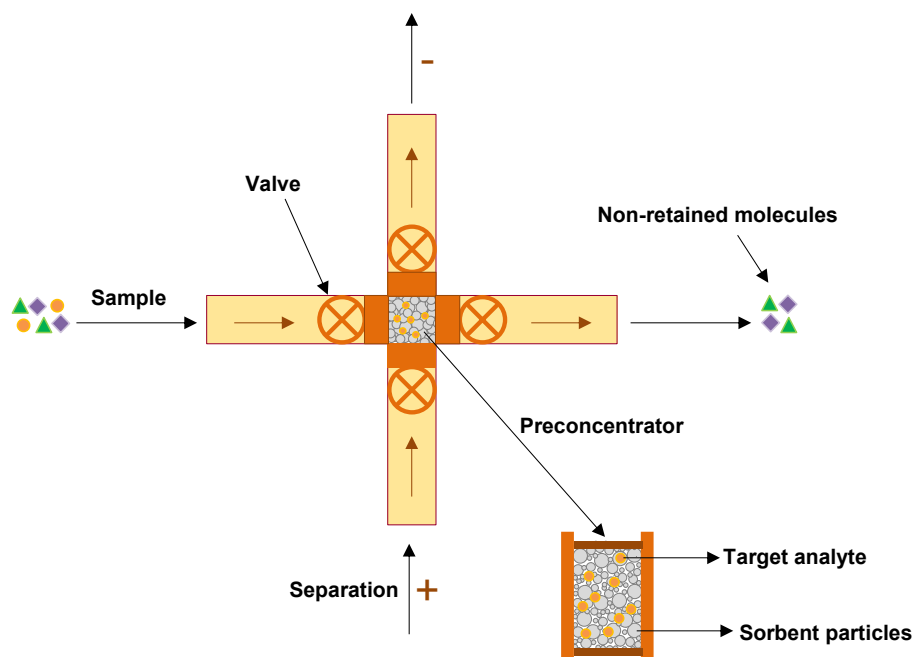


Figure 1.16. Cruciform design for SPE-CE. Sample introduction is orthogonal to the separation.

1.5.1. SPE-CE with conventional chromatographic sorbents

Nowadays, conventional chromatographic sorbents, especially reversed phase and ionic exchange sorbents, are the most common ones for SPE-CE, as they fulfill most of the requirements for optimum performance. On the one hand, they provide a large active surface area, without interfering with the on-line electrophoretic separation [201,203]. On the other hand, they are commercially available, have been optimized and widely used for the analysis of a great variety of compounds and can be purchased at a reasonable price. In particular, silica-based sorbents (e.g. C_8 and C_{18}) are widely recognized for their high efficiency and good extraction capacities, being C_{18} the most common chromatographic sorbent used for SPE-CE applications involving peptides [201,203,204].

However, the major drawback of these conventional chromatographic sorbents is their limited selectivity, which precludes the direct analysis of complex samples such as biological fluids. In such cases, a previous clean-up pretreatment, which can be more or less laborious, is required to purify and enrich the target analytes in order to prevent microcartridge saturation [201,203]. SPE-CE performance can also be improved by exploring the use of sorbents with higher extraction selectivity, such as immobilized metal ion affinity chromatography (IMAC) sorbents, aptamers, lectin-based sorbents and, especially, immunoaffinity (IA) sorbents.

1.5.2. SPE-CE with immunoaffinity sorbents (IA-SPE-CE)

Since the introduction of immunoaffinity capillary electrophoresis (IACE) in the early nineties [205], several IACE applications have been developed employing conventional CE or microchip CE formats [197,198,206,207]. In particular, this thesis will present a study using on-line immunoaffinity solid-phase extraction capillary electrophoresis mass spectrometry (IA-SPE-CE-MS) for the targeted enrichment, clean-up and characterization of a large biomolecule, i.e. TTR, in serum. IA-SPE-CE is a variant of SPE-CE based on immunosorbents prepared by immobilization of antibodies or antibody fragments. Therefore, this hyphenated technology combines the high selectivity of immunocapture by IA-SPE with the high separation efficiency of CE, which can be also coupled on-line with MS detection for a reliable compound identification [208–211].

IA-SPE-CE has been primarily described with laser induced fluorescence detection (LIF) for preconcentration of small molecules, mostly peptides, from biological samples [197,198,206,207]. In contrast, IA-SPE-CE-MS has been demonstrated to a much lesser extent due to the difficulties of making compatible the requirements of IA-SPE-CE with on-line MS detection. The IA-SPE-CE-MS applications described to date in the literature are shown in **Table 1.4**.

Table 1.4. IA-SPE-CE-MS applications described to date in the literature. Abbreviations: **GnRH**, gonadotropin releasing hormone; **hEPO**, human erythropoietin; **NESP**, novel erythropoiesis stimulating protein; **End**, endomorphin; **Tf**, transferrin; **TTR**, transthyretin.

IA-SPE-CE-MS							
Analyte	Sample	SPE configuration	IA sorbent	Type of immobilization	Mass analyzer	LOD	[Ref.]
GnRH	Serum and urine	Microcartridge with frits	Polyclonal Fab' antibody fragments coupled to glass beads	Covalent	Quadrupole (Q)	1 ng/mL	[212]
hEPO and NESP peptides	Standards and tryptic digests	Microcartridge without frits	Polyclonal intact antibody coupled to a CNBr-Sepharose solid support	Covalent	Ion trap (IT)	25 mg/L	[213]
Opioid peptides: End 1 and End 2	Standards and human plasma	Microcartridge with frits	Polyclonal oxidized intact antibody coupled to hydrazide silica particles	Covalent	Time-of-flight (TOF)	Standards: 1 ng/mL Plasma: 100 ng/mL	[208]
Opioid peptides: End 1 and End 2	Standards and human plasma	Microcartridge with frits	Polyclonal Fab' antibody fragments coupled to succinimidyl silica particles	Covalent	Time-of-flight (TOF)	Standards: 0.5-5 ng/mL Plasma: 1-50 ng/mL	[209]
Tf	Standards and human serum	Microcartridge with frits	Polyclonal oxidized intact antibody coupled to hydrazide silica particles	Covalent	Time-of-flight (TOF)	-	[210]
TTR	Standards and human serum	Microcartridge without frits	Polyclonal intact antibody coupled to magnetic beads functionalized with amino reactive groups	Covalent	Time-of-flight (TOF)	1 µg/mL	[211]

Reports on the analysis of large biomolecules by “fully on-line” IA-SPE-CE-MS are especially scarce due to the additional issues related to low extraction efficiency, poor ionization efficiency and adsorption onto the inner capillary walls. Our research group has recently described two different IA-SPE-CE-MS methods based on immunosorbents prepared with intact antibodies immobilized on conventional silica particles or magnetic beads for the analysis of transferrin (Tf) and TTR in serum samples, respectively [210,211].

Another important issue in IA-SPE-CE is the limited commercial availability of compatible IA sorbents. Generally, researchers develop their own IA sorbents with the most suitable features for IA-SPE-CE. There is a wide variety of methods for preparing IA sorbents, which immobilize intact antibodies or their active fragments (obtained after enzymatic digestion) onto different supports [214,215]. Silica-based supports (commonly functionalized with diol, amino or hydrazide groups, among others) are generally preferred as a solid support because they fulfil most of the desired requirements to avoid flow restriction, bubble formation and current instability during the on-line electrophoretic separation by IA-SPE-CE [215].

Covalent immobilization is by far the most popular approach for attaching intact antibodies or their active fragments onto silica supports in IA-SPE-CE because of the improved stability of the supports, which can be then reused [215]. This process involves the activation of the solid support, as well as the activation of the antibody or its active fragments in order to generate the desired reactive groups so that the immobilization takes place through a specific region. Additionally, orientation of the antibodies or antibody fragments on the solid support is of great concern because the performance of the IA sorbent depends on the availability of the binding sites for the antigen-antibody interaction [215]. Several immobilization strategies have been used for covalent immobilization of intact antibodies onto silica supports (**Figure 1.17**). Reductive amination (commonly known as Schiff base), glutaraldehyde, carbonyldiimidazole (CDI), N-hydroxysuccinimide (NHS) and hydrazide methods are typically used [215]. However, only the hydrazide method ensures an appropriate orientation of the intact antibody,

which is always immobilized onto the activated silica support (with hydrazide groups) via its Fc region.

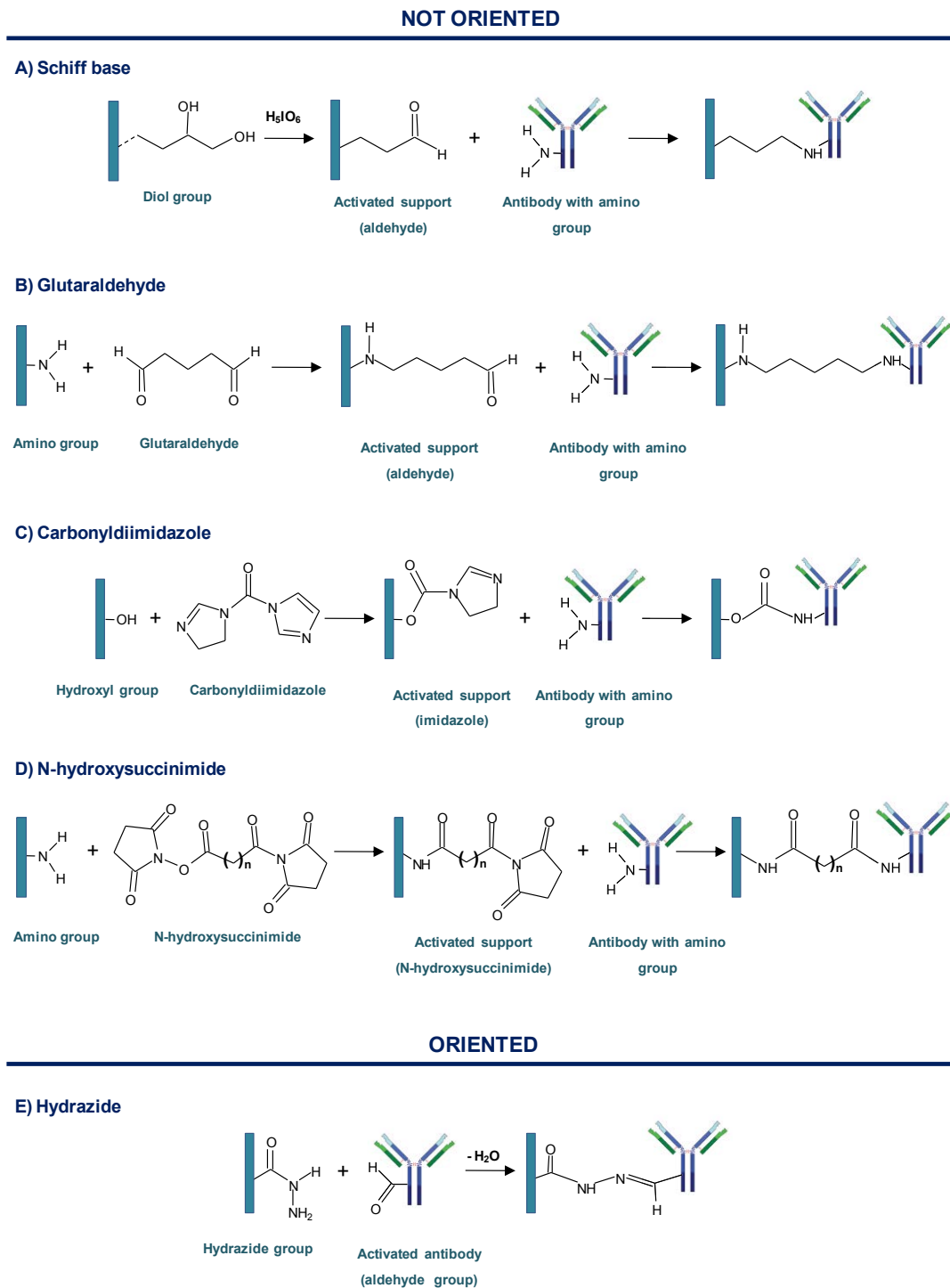


Figure 1.17. Immobilization procedures for the covalent attachment of intact antibodies to activated silica supports.

Additionally, the maleimide method, which has been used to prepare IA sorbents in the last decades, is commonly used for the covalent immobilization and appropriate orientation of antibody fragments (**Figure 1.18**). The preparation of the IA sorbent is based on the derivatization of aminopropyl silica particles with succinimidyl groups, and the fragmentation of the antibody (with pepsin) so as to obtain the Fab' antibody fragments with sulfhydryl groups covalently attached to the succinimidyl solid support. Following this immobilization method, the antigen binding sites of the Fab' fragments are supposed to be appropriately oriented, hence facilitating the immunocapture of the target analytes [215].

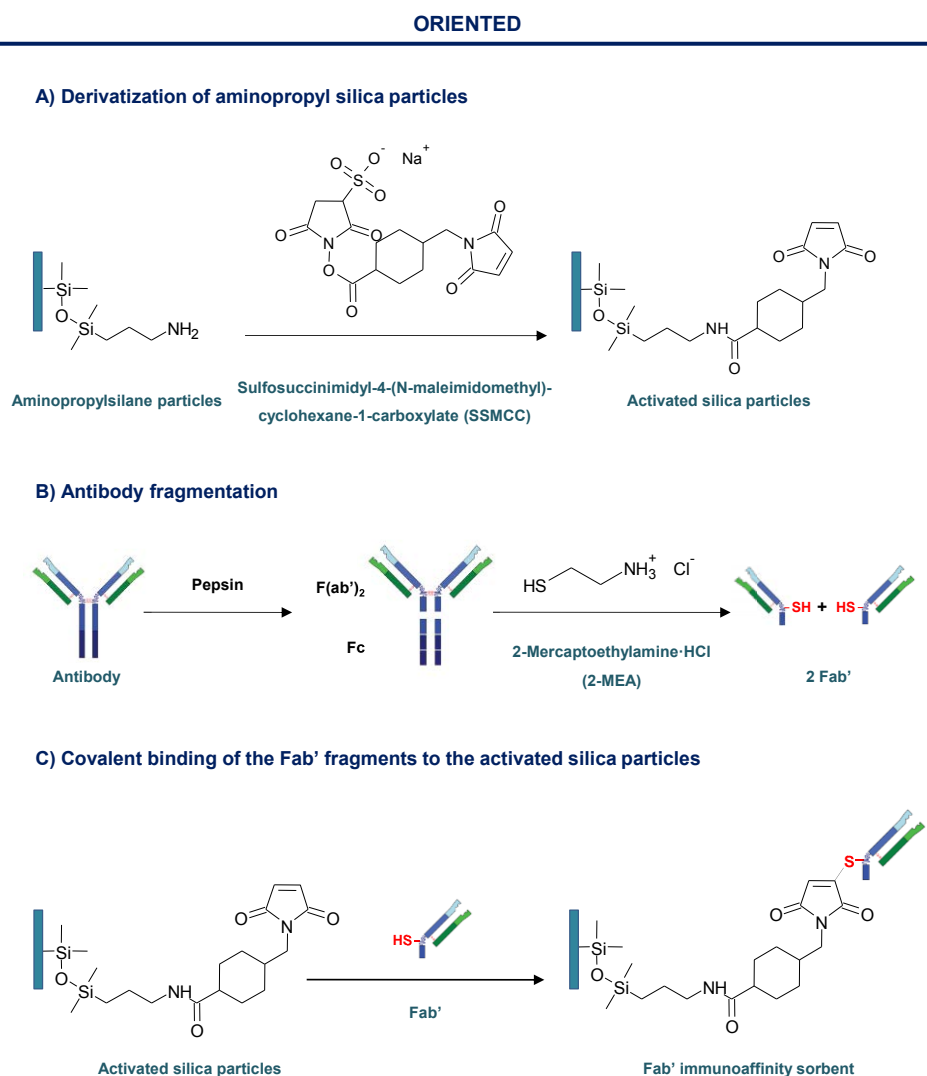


Figure 1.18. Maleimide method for the covalent attachment of Fab' antibody fragments to activated silica supports.

In particular, this thesis will present an IA-SPE-CE-MS method using a lab-made IA sorbent with Fab' antibody fragments for the analysis of a large biomolecule, TTR, in a complex biological sample, i.e. human serum.

1.6. Data analysis. Chemometric methods

Chemometric methods play a crucial role in data processing, exploration, and classification of the massive and complex datasets generated in proteomics and metabolomics studies, especially in untargeted analysis. Some of the studies presented in this thesis will show the benefits of using multivariate data analysis tools, in particular, multivariate curve resolution alternating least squares (MCR-ALS), principal component analysis (PCA) and partial least squares discriminant analysis (PLS-DA).

1.6.1. MCR-ALS

MCR-ALS has become a popular chemometric method for the resolution of multiple component responses in complex unresolved mixtures [216]. Specifically, this recognition is due to the great variety of data sets that can be analyzed by curve resolution methods; essentially, any multicomponent system that gives as a result data tables or data matrices that can be described by a bilinear model. This description includes all kind of processes and mixtures (e.g. chemical reactions, industrial processes, chromatographic separations, spectroscopic images) monitored by diverse multivariate responses, such as spectral information, electrochemical signals, composition profiles or others. Furthermore, other reasons for the great acceptance of MCR-ALS are its ability to deal with multiple data matrices simultaneously (reducing factor analysis intrinsic ambiguities [216,217] and/or data rank deficiencies [216,218]), the diversity and flexible application of constraints to help and improve the resolution results, and the utility to deal with data lacking of reproducibility. In this thesis, MCR-ALS has been used for the analysis of multicomponent systems with strongly overlapping contributions, such as those present in metabolomics studies by CE-MS. MCR-ALS can resolve overlapped electrophoretic peaks from the collected data and provide the separation profiles and mass spectra of the constituents in the analyzed samples. This approach allows overcoming problems such as

migration time shifts, background noise contributions, and differences in signal-to-noise ratios (S/Ns) among different injections [219].

MCR methods are based on a bilinear model, as deduced by the following equation:

$$\mathbf{D} = \mathbf{C}\mathbf{S}^T + \mathbf{E} \quad \text{Eq. 3}$$

The goal of MCR-ALS is the bilinear decomposition of the data matrix \mathbf{D} (e.g. the migration time \times m/z matrix obtained in a CE-MS analysis) into the pure response profiles associated with the variation of each contribution in the row and the column directions, represented by matrices \mathbf{C} and \mathbf{S}^T , respectively, which are responsible for the observed data variance. Specifically, in CE-MS separations, \mathbf{C} contains the electrophoretic profiles of the resolved contributions (components), matrix \mathbf{S}^T contains the corresponding mass spectra of the resolved contributions, and matrix \mathbf{E} contains the residuals unexplained by the model.

MCR-ALS solves iteratively **Eq. 3** by an alternating least squares (ALS) algorithm which calculates concentration \mathbf{C} and pure spectra \mathbf{S}^T matrices optimally fitting the experimental data matrix \mathbf{D} . This optimization is carried out for a proposed number of components (selected by singular value decomposition, SVD) and using initial estimates of either \mathbf{C} or \mathbf{S}^T . Initial estimates of \mathbf{C} or \mathbf{S}^T can be obtained either using evolving factor analysis [220] or simple to use interactive self modeling mixture analysis (SIMPLISMA) [221]. During the ALS optimization, several constraints can be applied to model the shapes of the \mathbf{C} and \mathbf{S}^T profiles, such as non-negativity, unimodality, closure, selectivity or/and other shape or hard-modeling constraints [222,223]. Convergence is achieved when in two consecutive iterative cycles, relative differences in standard deviations of the residuals between experimental and ALS calculated data values are less than a previously selected value, usually chosen as 0.1%. **Figure 1.19** summarizes the processes involved in MCR-ALS.

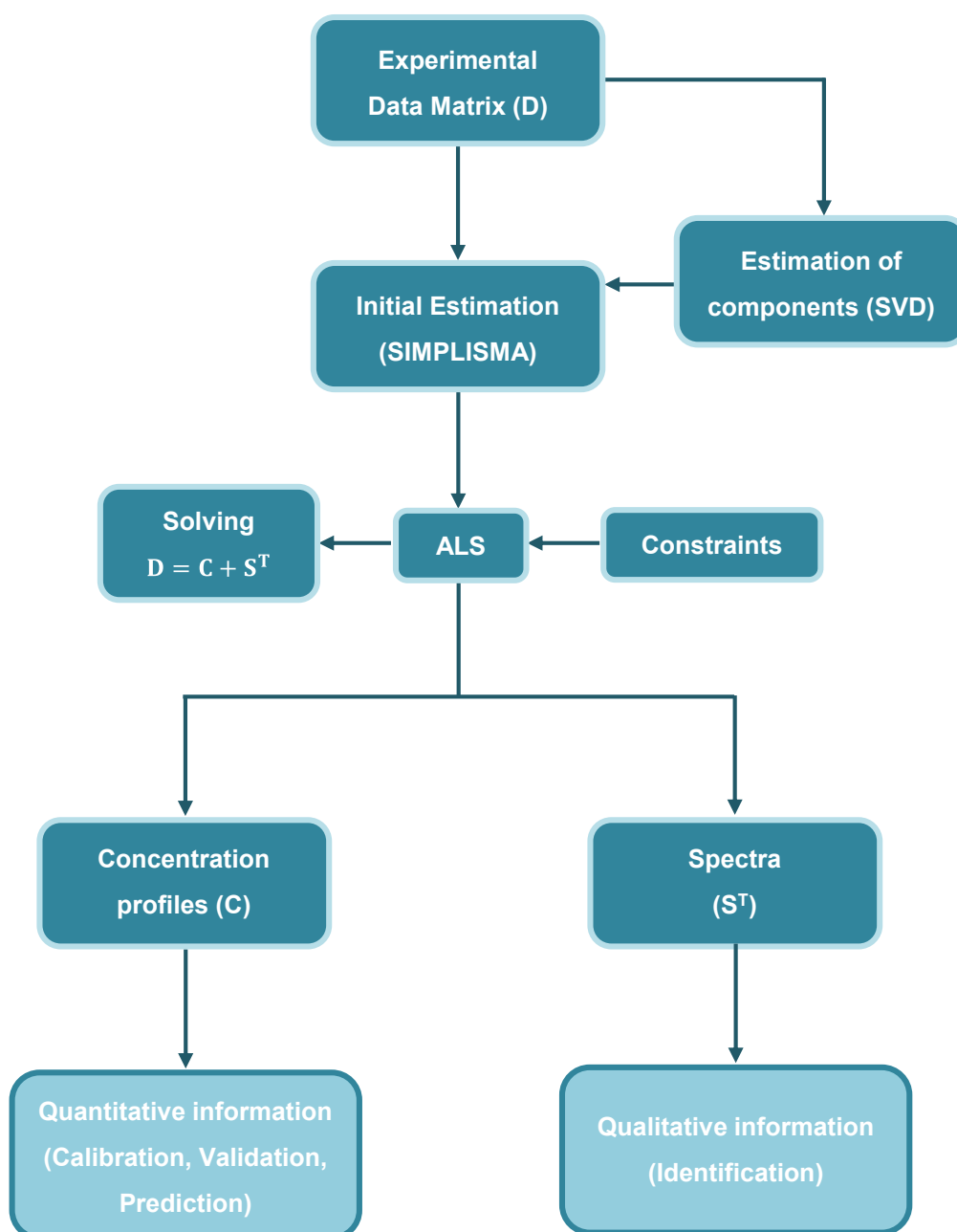


Figure 1.19. MCR-ALS for experimental data resolution.

Once MCR-ALS is performed, data analysis tools such as PCA and PLS-DA can be applied to the complex datasets in order to maximize class separation and identify the most important components to explain differences between groups, especially, in untargeted approaches.

1.6.2. PCA

PCA is a widely applied mathematical tool for unsupervised data decomposition and dimensionality reduction, which helps to understand and interpret large and complex datasets [224–226]. PCA is a technique that condenses all the information into a few number of components (principal components, PCs). Once the decomposition is performed, each score vector can be refolded to show the relative distribution map for each component. Therefore, it gives an abstract decomposition of experimental data, which maximizes the explained variance under the constraint of orthonormality of the components [226,227]. PCA decomposes the measurement matrix \mathbf{X} into the scores \mathbf{T} and loadings \mathbf{P}^T factor matrices, as deduced by the following equation:

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} \quad \text{Eq. 4}$$

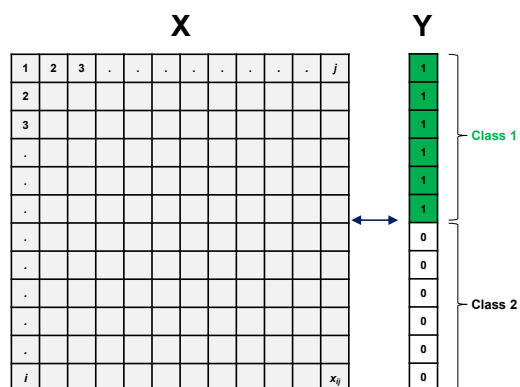
The aim of PCA is to maximize the explained variance in the data with a minimum number of components. \mathbf{T} and \mathbf{P}^T factor matrices represent a concise summary of the original data that, in most cases, can aid in interpreting the underlying data variance sources. Specifically, scores plots are quite useful for revealing patterns, such as clusters, trends and outliers, in the data. Additionally, loadings plots are mainly used to check whether there is covariance among variables or to explain and interpret the patterns observed in the scores plot. However, due to the applied constraints during the PCA bilinear decomposition (orthonormality, normalization and maximum variance), \mathbf{T} and \mathbf{P}^T profiles are not providing the profiles of the true variance sources, but a linear combination of them fulfilling the applied constraints [226,227].

1.6.3. PLS-DA

Introduced approximately three decades ago [228–230], PLS-DA also reduces the dimensionality of a dataset matrix by means of decomposition into a set of components, in this

case referred to as latent variables (LVs) [228–231]. In contrast to PCA, which is an unsupervised data decomposition method, PLS-DA is used for the supervised identification of trends and clustering of the data. The method is in fact an extension of partial least squares (PLS). In PLS, a matrix is composed of normalized weight vectors (\mathbf{WT}), which are calculated as the covariance between the response matrix \mathbf{Y} (i.e. groups, class membership) and the data matrix \mathbf{X} (i.e. raw data). Scores for the PLS components are calculated by projecting the spectral variables \mathbf{X} on \mathbf{WT} , whereas loadings are calculated by projecting \mathbf{X} on the resulting scores vectors [229]. When PLS is used as a supervised classification method, the response variable is just a binary vector of zeros and ones (in contrast to PCA, where only the matrix \mathbf{X} is present), which describes the class membership for each sample in the studied groups. In this case, the method is referred to as PLS-DA [232,233]. **Figure 1.20** shows a representation of matrices \mathbf{X} and \mathbf{Y} for PLS-DA models including two and three classes.

A) Two-class model



B) Three-class model

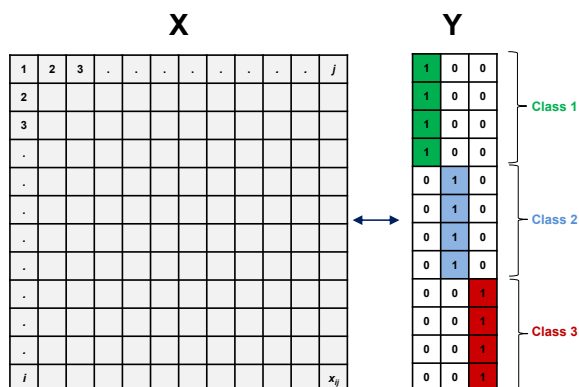


Figure 1.20. Representation of partial least squares discriminant analysis (PLS-DA) for models including (A) two classes and (B) three classes.

The main advantage of the PLS-DA approach is the availability and handling of highly collinear and noisy data, which are very common outputs from proteomics and metabolomics experiments. Furthermore, it can be performed after MCR-ALS in order to maximize class separation, especially in untargeted approaches. This method provides several statistics such as

the loading weight, the regression coefficient and the variable importance in the projection (VIP), which can be used to identify the most important variables and their importance to explain differences between classes. As in PCA, this technique provides a visual interpretation of complex datasets through a low-dimensional, easily interpretable scores plot that illustrates the separation between different groups. Comparison of loadings and scores plots supports investigations in terms of the relationship between important variables that can be specific to the group of interest [232,233].

In particular, this thesis will present some PLS-DA studies after outlier and class exploration by PCA (in combination, or not, with MCR-ALS) for the identification of the most relevant compounds to differentiate between two groups of complex samples generated in proteomics and metabolomics experiments, for both targeted and untargeted approaches.

Chapter 2. Targeted analysis of protein biomarkers. Top-down proteomics

In recent years, an increased emphasis has been placed on the separation, detection, characterization and quantification of normal and variant proteoforms of different proteins as indicators of an ongoing disease state. Unfortunately, despite recent advances in separation techniques and mass spectrometry (MS) instrumentation, the goal of profiling proteins still remains a complex analytical problem, mainly due to the mixture of large proteoforms resulting from protein microheterogeneity, the extreme complexity typically present in proteomics samples and the wide dynamic range of protein concentrations. In this regard, the rise of top-down strategies, which are focused on the characterization of intact proteins, has ushered in a new age of promise and challenge for the reliable identification and characterization of protein biomarkers in complex biological samples, without the inconvenience of performing enzymatic digestions (bottom-up strategies). The analysis of proteins by MS allows for a direct characterization of post-translational modifications (PTMs) and avoids several issues associated with bottom-up proteomics.

The present chapter is focused on the **targeted top-down** analysis of transthyretin (TTR), a homotetrameric protein involved in familial amyloidotic polyneuropathy type I (FAP-I), which is the most common hereditary systemic amyloidosis. FAP-I is associated with a TTR variant that presents a single amino acid substitution of valine for methionine at position 30 of the sequence (Met30). FAP-I Met30 can be easily detected analyzing the monomeric proteoforms of the mutant protein. However, the mechanism of protein aggregation onset, which could be triggered by structural changes on the native tetrameric protein complex (e.g. conformational changes, protein complex dissociation, presence of certain proteoforms, etc.), remains relatively unknown.

In this chapter, we describe different off-line and on-line immunoprecipitation (IP) procedures under denaturing and non-denaturing conditions for the isolation of TTR from healthy controls and FAP-I human serum samples. Moreover, several MS-hyphenated techniques, such as capillary electrophoresis mass spectrometry (CE-MS), capillary liquid chromatography mass

spectrometry (CapLC-MS) and ion mobility mass spectrometry (IM-MS), are evaluated for the detection and characterization of normal and variant TTR proteoforms and oligomers. The obtained results permit us to propose complementary and reliable methods to screen, diagnose, and follow-up patients with suspected TTR amyloidosis, as well as to gain a novel insight into the structural changes on TTR proteoforms to understand the mechanisms underlying FAP-I.

This chapter includes the following publications:

- **Publication 2.1.** Analysis of transthyretin in human serum by capillary zone electrophoresis electrospray ionization time-of-flight mass spectrometry. Application to familial amyloidotic polyneuropathy type I. L. Pont, F. Benavente, J. Barbosa, V. Sanz-Nebot. *Electrophoresis* 36 (2015), 1265-1273.
- **Publication 2.2.** Comparison of capillary electrophoresis and capillary liquid chromatography coupled to mass spectrometry for the analysis of transthyretin in human serum. L. Pont, K. Poturcu, F. Benavente, J. Barbosa, V. Sanz-Nebot. *Journal of Chromatography A* 1444 (2016), 145-153.
- **Publication 2.3.** Characterisation of serum transthyretin by electrospray ionisation-ion mobility mass spectrometry: Application to familial amyloidotic polyneuropathy type I (FAP-I). L. Pont, F. Benavente, M. Vilaseca, E. Giménez, V. Sanz-Nebot. *Talanta* 144 (2015), 1216-1224.
- **Publication 2.4.** On-line immunoaffinity solid-phase extraction capillary electrophoresis mass spectrometry using Fab' antibody fragments for the analysis of serum transthyretin. L. Pont, F. Benavente, J. Barbosa, V. Sanz-Nebot. *Talanta* 170 (2017), 224-232.

Laura Pont
Fernando Benavente
José Barbosa
Victoria Sanz-Nebot

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

Received December 11, 2014
Revised March 3, 2015
Accepted March 3, 2015

Research Article

Analysis of transthyretin in human serum by capillary zone electrophoresis electrospray ionization time-of-flight mass spectrometry. Application to familial amyloidotic polyneuropathy type I

Transthyretin (TTR) is known to misfold and aggregate, causing different types of amyloidosis. Familial amyloidotic polyneuropathy type I (FAP-I), which is the most common hereditary systemic amyloidosis, is associated with a TTR variant that presents a single amino acid substitution of valine for methionine at position 30 (Met 30). To screen for TTR-related amyloidosis rapidly and reliably, we have developed a novel procedure based on the analysis of monomers from the homotetrameric protein (~56 kDa). First, we established a CZE-ESI-TOF-MS method to detect wild-type (normal) TTR with or without several PTMs, as well as an extra minor isoform in TTR standard solutions. Later, a sample pretreatment based on immunoprecipitation (IP) and centrifugal filtration was optimized to analyze serum samples from healthy controls and FAP-I patients (including an asymptomatic patient, a symptomatic patient, a liver-transplanted patient with the specific mutation, and a patient originally without the mutation who received a liver transplant from an FAP-I patient (iatrogenic FAP-I)). The mutant TTR (Met 30) variant with a relative molecular mass 32.07 higher than the wild-type TTR was found in the asymptomatic, the symptomatic and the iatrogenic FAP-I patients, who interestingly also presented the same concentration ratio between both variants of TTR (abnormal and normal). In contrast, as in the healthy controls, the abnormal TTR variant was not detected in the liver-transplanted patient with the specific mutation, which confirms the effectiveness of the treatment. The proposed procedure could be regarded as a suitable screening system for individuals with suspected TTR amyloidosis, and to gain insight into TTR structure, to understand the mechanism underlying the disease.

Keywords:

Familial amyloidotic polyneuropathy / Immunoprecipitation / Isoforms / PTMs / Transthyretin
DOI 10.1002/elps.201400590



Additional supporting information may be found in the online version of this article at the publisher's web-site

1 Introduction

The detection, characterization, and quantification of protein variants and abnormally modified proteins are important for

clinical diagnosis and for elucidation of the pathogenesis of various diseases [1, 2]. This is particularly the case for a group of diseases named amyloidosis, in which some organ functions are destroyed as a result of deposits of normally soluble proteins as stable insoluble fibrils [3–5]. Transthyretin (TTR) (formerly known as prealbumin because it migrates before albumin in typical electrophoresis separations) is one of at least 16 human proteins known to precipitate as amyloid fibrils [6–8]. Other examples are amyloid β -protein ($\Delta\beta$) and prion protein (PrP), which cause Alzheimer and Creutzfeldt-Jakob disease, respectively [7, 8]. TTR is predominantly expressed in the liver and in the choroid plexus of the brain,

Correspondence: Dr. Fernando Benavente, Department of Analytical Chemistry, University of Barcelona, Barcelona, Spain
E-mail: fbenavente@ub.edu
Fax: +34-934021233

Abbreviations: CSF, cerebrospinal fluid; DI, dimer; EIE, extracted ion electropherogram; FAP-I, familial amyloidotic polyneuropathy type I; HAc, acetic acid; HFor, formic acid; IP, immunoprecipitation; MO, monomer; M_r , relative molecular mass; SSA, senile systemic amyloidosis; TTR, transthyretin; TIE, total ion electropherograms

Colour Online: See the article online to view Figs. 1 and 2 in colour.

and it binds to thyroid hormones and transports thyroxine and retinol associated with retinol-binding protein. This protein is highly abundant in serum (200–400 $\mu\text{g TTR}\cdot\text{mL}^{-1}$) and cerebrospinal fluid (CSF) (10–40 $\mu\text{g TTR}\cdot\text{mL}^{-1}$). It is a homotetramer composed of four identical, noncovalently associated subunits. Each monomer consists of 127 amino acid residues (approximately 14 kDa) [9–11]. Today, about 100 point mutations are known in the TTR-gene. Several of the resulting genetic TTR variants are harmless, but many play crucial roles in different types of hereditary TTR amyloidosis [12]. In addition to mutations at protein sequence level (isoforms), this protein exists in several forms with different PTMs. The most abundant is the mixed disulfide with the amino acid cysteine at position 10 (TTR-Cys) [11, 13–15].

Amyloid fibrils of TTR cause two different pathogenic conditions: familial amyloidotic polyneuropathy (FAP) and senile systemic amyloidosis (SSA). FAP is a genetically inherited neurodegenerative disease associated with around 73 point mutations. It is hypothesized that amino acid substitutions alter the stability of the tetramer, presumably leading to misfolding, aggregation, or polymerization of TTR monomers to form amyloid fibrils. SSA is a sporadic cardiomyopathy disease that affects about 25% of the male population over the age of 80. In this case, fibrils are formed from wild-type TTR [16, 17].

Among familial amyloidotic polyneuropathy, FAP type I Met 30 (single amino acid substitution of valine for methionine at position 30 of the TTR monomer sequence) is the most common [18, 19]. The disease is characterized by a progressive painful peripheral neuropathy, although symptoms involving the gastrointestinal tract, heart, kidney, ocular tissues, and autonomic nervous system are also present [18]. FAP is a fatal disease and liver transplantation is the most promising therapy for patients today [20, 21].

Nowadays, the analysis of normal and variant forms of TTR in serum samples is based on immunoprecipitation (IP) methods, followed by MALDI-TOF-MS or ESI-MS, coupled on-line, or not, with HPLC [22–24]. For the first time to our knowledge, we propose CZE-ESI-TOF-MS as an alternative method. In general, such hyphenated separation techniques allow enhanced ionization efficiencies and LODs than MALDI-TOF-MS or direct infusion ESI-MS. The separation of target compounds as narrow peaks from the sample matrix avoids ion suppression and improves S/N. CZE-ESI-MS is widely used for the highly efficient separation and characterization of a number of biomolecules, including protein isoforms, glycoforms, oligomers, or drug-protein complexes [25–30]. In addition to high resolution capabilities, CZE has many other desirable characteristics, including a separation mechanism based on electrophoretic mobility (which is related to molecule charge and size ratio), low consumption of sample and reagents, reduced analysis times, instrumental simplicity, and full automation. Furthermore, experimental conditions can be adapted to run separations under non-denaturing conditions, in order to separate native protein oligomers. Nowadays, orthogonal acceleration-TOF mass spectrometers provide high mass accuracy and resolving

power, as well as the wide scanning ranges necessary for reliable characterization of large proteins and oligomers by CZE-ESI-TOF-MS [31–33].

In this paper, we describe a novel CZE-ESI-TOF-MS method for the detection and characterization of normal and variant forms of TTR in serum samples after IP from healthy controls and FAP-I patients. The results permit us to propose a simple, rapid, and reliable method to screen, diagnose, and follow-up patients with suspected TTR amyloidosis and to gain insight into TTR structure, to understand the mechanism underlying the disease.

2 Materials and methods

2.1 Chemicals and reagents

All the chemicals used in the preparation of buffers and solutions were of analytical reagent grade. Propan-2-ol ($\geq 99.9\%$), formic acid (HFor) (99.0%), acetic acid (HAc) (glacial), ammonia (25%), sodium hydrogenphosphate ($\geq 99.0\%$), sodium chloride ($\geq 99.5\%$), potassium dihydrogenphosphate ($\geq 99.0\%$), potassium chloride (99.0%), sodium hydroxide ($\geq 99.0\%$, pellets), and TTR ($\geq 95.0\%$) were purchased from Merck (Darmstadt, Germany). Rabbit antihuman TTR polyclonal antibody was supplied by Dako (Glostrup, Denmark). BSA ($\geq 99.0\%$) and ammonium acetate (NH_4Ac) ($\geq 99.9\%$) were provided by Sigma (St. Louis, MO, USA). PEG 8000 relative molecular mass (M_r) ($\sim 50\%$ in water) was purchased from Fluka (Buchs, Switzerland). Water with a conductivity value lower than $0.05 \mu\text{S}/\text{cm}$ was obtained using a Milli-Q water purification system (Millipore, Molsheim, France).

2.2 Electrolyte and sheath liquid solutions

The BGEs contained 1.0 M HAc (the measured pH was 2.30) or 10 mM NH_4Ac , adjusted to pH 7.00 with ammonia. Both BGEs were passed through a $0.45 \mu\text{m}$ nylon filter (MSI, Westboro, MA, USA) before analysis. The sheath liquid solution consisted of a mixture of 60:40 v/v propan-2-ol/water with a 0.05% v/v or 0.25% v/v of HFor, for the acidic and the neutral BGEs, respectively. The sheath liquid was degassed for 10 min by sonication before use. Solutions of 1.0% w/v of BSA in PBS (0.011 M sodium hydrogenphosphate, 0.0015 M potassium dihydrogenphosphate, 0.14 M sodium chloride, 0.0027 M potassium chloride, pH 7.20) and 5.0% v/v of PEG in water were used to evaluate passivation of the centrifugal filters [34].

2.3 Apparatus

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

in a cooled Rotanta 460 centrifuge (Hettich Zentrifugen, Tuttlingen, Germany) for centrifugation at controlled temperature (4 or 25°C). Agitation during sample incubation overnight was performed with a Vortex Genius 3 (Ika®, Staufen, Germany).

2.4 Procedures

2.4.1 Standard TTR solutions

An aqueous standard solution (1000 µg/mL) of TTR was prepared and stored in a freezer at –20°C when not in use. Excipients of low M_r were removed from the TTR standard solution by passage through 10 000 M_r cut-off cellulose acetate filters (Amicon Ultra-0.5, Millipore). The standard solution was centrifuged at 25°C for 10 min at 11 000 $\times g$, and the residue was washed three times for 10 min in the same way, with an appropriate volume of acidic or neutral BGE. The final residue was recovered by inverting the upper reservoir in a vial and spinning once more at a reduced centrifugal force (2 min at 300 $\times g$). Sufficient BGE (acidic or neutral) was added to adjust the TTR concentration to 1000 µg/mL. The 10 000 M_r filters were passivated before filtration with PEG solution to ensure optimum protein recoveries [34]. In order to passivate the filters, 500 µL of the passivation solution were pipetted into the reservoir of the centrifugal device, capped, and allowed to stand 24 h at room temperature. The reservoirs were then uncapped and rinsed thoroughly with tap water. A volume of 500 µL of purified water was added to remove any residue and the centrifuge was then spun to dead stop (30 min at 11 000 $\times g$). The remaining water was eliminated by inverting the reservoir in a vial and spinning once more at a reduced centrifugal force (2 min at 300 $\times g$).

2.4.2 Immunoprecipitation of serum samples

Human blood samples from two healthy controls and four FAP-I patients (one asymptomatic, one symptomatic (symptom onset was 4 years before blood collection), a liver-transplanted patient with the specific mutation (Met 30) (transplanted in 2013), and a iatrogenic patient originally without the mutation who was transplanted a liver from an FAP-I patient (transplanted 14 years before blood collection and symptom onset was 6 years before collection)) were kindly supplied by the Hospital Universitari de Bellvitge (HUB, Hospitalet de Llobregat, Spain). The assay was approved by the Ethics Committee of the HUB and written informed consent was obtained from all participants in the study. 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. After that, the blood was kept at 4°C for 12–16 h to improve the clot retraction. Then, the supernatant serum was separated from the clot with a Pasteur pipette and centrifuged at 1200 $\times g$ for 20 min at 4°C. Clear

serum was separated and aliquoted to store in a freezer at –20°C when not in use [35].

Fifty microliters of serum were incubated with the same volume of rabbit anti-TTR antibody overnight at 4°C in a shaker. After centrifugation of the sample at 9000 $\times g$ for 5 min at 4°C, the precipitate was washed twice with 100 µL of solution of 0.9% m/v sodium chloride in water (saline), and twice with 100 µL of purified water (by centrifugation as before). After the last centrifugation step, the supernatant was removed and the precipitate was dissolved by the addition of 50 µL of HAC 1.0 M. This sample was analyzed by CZE-ESI-TOF-MS with or without desalting by centrifugal filtration through 10 000 M_r cut-off cellulose acetate filters passivated with PEG (see the procedure for filtration of TTR standard solutions in Section 2.4.1). Alternatively, in order to separate the TTR from the antibody, the undesalted sample was passed through 100 000 M_r cut-off cellulose acetate filters passivated with PEG [34]. The sample was centrifuged at 11 000 $\times g$ for 10 min and washed with 20 µL of HAC 1.0 M for 8 min at the same centrifugal force. Sufficient HAC 1.0 M was added to the filtrate until the final volume was adjusted to 50 µL. TTR recovery by centrifugal filtration was evaluated before the experiments with serum samples. A standard TTR solution of 250 µg/mL was passed through 50 000 and 100 000 M_r cut-off cellulose acetate filters nonpassivated and passivated with BSA or PEG solution [34]. The filters were passivated as explained before in Section 2.4.1.

2.4.3 CZE-ESI-TOF-MS

All CZE-ESI-TOF-MS experiments were performed in an HP^{3D} CE system coupled with an orthogonal G1603A sheath-flow interface to a 6220 oaTOF LC/MS spectrometer (Agilent Technologies, Waldbronn, Germany). 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). ChemStation C.01.06 software (Agilent Technologies) was used for CE control and separation data acquisition (e.g. voltage, temperature, and current), and was run in combination with MassHunter B.04.00 workstation software (Agilent Technologies) for control, data acquisition, and analysis of the oaTOF mass spectrometer. m/z mass spectra were deconvoluted to obtain the M_r values and peak areas of the TTR forms. The deconvolution algorithm of maximum entropy computed the simplest zero charge mass spectra that could account for the observed m/z .

A 72 cm long (L_T) \times 75 µm id \times 360 µm od bare fused-silica capillary supplied by Polymicro Technologies (Phoenix, AZ, USA) was used for the separations. All capillary rinses were performed at high pressure (930 mbar). A new capillary was set every time the BGE pH was changed. New capillaries were flushed with 1.0 M NaOH (20 min), water (15 min), and BGE (30 min, acidic or neutral). The system was finally equilibrated by applying the separation voltage (25 kV) for 15 min. Between workdays, the capillary was conditioned by rising successively with 0.1 M NaOH (5 min), water (10 min), and

BGE (15 min). Both activation and conditioning procedures were performed off-line to avoid the unnecessary entrance of NaOH into the MS system. Between runs, the capillary was rinsed for 2 min with HAc 1.0 M, 1 min with water and 1 min with BGE (acidic or neutral). The working BGE was refreshed after every analysis to ensure optimum separation repeatability. Samples were hydrodynamically injected at 50 mbar for 10 s. A separation voltage of 25 kV (normal polarity, anode at the injection capillary end) was applied for electrophoretic separations at 25°C.

The oaTOF mass spectrometer was operated under optimum conditions in positive mode using the following parameters: capillary voltage 4000 V, drying gas temperature 300°C, drying gas flow rate 4 L/min, nebulizer gas 7 psig, fragmentor voltage 325 V, skimmer voltage 80 V, OCT 1 RF Vpp voltage 300 V. Data were collected in profile at 1 spectrum/s between 100 and 3200 m/z , with the mass range set to high resolution mode (4 GHz). The reference mass correction for internal mass recalibration was disabled for all the experiments for optimum sensitivity [33]. A standard tune and an external mass calibration were performed daily at the beginning of the day following the manufacturer instructions using the typical LC-MS sprayer and ESI-L tuning mix (Agilent Technologies).

2.4.4. Quality parameters in CZE-ESI-TOF-MS

All quality parameters in CZE-ESI-TOF-MS were calculated from data obtained by measuring the peak area and migration time from the extracted ion electropherogram (EIE) of wild-type (normal) TTR with the different modifications (considering the m/z of the most abundant molecular ions, i.e. ions with charges +16, +15, +14, +13). The deconvoluted average M_r and abundance of the detected forms for the normal TTR variant are shown in Table 1 (TTR standard 1000 $\mu\text{g/mL}$). Repeatability studies were performed by analysing ten times in one day a standard solution of TTR at a concentration of 250 $\mu\text{g/mL}$ ($n = 10/1$ day). This parameter was calculated as %RSD of peak areas and migration times (t_m). An estimation of the LODs was obtained by analysing low-concentration solutions of a standard of TTR until no peaks were detected (10 $\mu\text{g/mL}$). The S/N values were calculated referred to peak areas using the software provided by the manufacturer of the mass spectrometer. The calibration range was established by injecting standard solutions of TTR at concentrations between 50 and 1000 $\mu\text{g/mL}$.

3 Results and discussion

3.1 Development of a method for the analysis of TTR by CZE-ESI-TOF-MS

CZE-ESI-TOF-MS was evaluated by analysing standard solutions of a commercially available normal TTR, after removal of excipients by centrifugal filtration, in positive ESI mode

with HAc 1.0 M (pH 2.30) and 10 mM NH_4Ac (pH 7.00) BGEs. Based on our previous studies with the homodimeric enzyme superoxide dismutase I, the acidic BGE was supposed to disrupt the native tetrameric structure generating the monomer (MO) while providing enhanced sensitivity, because it was more volatile and allowed maximum protonation of the amino acid ionizable groups [33]. In contrast, a neutral BGE was supposed to be more suitable for detecting the TTR native tetrameric structure, but with lower S/Ns. Figure 1A and B shows the total ion electropherograms (TIEs) (i), the mass spectra (ii) and the deconvoluted mass spectra (iii, iv) of a 1000 $\mu\text{g/mL}$ standard solution of normal TTR using acidic and neutral BGEs, respectively. The electropherogram shows a broad single peak for the neutral BGE (Fig. 1B), while the single peak was sharper for the acidic BGE (Fig. 1A). As expected, the sensitivity was better with the acidic than with the neutral BGE (compare the y-axis of both TIEs and mass spectra). Even with the neutral BGE, the percentage of HFor in the sheath liquid needed to be increased up to 0.25% v/v to detect the protein. At acidic pH, the analysis time was slightly shorter because, although the EOF was lower, the effective electrophoretic mobility of TTR was high and directed toward the cathode ($pI = 5.4$). A careful analysis of the cluster of multiply charged ions obtained in the m/z spectra of both electrophoretic peaks indicated the presence of protein MO with the acidic BGE (with m/z of the most abundant molecular ion corresponding to the MO with charge +15), and the presence of protein MO and dimer (DI) with the neutral BGE (this is probably the reason for the wider electrophoretic peak). As can be observed in the different ion labels of the mass spectrum of Fig. 1B (see the zoomed area), the ions with nominal m/z values of 2314 and 2777 were the summed contribution of the TTR MO and DI, while the ion of 2524 was exclusively detected for the TTR DI (+11 DI). Unfortunately, the DI ions were detected with very low abundance and the presence of the tetramer could not be demonstrated by scanning until 3200 m/z . This was probably due to tetramer disruption during detection, because of the acidic sheath liquid or the vacuum-pressure inside the mass spectrometer. However, as indicated before, a sheath liquid with 0.25% v/v of HFor was the minimum amount of acid necessary to achieve appropriate sensitivity in positive ion mode with the neutral BGE, while the vacuum pressure in the first stages of our mass spectrometer could not be tuned by the operator. Figure 1A and B (iii) shows the deconvoluted mass spectra for the MO using acidic and neutral BGEs, respectively. As can be observed, the TTR isoforms and PTMs for the MO were the same in both cases. Table 1 shows the $M_{r \text{ exp}}$, the $M_{r \text{ theo}}$, and the mass error (E_r) for the detected monomeric TTR forms. The main PTM detected corresponded to TTR showing a mixed disulfide with the amino acid cysteine at position 10 of the sequence (TTR-Cys, $M_{r \text{ theo}} = 13\,880.4022$, Table 1). The rest of the PTMs and isoforms corresponded to the free TTR (Free-TTR, $M_{r \text{ theo}} = 13\,761.2640$), the phosphorylated or sulfonated TTR (TTR-Phosphorylated, $M_{r \text{ theo}} = 13\,841.2439$ and TTR-Sulfonated, $M_{r \text{ theo}} = 13\,841.3283$, respectively), the dehydroxylated or the

Table 1. Theoretical and deconvoluted average M_r , abundance, repeatability, calibration range, and LOD for the detected isoform and PTMs of a TTR standard solution by CZE-ESI-TOF-MS using HAC 1.0 M as BGE

N	Monomer transthyretin forms	% Abundance ^{a)}	Deconvoluted average M_r		E_r ^{b)} (ppm)	Peak area Repeatability 250 $\mu\text{g/mL}$ ($n = 10$), %RSD	Migration times (t_m)		Calibration range: 50–1000 $\mu\text{g/mL}$, $r^2 > 0.991$	LOD ^{c)} ($\mu\text{g/mL}$)
			Theoretical	Experimental			Average ($n = 10$)	Repeatability 250 $\mu\text{g/mL}$ ($n = 10$), %RSD		
1	TTR-Cys	100	13 880.4022	13 880.8583	33	9	10.5	1.2	$A = 9 \times 10^3 C + 1 \times 10^5$	<50 (S/N = 471)
2	Free-TTR	92	13 761.2640	13 761.8138	40	6	10.8	1.7	$A = 9 \times 10^3 C + 2 \times 10^5$	<50 (S/N = 407)
3	TTR-Phosphorylated	49	13 841.2439	13 841.5633	23	9	10.8	1.1	$A = 4 \times 10^3 C + 1 \times 10^5$	<50 (S/N = 117)
4	TTR-Sulfonated	49	13 841.3283	13 841.5633	17	9	10.8	1.1	$A = 4 \times 10^3 C + 1 \times 10^5$	<50 (S/N = 115)
	TTR-Dehydroxylated or TTR-Sulfinic	37	13 793.2628	13 793.5782	23	12	10.8	1.2	$A = 3 \times 10^3 C + 1 \times 10^5$	<50 (S/N = 100)
5	(10) C-G	37	13 715.1713	13 715.7959	46	7	10.4	1.1	$A = 4 \times 10^3 C + 4 \times 10^4$	<50 (S/N = 97)

a) The % abundance was calculated normalizing to the area value of the most abundant form.

b) Relative error (E_r) was calculated in ppm as: $(M_{r, \text{exp}} - M_{r, \text{theo}})/M_{r, \text{theo}} \times 10^6$ (exp = experimental and theo = theoretical).

c) Concentration was referred to the total amount of protein in the standard solution. The S/N for a certain form at this concentration was calculated from the EIE using the software provided by the manufacturer of the mass spectrometer.

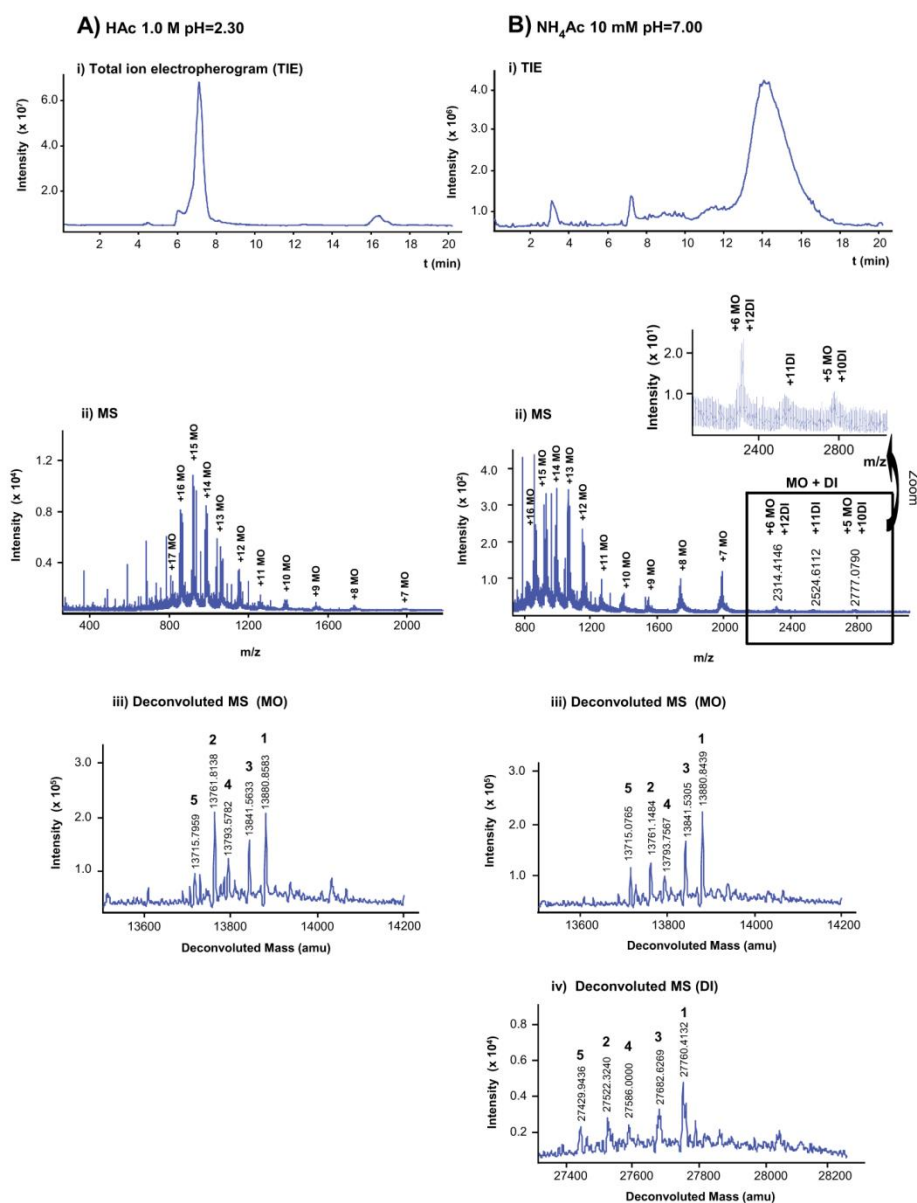


Figure 1. CZE-ESI-TOF-MS for a TTR standard at a concentration of 1000 $\mu\text{g/mL}$ using (A) HAC 1.0 M (pH 2.30) and (B) 10 mM NH₄Ac (pH 7.00 adjusted with ammonia). (MO, monomer; DI, dimer) (1: TTR-Cys, 2: Free-TTR, 3: TTR-Phosphorylated or TTR-Sulfonated, 4: TTR-Dehydroxylated or TTR-Sulfinic, 5: (10) C-G)).

conjugated cysteine sulfinic acid TTR (TTR-Dehydroxylated or TTR-Sulfinic, $M_{r\text{theo}} = 13\,793.2628$) and the isoform caused by the single amino acid substitution of a cysteine for glycine at position 10 ((10) C-G, $M_{r\text{theo}} = 13\,715.1713$ Da). As can be observed, the accuracy of the M_r determination did not allow differentiation between the TTR-Phosphorylated and TTR-Sulfonated forms (peak 3, Fig. 1A and B (iii)) and Table 1) (E_r were 23 and 17 ppm, respectively, Table 1). Similarly, we could not differentiate between TTR-Dehydroxylated or TTR-Sulfinic that were isobars and were not separated by electrophoresis. Especially in these last two cases, reliability of the identification would improve running CZE-MS/MS experiments with an appropriate mass spectrometer. The results for the commercial TTR standard solution agree with those reported by other authors [11, 13–15]. TTR has been proven to be predominantly the cysteine-conjugated and free forms by ESI-MS, although other isoforms and PTMs have also been described (TTR-Phosphorylated, TTR-Sulfonated, TTR-Dehydroxylated, TTR-Sulfinic, (10) C-G, and other PTMs observed when the cysteine residue on position 10 makes a mixed disulfide with the peptide cysteinyl-glycine or the peptide glutathione) [11, 13–15]. With regard to the neutral BGE and the DI (Fig. 1B (iv)), the detected forms were the same as for the MO considering a homodimer. No information has been previously reported about the main isoforms and PTMs observed in the TTR DI or tetramer.

The use of a neutral BGE is generally advantageous when it is necessary to work under nondenaturing conditions. However, this was not the case with TTR because, as explained before, the tetrameric native structure was not detected. Furthermore, the concentration of TTR in serum samples is low enough to require the acidic BGE that produces the best sensitivity. In addition, the immunoprecipitate from serum samples is in general dissolved by the addition of an acidic solution (HAc 1.0 M), which could also disrupt the tetramer. Therefore, the acidic BGE was selected for subsequent experiments. Table 1 summarizes the %RSD obtained for the repeatability of the peak areas and migration times ($n = 10/1$ day), LODs, and calibration ranges for the normal TTR using acidic BGE. As mentioned in Section 2.4.4, the differentiation between forms was achieved measuring the peak areas and migration times from the EIEs. The %RSD values for the repeatability of the different TTR forms ranged between 6 and 12% for peak areas, and 1.1 and 1.7% for migration times (t_m). The repeatability for the peak areas was good, because protein adsorption was reduced by the appropriate balance between silanol ionization and protein charge at this pH value. The average values of t_m show that there was only a slight separation between the TTR forms (e.g. 0.3 min between TTR-Cys and Free-TTR, Table 1). Area versus concentration regression lines were established in all cases over the calibration range between 50 and 1000 $\mu\text{g/mL}$ ($r^2 > 0.99$) and the LODs for all the identified forms were lower than 50 $\mu\text{g/mL}$. At 50 $\mu\text{g/mL}$ the S/N ratios were higher than 10 in all cases, and the highest values were obtained for the TTR-Cys and the Free-TTR, Table 1. The LODs obtained using

the neutral BGE were as indicated before higher, at around 250 $\mu\text{g/mL}$.

3.2 Analysis of TTR in serum by CZE-ESI-TOF-MS

A sample pretreatment based on IP of TTR with a polyclonal anti-TTR rabbit antibody was applied to isolate the protein from serum samples. Various serum/antibody volume ratios were tested, from 50:20 to 50:100 μL , and it was observed that precipitation of the TTR-antibody complex was quantitative from a volume ratio higher than 50:20. First, we directly analyzed the TTR-antibody mixture obtained after dissociating the antigen-antibody complex with HAc 1.0 M. However, the presence of different components from the solutions required for the IP had a negative influence on the current stability of the electrophoretic separation, and breakdown occurred at the beginning of the runs, probably due to precipitation of proteins with partial capillary clogging. To solve this problem, the TTR-antibody mixture was passed through 10 000 M_r cut-off cellulose acetate filters passivated with PEG to remove salts and other low M_r compounds [34]. Figure 2A shows the TIE obtained for a serum sample from a healthy control after IP and centrifugal filtration with 10 000 M_r cut-off filters. The three separated peaks corresponded to the polymeric gel used in the blood collection tubes as a clot activator for the serum separation, the anti-TTR antibody and the TTR (see the mass spectra and the deconvoluted mass spectra in the Supporting Information Fig. 1). The deconvoluted mass spectrum for the TTR (Fig. 2A) shows that now only the most abundant TTR forms were detected (TTR-Cys and Free-TTR). This was probably because the TTR concentration in serum-treated samples after the IP method and centrifugal filtration was low, ranging between 50 and 150 $\mu\text{g/mL}$, depending on the individual (see explanation below).

3.2.1 Optimization of centrifugal filtration of immunoprecipitated samples

When we applied this filtration procedure, the presence of the antibody and the polymeric gel activator continued to have negative effects on the CZE current and stability, so that the method was not reproducible. For this reason, we used 100 000 M_r cut-off centrifugal filters to separate TTR from the antibody prior to CZE-ESI-TOF-MS. Filtration with 100 000 M_r cut-off centrifugal filters has been applied for this purpose in TTR immunoprecipitation procedures with different degrees of success. While some authors were not able to separate TTR from the antibody with filters from a certain manufacturer [18, 19, 24], others demonstrated that separation was possible with filters from a different manufacturer, but no recoveries were reported [17]. In the present study, to evaluate the TTR recovery in the filtration step with filters of different M_r cut-off, a standard TTR solution of 250 $\mu\text{g/mL}$

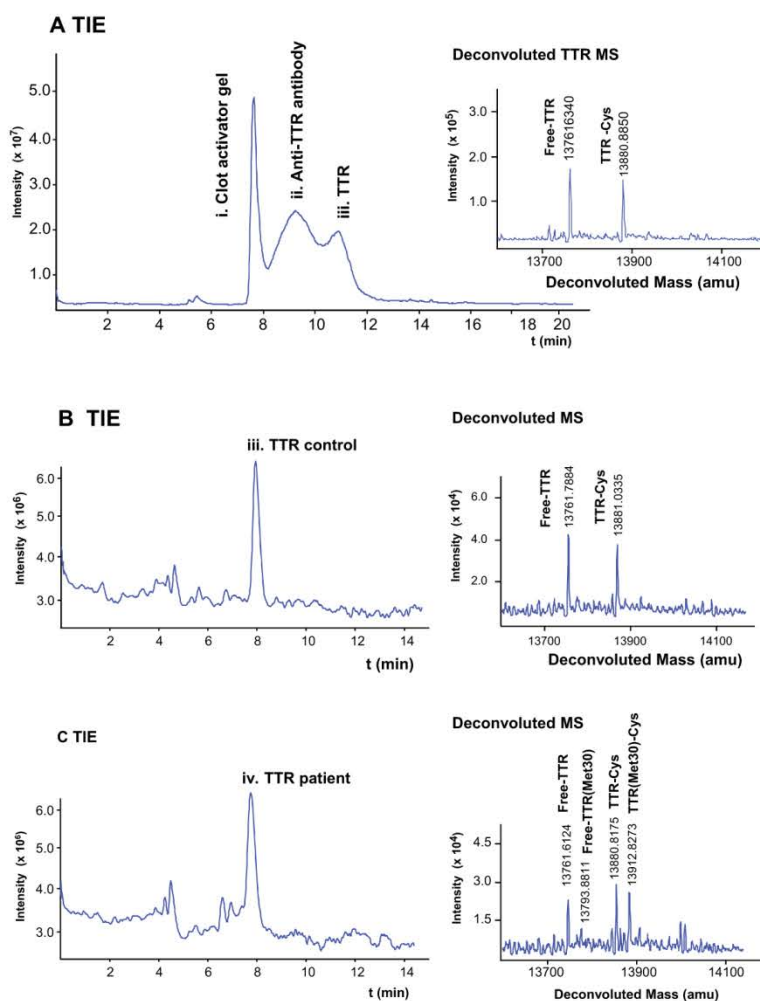


Figure 2. CZE-ESI-TOF-MS for (A) a control sample using the IP method followed by centrifugal filtration with 100 000 M_r filters and (B) a control sample and (C) an FAP-I sample using the IP method followed by centrifugal filtration with 100 000 M_r filters.

was passed through 50 000 and 100 000 M_r cut-off filters and then analyzed by CZE-ESI-TOF-MS. The filters were used without passivation and after passivation with BSA or PEG, because some manufacturers recommend passivation when handling highly diluted analyte solutions to prevent analyte adsorption on the plastic walls [34]. When we used 50 000 M_r nonpassivated filters, and 50 000 M_r filters passivated with PEG and BSA, TTR was detected in the filtrates with recovery values of 43% ($n = 3$, %RSD = 5), 55% ($n = 3$, %RSD = 1), and 61% ($n = 3$, %RSD = 4), respectively. These recovery values were lower in all cases than with 100 000 M_r filters (55% ($n = 3$, %RSD = 2), 76% ($n = 3$, %RSD = 3), and 62% ($n = 3$, %RSD = 5), respectively), probably because TTR was retained to a larger extent in the membrane of smaller pore size. In all cases, recoveries were improved with passivation. The 100 000 M_r filters passivated with PEG were used in the rest of the study, because the recovery values were the best (76%), hence ensuring enough TTR concentration for a reliable characterization. Figure 2B and C shows the TIE and deconvoluted mass spectra obtained for a serum sample from a healthy control and an FAP-I patient (asymptomatic) after IP and centrifugal filtration with 100 000 M_r cut-off.

As shown, the observed peak corresponded to monomeric TTR, while the antibody and the clot activator gel for the serum separation were removed from the sample using the 100 000 M_r filters. The deconvoluted mass spectrum for the healthy control (Fig. 2B) shows that the main forms detected were TTR-Cys and Free-TTR, as in the IP method followed by centrifugal filtration with 100 000 M_r cut-off (Fig. 2A). In the case of the FAP-I patient (asymptomatic) (Fig. 2C), the main forms corresponded to the TTR-Cys and the Free-TTR for the normal TTR, and the TTR(Met30)-Cys and the Free-TTR(Met30) for the abnormal (or mutant) TTR, with a M_r 32.07 higher than the normal forms of TTR ($\Delta M_r = +32.07$). Repeatability in the TTR analysis after the serum sample pretreatment was similar to that obtained for the TTR standard solution (Table 1) and, for example, values ranging between 1.0–1.4% and 7–11% were obtained for migration times and peak areas of normal and mutant forms from the asymptomatic FAP-I patient ($n = 10$). The reproducibility of the sample pretreatment method was also good, and for the asymptomatic FAP-I patient values ranged between 3.6–3.7% and 3–11% ($n = 3$, three independent samples), for migration times and peak areas, respectively.

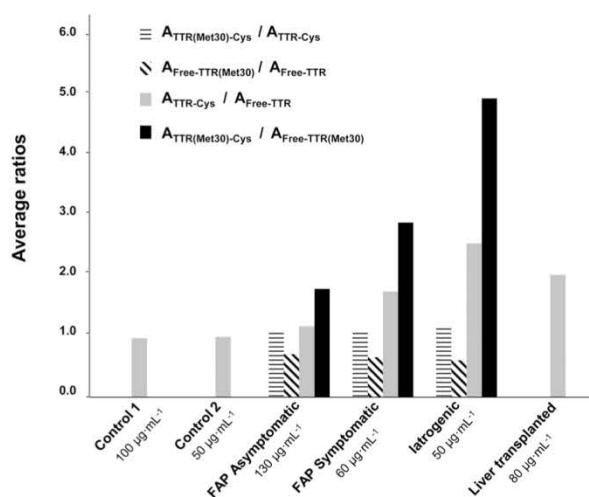


Figure 3. Bar graph with the average ratios between the areas of the detected normal and FAP-I protein forms ($n = 3$). The area was measured from the extracted ion electropherogram (EIE) obtained from the most abundant molecular ions (+16, +15, +14, +13). The total concentration of TTR appears as insets (it was estimated by comparing the area of the TTR peak observed in the total ion electropherograms (TIEs) from TTR standards and serum samples).

3.2.2 Analysis of clinical samples

The serum samples from two healthy controls and four patients suffering from FAP-I (one asymptomatic, one symptomatic, a liver-transplanted patient with the specific mutation, and an iatrogenic patient originally without the mutation who had a liver transplanted from an FAP-I patient) were analyzed in triplicate by CZE-ESI-TOF-MS. It has been reported that the total concentration of TTR in serum and cerebrospinal fluid is variable and depends on the individual [19]. The total concentration of TTR in our samples ranged between 50 and 130 µg/mL (see the insets on the x-axis of Fig. 3). Figure 3 shows a bar graph with the average ratios between the areas of FAP-I forms and normal forms of TTR ($TTR(Met30)-Cys/TTR-Cys$ and $Free-TTR(Met30)/Free-TTR$), as well as the ratio between the areas of $TTR-Cys$ and $Free-TTR$ ($TTR-Cys/Free-TTR$ and $TTR(Met30)-Cys/Free-TTR(Met30)$). As expected, mutant $Free-TTR(Met30)$ and $TTR(Met30)-Cys$ were found in the asymptomatic, symptomatic, and iatrogenic FAP-I patients, whereas only the normal forms were detected in the liver-transplanted, which indicates the effectiveness of the treatment. Further experiments would be necessary to confirm if the optimized method could be used to monitor the progression of transplanted patients and, eventually, to improve patient prognosis, because this liver-transplanted patient with the specific mutation had received the transplant only 10 months before the blood was collected, while the iatrogenic FAP-I patient received the transplant a long time earlier (14 years and the patient already showed FAP-I symptoms after 8 years of transplantation).

There is some controversy about the importance of TTR oxidized forms (e.g. $TTR-Cys$), as biomarkers of the oxidative

stress involved in amyloid deposit formation in FAP-I and other amyloid diseases [19, 36–39]. The level of $TTR-Cys$ in plasma samples of FAP-I symptomatic patients was found elevated when compared to the $Free-TTR$ [36, 37]. Similarly, the concentration of these oxidized TTR forms in CSF were recently found elevated in Alzheimer's disease and mild cognitive impairment patients [39], while a previous study found less TTR oxidation in a similar study with different techniques [38]. In this preliminary study, as can be observed in Fig. 3, $TTR-Cys/Free-TTR$ ratios for the normal (gray bars) and mutant TTR (black bars) were higher in symptomatic patients than in controls, especially for mutant TTR in the iatrogenic patient (last black bar). Our results also indicate that the $TTR(Met30)-Cys/Free-TTR(Met30)$ ratio (black bars) could be related to the time from symptom onset, because it was lowest for the FAP-I asymptomatic patient, medium for the symptomatic FAP-I patient (symptom onset was 4 years before blood collection), and highest for the iatrogenic patient (symptom onset was 6 years before blood collection). With regard to the ratios between mutant and normal forms (slashed bars) in the asymptomatic, symptomatic, and iatrogenic patients, they were in all cases around 0.7 and 1.1 for the free and the cysteine-conjugated forms, respectively. As the values are similar in the three cases, both ratios would be maintained during FAP-I progression. Again, this is a preliminary study and further analyses of a larger set of samples would be necessary to determine whether the optimized method is useful to reliably gain insight into FAP-I. The trigger for the TTR misfolding and aggregation could be related to the $TTR-Cys$ oxidized form; but also to other TTR forms found at lower concentrations that are not detected with this method; or an external factor that would alter the stability of the tetramer.

4 Concluding remarks

We established a novel CZE-ESI-TOF-MS method to detect wild-type TTR in standard solutions with several PTMs, and an extra minor isoform ((10) C-G) using an acidic and a neutral BGE. Although the neutral BGE was supposed to be more suitable for detecting the TTR native structure, the tetramer was disrupted because of the acidic sheath liquid or the high vacuum inside the mass spectrometer. We also optimized a simple and reliable sample treatment method based on IP followed by centrifugal filtration to purify TTR from serum before the CZE-ESI-TOF-MS analysis of samples from control individuals and FAP-I patients. No mutant TTR forms were detected in the liver-transplanted patient or in the control samples. The main mutant TTR forms ($TTR(Met30)-Cys$ and $Free-TTR(Met30)$) were only found in the asymptomatic, the symptomatic, and the iatrogenic patient. The $TTR(Met30)-Cys/Free-TTR(Met30)$ ratio was higher than any other protein ratio in all control and symptomatic patients, and it could be related to the time from symptom onset. In contrast, the ratios between mutant and normal forms of TTR were similar in all the symptomatic patients, which shows that these ratios would be maintained during FAP-I progression. Further

analyses of a larger set of samples would be necessary to determine whether the optimized method can help to gain insight into FAP-I.

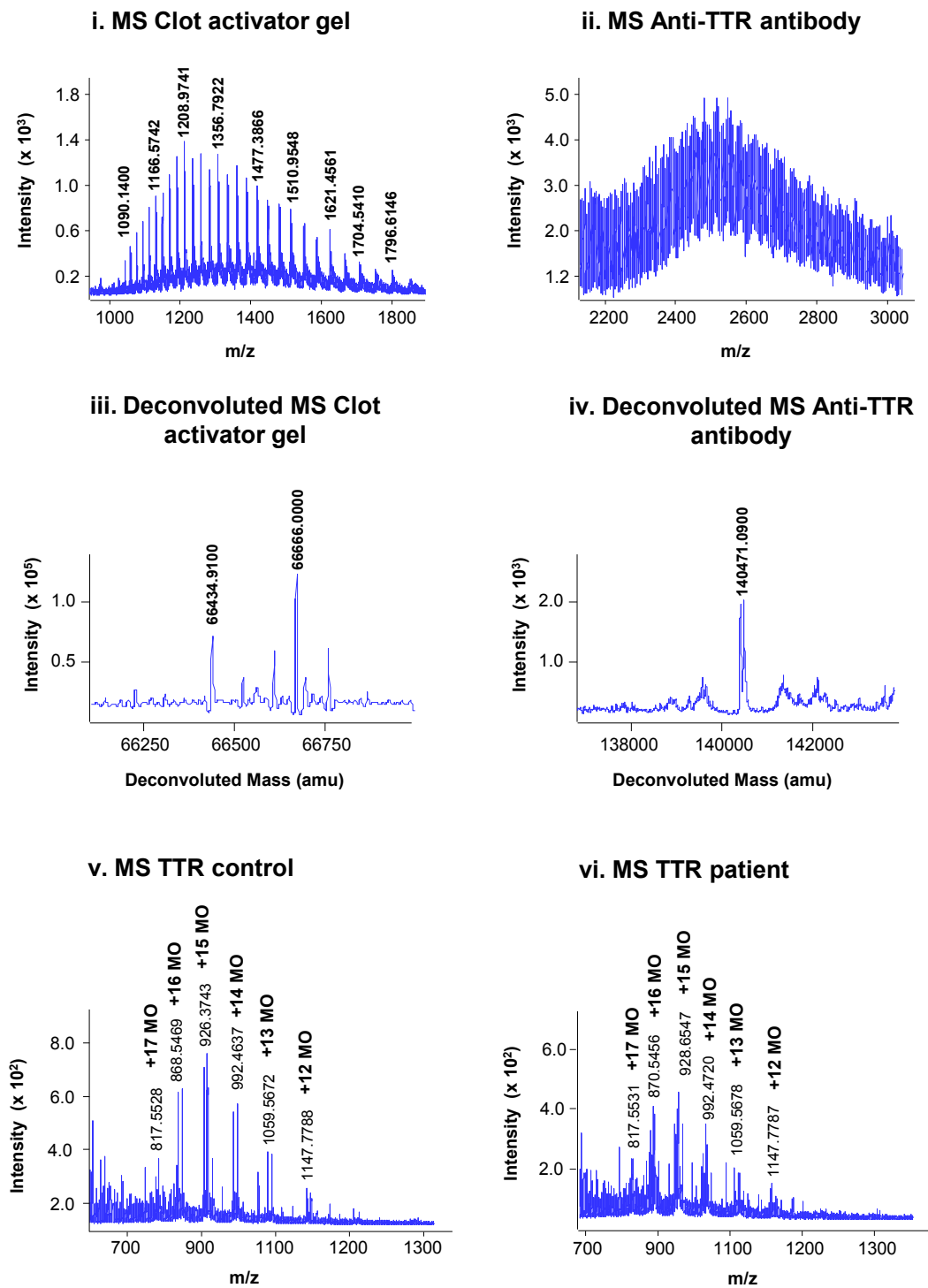
Laura Pont acknowledges the Spanish Ministry of Economy and Competitiveness for a FPI fellowship. This study was supported by a grant from the Spanish Ministry of Education and Science (CTQ2011-27130). We also thank Dr. C. Casasnovas and Dr. M. A. Alberti (Hospital Universitari de Bellvitge, HUB, Hospitalet de Llobregat, Spain) for providing the blood samples. The assay was approved by the Ethics Committee of the HUB and written informed consent was obtained from all participants in the study.

The authors have no conflict of interest to declare.

5 References

- [1] Shimizu, A., Nakanishi, T., Kishikawa, M., Miyazaki, A., *J. Chromatogr. B* 2002, 776, 15–30.
- [2] Shimizu, A., Nakanishi, T., Miyazaki, A., *Mass Spectrom. Rev.* 2006, 25, 686–712.
- [3] Pepys, M. B., Herbert, J., Hutchinson, W. L., Tennent, G. A., Lachmann, H. J., Gallimore, J. R., Lovat, L. B., Bartfai, T., Alanine, A., Hertel, C., Hoffmann, T., Jakob-Roetne, R., Norcross, R. D., Kemp, J. a, Yamamura, K., Suzuki, M., Taylor, G. W., Murray, S., Thompson, D., Purvis, A., Kolstoe, S., Wood, S. P., Hawkins, P. N., *Nature* 2002, 417, 254–259.
- [4] Dobson, C. M., *Nature* 2003, 426, 884–890.
- [5] Loizos, S., Shiakalli Chrysa, T., Christos, G. S., *Semin. Ultrasound CT MRI* 2014, 35, 225–239.
- [6] Petersen, R. B., Goren, H., Cohen, M., Richardson, S. L., Tresser, N., Lynn, A., Gali, M., Estes, M., Gambetti, P., *Ann. Neurol.* 1997, 41, 307–313.
- [7] Hörnberg, A., Eneqvist, T., Olofsson, A., Lundgren, E., Sauer-Eriksson, E., *J. Mol. Biol.* 2000, 302, 649–669.
- [8] Johnson, S. M., Connelly, S., Fearn, C., Powers, E. T., Kelly, J. W., *J. Mol. Biol.* 2012, 421, 185–203.
- [9] Zheng, W., Lu, Y. M., Lu, G. Y., Zhao, Q., Cheung, O., Blaner, W. S., *Toxicol. Sci.* 2001, 61, 107–114.
- [10] Bergen, H. R., Zeldenrust, S. R., Naylor, S., *Amyloid J. Protein Fold. Disord.* 2003, 10, 190–197.
- [11] Poulsen, K., Bahl, J. M. C., Tanassi, J. T., Simonsen, A. H., Heegaard, N. H. H., *Methods* 2012, 56, 284–292.
- [12] Connors, L. H., Lim, A., Prokaeva, T., Roskens, V. A., Costello, C. E., *Amyloid J. Protein Fold. Disord.* 2003, 10, 160–184.
- [13] Terazaki, H., Ando, Y., Suhr, O., Ohlsson, P. I., Obayashi, K., Yamashita, T., Yoshimatsu, S., Suga, M., Uchino, M., Ando, M., *Biochem. Biophys. Res. Commun.* 1998, 249, 26–30.
- [14] Nakanishi, T., Sato, T., Sakoda, S., Yoshioka, M., Shimizu, A., *Biochim. Biophys. Acta* 2004, 1698, 45–53.
- [15] Gericke, B., Raila, J., Sehoul, J., Haebel, S., Könsgen, D., Mustea, A., Schweigert, F. J., *BMC Cancer* 2005, 5, 133–141.
- [16] Falk, R. H., Comenzo, R. L., Skinner, M., *N. Engl. J. Med.* 1997, 337, 898–909.
- [17] Lim, A., Prokaeva, T., McComb, M. E., O'Connor, P. B., Théberge, R., Connors, L. H., Skinner, M., Costello, C. E., *Anal. Chem.* 2002, 74, 741–751.
- [18] Ando, Y., Ohlsson, P., Suhr, O., Nyhlin, N., Yamashita, T., Holmgren, G., Danielsson, A., Sandgren, O., Uchino, M., Ando, M., *Biochem. Biophys. Res. Commun.* 1996, 228, 480–483.
- [19] Ando, Y., Suhr, O., Yamashita, T., Ohlsson, P. I., Holmgren, G., Obayashi, K., Terazaki, H., Mambule, C., Uchino, M., Ando, M., *Neurosci. Lett.* 1997, 238, 123–126.
- [20] Pomfret, E. A., Lewis, W. D., Jenkins, R. L., Bergethon, P., Dubrey, S. W., Reisinger, J., Falk, R. H., Skinner, M., *Transplantation* 1991, 65, 918–925.
- [21] Okamoto, S., Wixner, J., Obayashi, K., Ando, Y., Uchino, M., Suhr, O. B., *Liver Transplant.* 2009, 15, 1229–1235.
- [22] Théberge, R., Connors, L., Skinner, M., Skare, J., Costello, C. E., *Anal. Chem.* 1999, 71, 452–459.
- [23] Théberge, R., Connors, L. H., Skinner, M., Costello, C. E., *J. Am. Soc. Mass Spectrom.* 2000, 11, 172–175.
- [24] Heegaard, N. H. H., Hansen, M. Z., Sen, J. W., Christiansen, M., Westermarck, P., *J. Sep. Sci.* 2006, 29, 371–377.
- [25] Hernández-Borges, J., Neusüss, C., Cifuentes, A., Pelzing, M., *Electrophoresis* 2004, 25, 2257–2281.
- [26] El Rassi, Z., *Electrophoresis* 2010, 31, 174–191.
- [27] Haselberg, R., de Jong, G. J., Somsen, G. W., *Electrophoresis* 2013, 34, 99–112.
- [28] Righetti, P. G., Sebastiano, R., Citterio, A., *Proteomics* 2013, 13, 325–340.
- [29] Kašička, V., *Electrophoresis* 2014, 35, 69–95.
- [30] Zhao, S. S., Chen, D. D. Y., *Electrophoresis* 2014, 35, 96–108.
- [31] Staub, A., Schappler, J., Rudaz, S., Veuthey, J.-L., *Electrophoresis* 2009, 30, 1610–1623.
- [32] Giménez, E., Ramos-Hernan, R., Benavente, F., Barbosa, J., Sanz-Nebot, V., *Anal. Chim. Acta* 2012, 709, 81–90.
- [33] Borges-Alvarez, M., Benavente, F., Barbosa, J., Sanz-Nebot, V., *Electrophoresis* 2012, 33, 2561–2569.
- [34] Pont, L., Benavente, F., Barbosa, J., Sanz-Nebot, V., *J. Sep. Sci.* 2013, 36, 3896–3902.
- [35] Ammerlaan, W., Trezzi, J.P., Lescuyer, P., Mathay, C., Hiller, K., Betsou, F., *Biopreserv. Biobank.* 2014, 12, 269–280.
- [36] Suhr, O. B., Ando, Y., Ohlsson, P. I., Olofsson, A., Andersson, K., Lundgren, E., Ando, M., Holmgren, G., *Eur. J. Clin. Invest.* 1998, 28, 687–692.
- [37] Suhr, O. B., Svendsen, I. H., Ohlsson, P.-I., Lendoire, J., Trigo, P., Tashima, K., Ranlov, P. J., Ando, Y., *Amyloid Int. J. EXD. Clin. Invest.* 1999, 6, 187–191.
- [38] Biroccio, A., Del Boccio, P., Panella, M., Bernardini, S., Di Ilio, C., Gambi, D., Stanzione, P., Sacchetta, P., Bernardi, G., Martorana, A., Federici, G., Stefani, A., Urbani, A., *Proteomics* 2006, 6, 2305–2313.
- [39] Poulsen, K., Bahl, J. M., Simonsen, A. H., Hasselbalch, S. G., Heegaard, N. H., *Clin. Proteomics* 2014, 11, 12.

Supplementary Figures



Supplementary Figure S-1. Mass spectra (i, ii, v, vi) and deconvoluted mass spectra (iii, iv) of the peaks of Figure 2.



Contents lists available at ScienceDirect

Journal of Chromatography A

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

Comparison of capillary electrophoresis and capillary liquid chromatography coupled to mass spectrometry for the analysis of transthyretin in human serum[☆]



Laura Pont^a, Kader Poturcu^b, Fernando Benavente^a, José Barbosa^a,
Victoria Sanz-Nebot^{a,*}

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

^b Department of Chemistry, Suleyman Demirel University, Isparta, Turkey

ARTICLE INFO

Article history:

Received 30 November 2015

Received in revised form 2 February 2016

Accepted 19 March 2016

Available online 24 March 2016

Keywords:

Capillary electrophoresis
Capillary liquid chromatography
Immunoprecipitation
Magnetic beads
Mass spectrometry
Transthyretin

ABSTRACT

Capillary electrophoresis and capillary liquid chromatography coupled to mass spectrometry (CE-MS and CapLC-MS, respectively) are nowadays very suitable techniques for the separation and characterization of intact proteins in biological fluids. In this paper, we compare the performance of both techniques for the analysis of transthyretin (TTR), which is a homotetrameric protein (relative molecular mass (M_r) ~56,000) involved in different types of amyloidosis. Furthermore, it is also presented a novel sample pretreatment based on immunoprecipitation (IP) using Protein A Ultrarapid AgaroseTM (UAPA) magnetic beads (MBs) to purify TTR from serum samples. This novel IP based on MBs allowed the detection of TTR monomeric proteoforms that were not possible to analyze by conventional IP in solution. In addition, UAPA MBs provided many other desirable advantages including higher selectivity and minimal unspecific binding of other proteins. CE-MS and CapLC-MS were applied to analyze serum samples from healthy controls and familial amyloidotic polyneuropathy type I (FAP-I) patients, who suffered from the most common hereditary systemic amyloidosis. Both techniques allowed detecting the same TTR proteoforms, including the mutant TTR (Met 30) variant (variation in relative molecular mass (ΔM_r) was +32.07, from wild-type TTR). Migration/retention times and relative quantitation of the different proteoforms were similar and reproducible in both cases, but the limits of detection (LODs) achieved by CE-MS were slightly lower (2–2.5-fold). Some other differences were also found on separation selectivity (migration orders and separation of antibody), peak efficiency, total analysis time, calibration ranges and experimental M_r accuracy.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Measurements of the content of normal and variant forms (proteoforms) of serum proteins are important for clinical diagnosis and for elucidation of the pathogenesis of various diseases [1–4]. Research progress on protein isolation and purification, as well as the introduction of highly sensitive analytical methods, in particular mass spectrometry, have been crucial for an increase in the accuracy and reproducibility of protein profiling in total

blood serum to identify post-translational modifications (PTMs) and mutations at the protein sequence level (isoforms) [5–8].

Nowadays, capillary electrophoresis mass spectrometry (CE-MS) and capillary liquid chromatography mass spectrometry (CapLC-MS) are very suitable techniques for the separation and characterization of intact proteins in biological fluids [9–12]. Both of these miniaturized techniques provide many advantages, such as low consumption of sample and reagents in contrast to conventional LC or electrophoresis methods, reduced ion suppression and better sensitivity than direct MS analysis, good separation efficiency, recovery, short analysis time, more robustness than nanoLC or microchip technologies, instrumental simplicity and full automation. Furthermore, in the most commonly used modes, namely reversed phase C_{18} CapLC and capillary zone electrophoresis (CZE), the separation mechanisms are considered complementary, because they are dependent on the hydropho-

[☆] Selected paper from the 22nd International Symposium on Electro- and Liquid Phase-Separation Techniques (ITP2015) and the 8th Nordic Separation Science (NoSSS) Symposium, 30 August–3 September 2015, Helsinki, Finland.

* Corresponding author.

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

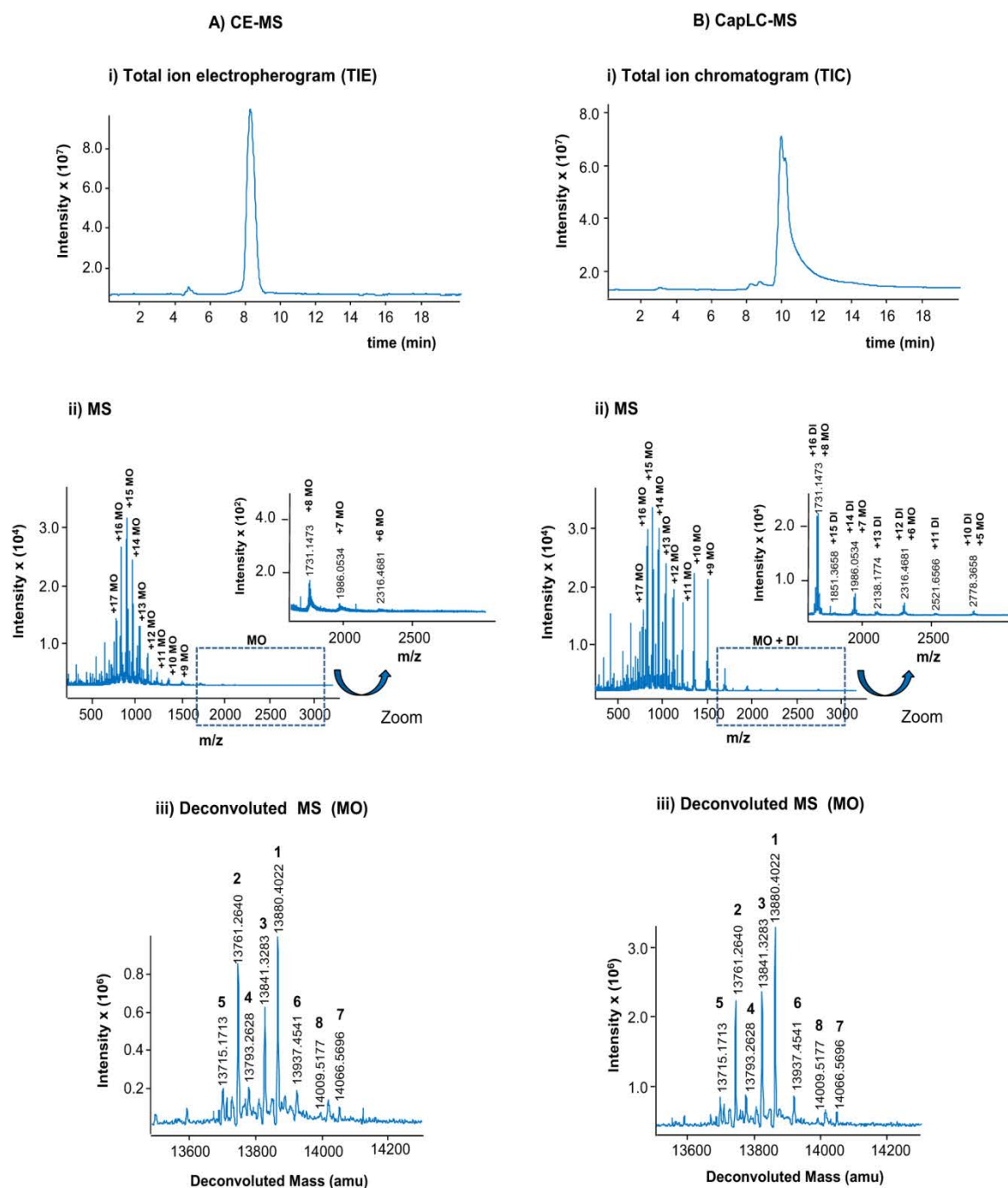


Fig. 1. (A) CE-MS and (B) CapLC-MS for a TTR standard at a concentration of 1000 $\mu\text{g mL}^{-1}$ (MO = monomer, DI = Dimer) ((i) total ion electropherogram/total ion chromatogram, (ii) TTR mass spectrum and (iii) TTR deconvoluted mass spectrum) (1: TTR-Cys, 2: Free-TTR, 3: TTR- Phosphorylated or TTR-Sulfonated, 4: TTR-Dehydroxylated or TTR-Sulfonic, 5: (10) C-G, 6: TTR-CysGly, 7: TTR-Glutathione, 8: TTR-CysGlu).

bicity and the charge-to-size ratios of the analytes, respectively [13–15]. In both cases, in order to achieve the best performance, a sample pretreatment method is necessary to purify and preconcentrate the target protein from the rest of serum components before the separation. The classical isolation procedures of serum proteins based on affinity chromatography, size exclusion chromatography or ion exchange chromatography are labor-intensive and time-consuming [16–18]. As an alternative, some authors have described conventional immunoprecipitation (IP) methods in solution, which are very simple and selective, allowing protein preparations of a high degree of purity [19–21]. In recent years, the emergence and

development of magnetic particles (MBs) for analytical purposes have led to an evolution of these methods. Nowadays, IP with MBs is becoming widely applied to avoid antibody elution, improve protein recovery, reproducibility and minimize unspecific binding of other proteins [22–24].

In the present study, CE-MS and CapLC-MS were compared for the analysis of serum transthyretin (TTR). TTR is a homotetramer composed of four identical monomers (MO) (relative molecular mass (M_r) $\sim 14,000$) with different proteoforms (isoforms and PTMs) [25–27]. TTR is known to misfold and aggregate as stable insoluble fibrils due to mutations and conformational changes,

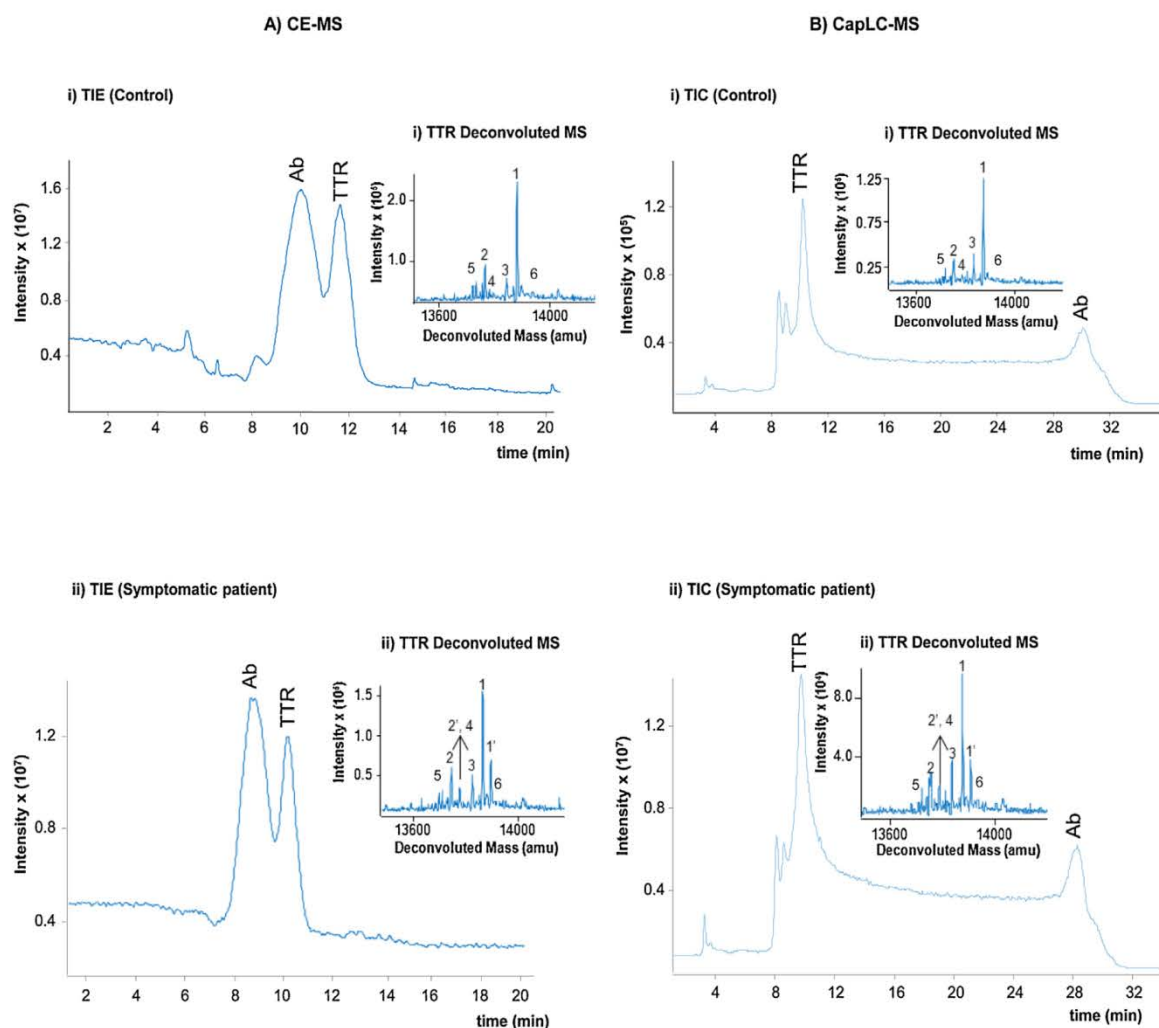


Fig. 2. (A) CE-MS and (B) CapLC-MS for a healthy control (i) and a symptomatic FAP-I patient (ii) using UAPA MBs (Ab: Anti-TTR antibody) (1: TTR-Cys, 1': TTR(Met30)-Cys, 2: Free-TTR, 2': Free-TTR(Met30), 3: TTR-Phosphorylated or TTR-Sulfonated, 4: TTR-Dehydroxylated or TTR-Sulfinic, 5: (10) C-G, 6: TTR-CysGly).

causing different neurodegenerative diseases known as amyloidosis [28,29]. Familial amyloidotic polyneuropathy type I (FAP-I), which is the most common type of hereditary amyloidosis, is associated with a TTR variant that presents a single amino acid substitution of valine for methionine at position 30 (Met 30) [30,31]. For a fair and systematic comparison, which is difficult to find in the literature, the same standard protein solutions and serum samples were analyzed using the same time-of-flight (TOF) mass spectrometer, which was successively coupled to the CE and CapLC systems. Furthermore, a novel sample pretreatment based on IP using Protein A Ultrarapid Agarose™ (UAPA) MBs is described, for the first time to our knowledge, to purify TTR from serum samples. Both techniques showed advantages and disadvantages and combined with the novel sample pretreatment permitted us to obtain simply and rapidly a rich global profile of TTR PTMs and isoforms in serum, which proved to be useful to differentiate normal and variant forms of TTR in serum samples from healthy controls and FAP-I patients.

2. Materials and methods

2.1. Chemicals and reagents

All the chemicals used in the preparation of buffers and solutions were of analytical reagent grade. Propan-2-ol ($\geq 99.9\%$), formic acid

(HFor) (99.0%), acetic acid (HAc) (glacial), hydrochloric acid (HCl) (25.0%), sodium hydrogenphosphate ($\geq 99.0\%$), sodium chloride ($\geq 99.5\%$), sodium hydroxide ($\geq 99.0\%$, pellets), potassium dihydrogenphosphate ($\geq 99.0\%$), potassium chloride (99.0%), glycine (99.7%) and TTR ($\geq 95.0\%$) were purchased from Merck (Darmstadt, Germany). Water (LC-MS grade) and acetonitrile (MeCN) (LC-MS grade) were supplied by Sigma (Madrid, Spain). Rabbit antihuman TTR polyclonal antibody was purchased from Dako (Glostrup, Denmark). Polyethylene glycol (PEG) 8000 M_r ($\sim 50\%$ in water) was supplied by Fluka (Buchs, Switzerland). Protein A Ultrarapid Agarose™ magnetic beads (UAPA MBs) were purchased from Lab on a Bead (Uppsala, Sweden). Water with a conductivity value lower than $0.05 \mu S cm^{-1}$ was obtained using a Milli-Q water purification system (Millipore, Molsheim, France).

2.2. Apparatus

pH measurements were made with a Crison 2002 potentiometer and a Crison electrode 52-03 (Crison instruments, Barcelona, Spain). Centrifugal filtration at 4 or 25 °C was carried out in a cooled Rotanta 460 centrifuge (Hettich Zentrifugen, Tuttlingen, Germany). Agitation during sample incubation was performed with a Vortex Genius 3 (Ika®, Staufen, Germany).

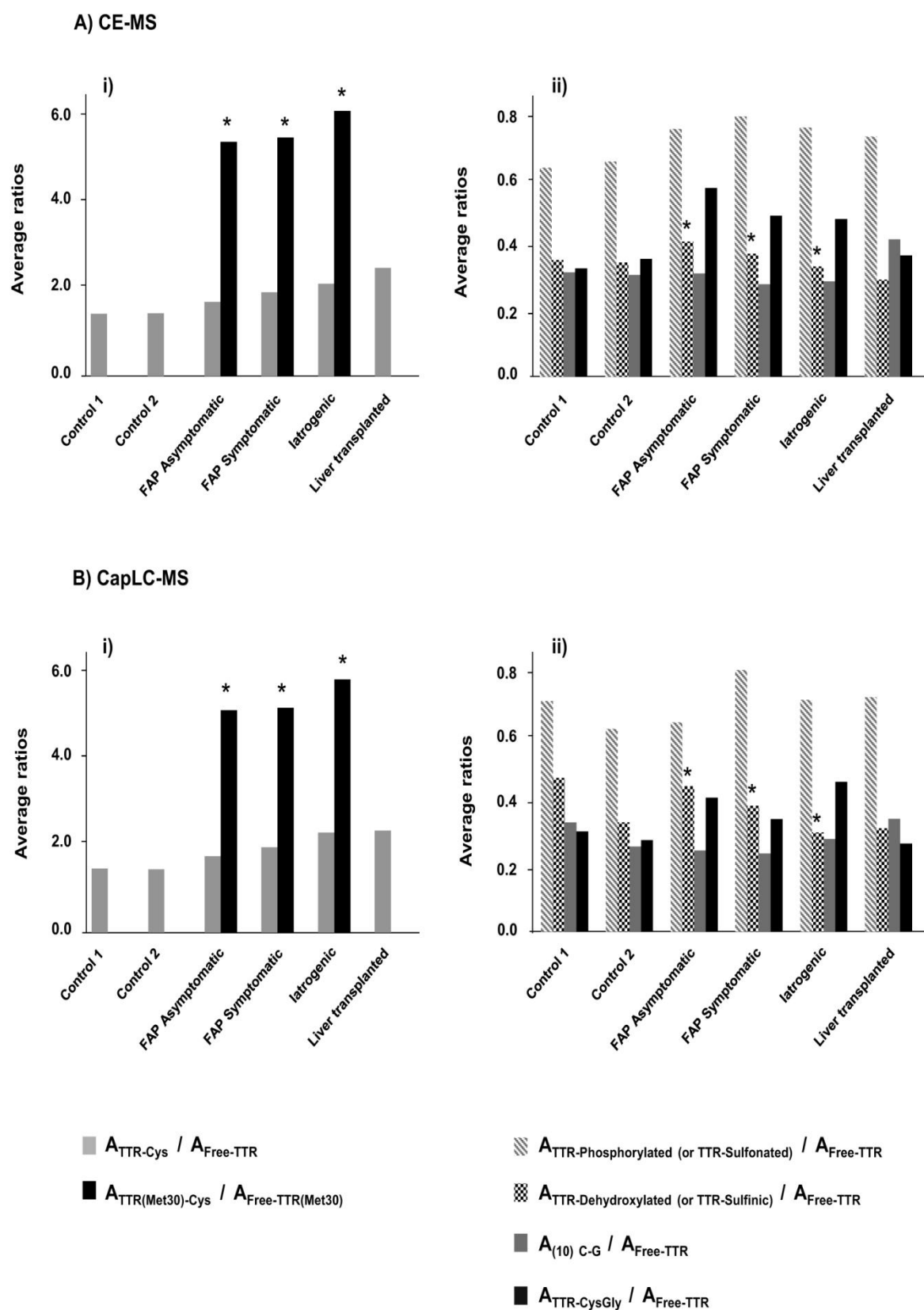


Fig. 3. Bar graphs with the average ratios ($n=3$) between the areas of the detected normal and FAP-I TTR proteoforms with regard to Free-TTR and Free-TTR(Met30) by (A) CE-MS and (B) CapLC-MS ((i) Main proteoforms and (ii) Minor proteoforms). The area was measured from the extracted ion electropherograms (EIEs) and extracted ion chromatograms (EICs) obtained from the most abundant molecular ions (+16, +15, +14, +13). *In patient samples the area of TTR-Dehydroxylated (or TTR-Sulfinic) and Free-TTR(Met30) was estimated taking into account an increase of 52.5% on the peak area of this mixture of proteoforms with regard to the controls, where the mutant proteoform was not present.

2.3. Procedures

2.3.1. Standard protein solutions

An aqueous standard solution of TTR ($1000 \mu\text{g mL}^{-1}$) was prepared as described elsewhere [20]. Excipients of low M_r were removed from the TTR standard solution by passage through $10,000 M_r$ cut-off cellulose acetate filters (Amicon Ultra-0.5, Millipore). The standard solution was centrifuged at 25°C for 10 min at $11,000g$, and the residue was washed three times for 10 min in the same way, with an appropriate volume of HAc $1.0 M$. The final residue was recovered by inverting the upper reservoir in a vial and spinning once more at a reduced centrifugal force (2 min at $300g$). Sufficient HAc $1.0 M$ was added to adjust the concentration to $1000 \mu\text{g mL}^{-1}$. The $10,000 M_r$ filters were passivated before filtration with PEG solution to ensure optimum protein recoveries [32].

2.3.2. Purification of TTR with Protein A Ultrarapid Agarose magnetic beads

Human blood samples from two healthy controls and four FAP-I patients were kindly supplied by the Hospital Universitari de Bellvitge (HUB, Hospitalet de Llobregat, Spain). The four FAP-I samples corresponded to an asymptomatic patient, a symptomatic patient (symptom onset was 4 years before blood collection), a liver transplanted patient with the specific mutation (transplanted in 2013), and a iatrogenic patient originally without the mutation who was transplanted a liver from an FAP-I patient (transplanted 14 years before blood collection and symptom onset was 6 years before collection). The assay was approved by the Ethics Committee of the HUB and written informed consent was obtained from all participants in the study. Serum was prepared as described in our previous work [20]. Serum aliquots were stored in a freezer at -20°C when not in use.

The procedure for purification of TTR from serum samples using UAPA MBs is described in detail below. First, the solvent of the commercial Anti-TTR antibody was changed to phosphate buffered saline (PBS, $0.01 M$ sodium hydrogenphosphate, $0.0015 M$ potassium dihydrogenphosphate, $0.14 M$ sodium chloride, $0.0027 M$ potassium chloride, pH 7.40) by filtration using $10,000 M_r$ cut-off cellulose acetate filters (as explained in Section 2.3.1 for TTR standard). After that, $50 \mu\text{L}$ of UAPA MBs were vortexed and the supernatant (storage solution) was removed after using a magnet to sediment the particles (magnetic separation). The MBs were washed with $100 \mu\text{L}$ of PBS twice and resuspended in $50 \mu\text{L}$ of PBS. Fifty μL of Anti-TTR antibody were then added to the MBs suspension. The mixture was moderately shaken for 40 min at room temperature. The supernatant was removed by magnetic separation and the MBs were subsequently washed three times with $100 \mu\text{L}$ of PBS. In order to immunoprecipitate TTR, $75 \mu\text{L}$ of serum sample were added to the MBs and the mixture was incubated for 20 min at room temperature with gentle orbital shaking. Again, the supernatant was removed and the MBs were then washed three times with $200 \mu\text{L}$ of PBS. Finally, in order to elute the TTR and the Anti-TTR antibody, $50 \mu\text{L}$ of $50 mM$ glycine (adjusted to pH 2.80 with HCl) were added and the mixture was then incubated for 5 min at room temperature with gentle orbital shaking. The supernatant containing the eluted Anti-TTR antibody and the TTR was collected and transferred into a clean eppendorf. Excipients of low M_r were removed from the sample by passage through $10,000 M_r$ cut-off cellulose acetate filters (see Section 2.3.1). The protein samples were kept in the freezer and thawed immediately before the analysis to prevent precipitation and spray instability during the CE-MS and CapLC-MS runs.

2.3.3. CE-MS

All CE-MS experiments were performed in an HP^{3D} CE system coupled with an orthogonal G1603A sheath-flow interface to a 6220 oa-TOF LC/MS spectrometer (Agilent Technologies, Waldbronn, Germany). A sheath liquid solution consisting of a mixture of 60:40 v/v propan-2-ol: water with a 0.05% v/v of HFor was delivered at a flow rate of $3.3 \mu\text{L min}^{-1}$ by a KD Scientific 100 series infusion pump (Holliston, MA, USA). ChemStation C.01.06 software (Agilent Technologies) was used for CE control and separation data acquisition (e.g. voltage, temperature and current), and was run in combination with MassHunter B.04.00 workstation software (Agilent Technologies) for control, data acquisition, and analysis of the TOF mass spectrometer. m/z mass spectra were deconvoluted to obtain the M_r values of the TTR proteoforms. The deconvolution algorithm of maximum entropy computed the simplest zero charge mass spectra that could account for the observed m/z .

A 72 cm long (L_T) \times $75 \mu\text{m}$ id \times $360 \mu\text{m}$ od bare fused-silica capillary supplied by Polymicro Technologies (Phoenix, AZ, USA) was used for the separations. All capillary rinses were performed at high pressure (930 mbar). New capillaries were flushed with $1.0 M$ NaOH (20 min), water (15 min) and background electrolyte (BGE) (30 min, HAc $1.0 M$). The system was finally equilibrated by applying the separation voltage (25 kV) for 15 min. Between workdays, the capillary was conditioned by rising successively with $0.1 M$ NaOH (5 min), water (10 min) and BGE (15 min). Both activation and conditioning procedures were performed off-line to avoid the unnecessary entrance of NaOH into the MS system. Between runs, the capillary was rinsed for 2 min with HAc $1.0 M$, 1 min with water and 1 min with BGE. Samples were hydrodynamically injected at 50 mbar for 10 s. A separation voltage of 25 kV (normal polarity, anode at the injection capillary end) was applied for the electrophoretic separations at 25°C .

The TOF mass spectrometer was operated under the optimized conditions in positive mode using the following parameters: capillary voltage 4000 V , drying gas temperature 300°C , drying gas flow rate 4 L min^{-1} , nebulizer gas 7 psig, fragmentor voltage 325 V , skimmer voltage 80 V , OCT 1 RF Vpp voltage 300 V . Data were collected in profile at 1 spectrum/s between 100 and $3200 m/z$, with the mass range set to high resolution mode (4 GHz). The reference mass correction for internal mass recalibration was disabled for all the experiments for optimum sensitivity. A standard tune and an external mass calibration were performed daily at the beginning of the day following the manufacturer instructions using the typical LC-MS sprayer and ESI-L tuning mix (Agilent Technologies).

2.3.4. CapLC-MS

All CapLC-MS experiments were performed in a 1200 series capillary liquid chromatography system coupled to a 6220 oa-TOF LC/MS mass spectrometer with an orthogonal G1385-44300 interface (Agilent Technologies). LC and MS control, separation, data acquisition and processing were performed using MassHunter B.04.00 workstation software (Agilent Technologies). The TOF mass spectrometer measurement parameters were as described for CE-MS, except for the nebulizer gas (N_2), which in this case was 15 psig under the optimized conditions.

For separation, a Zorbax 300SB-C₁₈ column ($150 \text{ mm } L_T \times 0.3 \text{ mm id}$, $3.5 \mu\text{m}$ particle diameter, 300 \AA pore diameter, Agilent Technologies) was used. Experiments were performed at room temperature with gradient elution at a flow rate of $4 \mu\text{L min}^{-1}$. Eluting solvents were (A) water and (B) MeCN (both with 0.1% (v/v) HFor). Solvents were degassed for 10 min by sonication before use. The optimum elution gradient was: solvent B from 5% to 40% (v/v) in 5 min and then from 40% to 55% (v/v) in 20 min, followed by cleaning and re-equilibration steps from 55% to 100% (v/v) (5 min), 100% (v/v) (10 min), from 100% to 5% (v/v) (5 min) and finally 5% (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 11,000 g for 4 min. Sample injection was performed with an autosampler refrigerated at 4 °C and the injection volume was 0.15 μL .

2.3.5. Quality parameters in CE-MS and CapLC-MS

All quality parameters in CE-MS and CapLC-MS were calculated from data obtained by measuring the peak area and migration or retention time ($t_{\text{m}}-t_{\text{r}}$) from the extracted ion electropherograms (EIEs) and extracted ion chromatograms (EICs) of TTR proteoforms with the different modifications (considering the m/z of the most abundant molecular ions, i.e. ions with charge +16, +15, +14, +13). The deconvoluted average M_{r} and abundance of the detected proteoforms for a standard of normal TTR are shown in Tables 1 and 2 for CE-MS and CapLC-MS, respectively (TTR standard 1000 $\mu\text{g mL}^{-1}$). Repeatability studies were performed by analyzing 10 times in one day a standard solution of TTR at a concentration of 250 $\mu\text{g mL}^{-1}$ ($n = 10/1$ day). This parameter was calculated as a percentage of the relative standard deviation (RSD%) of peak areas and $t_{\text{m}}-t_{\text{r}}$. An estimation of the limits of detection (LODs) was obtained by analyzing low concentration solutions of a standard of TTR until no peaks were detected (1 and 10 $\mu\text{g mL}^{-1}$ for CE-MS and CapLC-MS, respectively). The S/N values were calculated referred to peak areas using the software provided by the manufacturer of the mass spectrometer. The calibration range was established by injecting standard solutions of TTR at concentrations between 10 and 1000 $\mu\text{g mL}^{-1}$.

3. Results and discussion

3.1. Analysis of TTR standard by CE-MS and CapLC-MS

CE-MS was evaluated by analyzing standard solutions of a commercially available normal TTR in positive ESI mode with a BGE of HAc 1.0M (pH 2.30). In a previous study [20], we demonstrated that this acidic BGE disrupted TTR tetrameric structure generating only the monomer (MO), but that provided better sensitivity than ammonium acetate neutral BGEs. CapLC-MS conditions in ESI positive mode were optimized in this work by analyzing the TTR standard solutions using MeCN:water (with 0.1% (v/v) HFor) mobile phases. In preliminary experiments, two columns (Zorbax 300SB-C₁₈ and Zorbax 300SB-C₈) and different gradients were tested. The best peak shape, efficiency and separation of TTR proteoforms, as well as the shortest retention times, were obtained using the C₁₈ column with the optimized gradient described in the experimental Section 2.3.4. In order to improve detection sensitivity and peak shape, different percentages of HFor in the mobile phases (between 0.1%–0.5% v/v) were tested. As the highest S/N values were obtained using 0.1% (v/v) HFor and the peak shape was similar in all the cases, this percentage was chosen for further analysis.

Fig. 1A and B show the total ion electropherogram (TIE) and the total ion chromatogram (TIC) (i), the mass spectra (ii) and the deconvoluted mass spectra (iii) of a 1000 $\mu\text{g mL}^{-1}$ standard solution of normal TTR by CE-MS and CapLC-MS, respectively. As it is shown, the sensitivity achieved with both techniques was quite similar (compare the y-axis of TIE/TIC and mass spectra), but peak shape was better and total analysis times slightly lower by CE-MS. A conscientious analysis of the cluster of multiply charged ions obtained in the m/z spectra of both peaks showed the presence of protein MO with CE-MS (the most abundant molecular ion had +15 charge), and the presence of MO and a very small amount of dimer (DI) using CapLC-MS. As can be observed in Fig. 1B (ii) (see the zoomed area), ions with nominal m/z values of 1731, 1986, 2316 and 2778 would be the summed contribution of the MO and DI, while 1851, 2138 and 2522 ions would be only detected for the DI (+15, +13 and +11 ions, respectively). In our previous study with TTR

Table 1
Theoretical and deconvoluted average M_{r} , abundance, repeatability, calibration range and LOD for the detected forms of a TTR standard solution by CE-MS.

N	Monomer Transthyretin Forms	% Abundance ^a	Deconvoluted Average M_{r}	E_{r}^b (ppm)	Peak area		Migration times (t_{m})	Calibration range ^c : 50–1000 $\mu\text{g mL}^{-1}$, $r^2 > 0.998$	LOD ^d ($\mu\text{g mL}^{-1}$)
					Theoretical	Experimental			
1	TTR-Cys	100	13880.4022	36					
2	Free-TTR	91	13761.2640	38					
3	TTR-Phosphorylated	67	13841.2439	29					
	TTR-Sulfonated	66	13841.3283	23					
4	TTR-Dehydroxylated or TTR-Sulfinic	27	13793.2628	17					
5	(10) C-G	24	13715.1713	34					
6	TTR-CysGly	11	13937.4541	3					
7	TTR-Glutathione	10	14066.5696	28					
8	TTR-CysGlu	9	14009.5177	11					
					Repeatability 250 $\mu\text{g mL}^{-1}$ ($n = 10$), RSD%	Average ($n = 10$)	Repeatability 250 $\mu\text{g mL}^{-1}$ ($n = 10$), RSD%		
					0.2	7.9	0.5	$A = 1 \times 10^4 \text{C}-2 \times 10^4$	10 (S/N=457)
					1.7	8.0	0.2	$A = 1 \times 10^4 \text{C}-2 \times 10^4$	10 (S/N=430)
					0.2	8.1	0.3	$A = 7 \times 10^3 \text{C}-2 \times 10^4$	10 (S/N=301)
					1.0	8.1	0.2	$A = 7 \times 10^3 \text{C}-2 \times 10^4$	10 (S/N=286)
					2.5	8.1	0.3	$A = 3 \times 10^3 \text{C}-1 \times 10^4$	25 (S/N=264)
					0.6	7.9	0.6	$A = 2 \times 10^3 \text{C}-4 \times 10^4$	25 (S/N=181)
					4.7	7.9	0.9	$A = 2 \times 10^3 \text{C}-8 \times 10^4$	100 (S/N=780)
					1.5	8.0	0.1	$A = 1 \times 10^3 \text{C}-2 \times 10^4$	100 (S/N=758)
					7.1	8.0	0.1	$A = 9 \times 10^2 \text{C}-2 \times 10^4$	100 (S/N=447)

^a The % Abundance was calculated normalizing to the area value of the most abundant form.

^b Relative error (E_{r}) was calculated in ppm as: $|M_{\text{r,exp}} - M_{\text{r,theo}}|/M_{\text{r,theo}} \times 10^6$ (exp = experimental and theo = theoretical).

^c Calibration range for TTR-CysGly, TTR-Glutathione and TTR-CysGlu ranged between 100–1,000 $\mu\text{g mL}^{-1}$.

^d Concentration was referred to the total amount of protein in the standard solution. The S/N for a certain form at this concentration was calculated from the EIE using the software provided by the manufacturer of the mass spectrometer.

Table 2
Theoretical and deconvoluted average M_r , abundance, repeatability, calibration range and LOD for the detected forms of a TTR standard solution by CapLC–MS.

N	Monomer	Transthyretin Forms	% Abundance ^a	Deconvoluted Average M _r		Peak area	Retention times (t _r)	Calibration range ^c : 100-1000 µg mL ⁻¹ , r ² > 0.998	LOD ^d (µg mL ⁻¹)		
				Theoretical	Experimental	E _r ^b (ppm)	Repeatability 250 µg mL ⁻¹ (n = 10), RSD%	Average (n = 10)	Repeatability 250 µg mL ⁻¹ (n = 10), RSD%		
1	TTR-Cys		100	13880.4022	13880.7500	3	1.8	10.3	0.1	A = 2 × 10 ⁴ C + 9 × 10 ⁴	25 (S/N = 46)
2	Free-TTR		88	13761.2640	13761.4500	1	2.0	10.1	0.1	A = 1 × 10 ⁴ C + 6 × 10 ⁴	25 (S/N = 31)
3	TTR-Phosphorylated		67	13841.2439	13841.5700	2	0.3	10.3	0.2	A = 1 × 10 ⁴ C + 4 × 10 ⁴	25 (S/N = 14)
	TTR-Sulfonated		66	13841.3283	13841.5700	2	1.7	10.3	0.2	A = 1 × 10 ⁴ C + 4 × 10 ⁴	25 (S/N = 11)
4	TTR-Dehydroxylated orTTR-Sulfinic		25	13793.2628	13793.9300	5	2.2	10.2	0.3	A = 5 × 10 ³ C + 2 × 10 ⁴	50 (S/N = 18)
5	(10) C-G		21	13715.1713	13715.3300	1	2.6	10.0	0.1	A = 4 × 10 ³ C + 2 × 10 ⁴	50 (S/N = 13)
6	TTR-CysGly		12	13937.4541	13937.8900	3	2.2	10.0	0.1	A = 4 × 10 ³ C + 3 × 10 ⁴	250 (S/N = 220)
7	TTR-Glutathione		10	14066.5696	14066.6700	1	0.6	10.2	0.2	A = 2 × 10 ³ C + 1 × 10 ⁴	250 (S/N = 184)
8	TTR-CysGlu		7	14009.5177	14009.3500	1	2.9	10.2	0.3	A 1 × 10 ³ C + 8 × 10 ⁴	250 (S/N = 118)

^a The % Abundance was calculated normalizing to the area value of the most abundant form.^b Relative error (E_r) was calculated in ppm as: $|M_{r,\text{exp}} - M_{r,\text{theo}}|/M_{r,\text{theo}} \times 10^6$ (exp = experimental and theo = theoretical).^c Calibration range for TTR-CysGly, TTR-Glutathione and TTR-CysGlu ranged between 250–1,000 $\mu\text{g mL}^{-1}$.^d Concentration was referred to the total amount of protein in the standard solution. The S/N for a certain form at this concentration was calculated from the EIC using the software provided by the manufacturer of the mass spectrometer.

using CE-MS, we already observed that under these analytical conditions sensitivity was excellent but the DI or the tetramer were not detected [20]. In that case, we proposed that the oligomeric structures were disrupted due to the acidic BGE, the acidic sheath liquid or the vacuum-pressure inside the spectrometer (which could not be tuned by the operator). Now, the presence of a very small amount of DI in CapLC–MS suggested that the optimized separation conditions preserved to some extent the TTR oligomeric structure. This is probably due to the higher acidity of the BGE in CE-MS compared to the mobile phase in CapLC–MS ($\approx 6\%$ vs 0.1% v/v of HAc and HFor, respectively).

Figs. 1A and B(iii) show the deconvoluted mass spectra for the MO using CE-MS and CapLC–MS, respectively. As can be observed in these figures, the same proteoforms were detected with both techniques. In addition, as it is shown in Tables 1 and 2, the relative abundances for all the monomeric TTR forms, which were calculated taking into account the peak areas of the EIEs and EICs, were also similar. Tables 1 and 2 show the experimental M_r ($M_{r,\text{exp}}$), the theoretical M_r ($M_{r,\text{theo}}$) and the mass error (E_r) values for the proteoforms. The detected TTR PTMs and isoforms corresponded (in order of abundance) to the TTR showing a mixed disulfide with the amino acid cysteine at position 10 of the sequence (TTR-Cys, $M_{r,\text{theo}} = 13880.4022$), the free TTR (Free-TTR, $M_{r,\text{theo}} = 13761.2640$), the phosphorylated or sulfonated TTR (TTR-Phosphorylated, $M_{r,\text{theo}} = 13841.2439$ and TTR-Sulfonated, $M_{r,\text{theo}} = 13841.3283$, respectively), the dehydroxylated or the conjugated cysteine sulfinic acid TTR (TTR-Dehydroxylated or TTR-Sulfinic, $M_{r,\text{theo}} = 13793.2628$), the isoform caused by the single amino acid substitution of cysteine for glycine at position 10 ((10) C-G, $M_{r,\text{theo}} = 13715.1713$), and the PTMs resulting when the cysteine residue on position 10 makes a mixed disulfide with the peptide cysteinyl-glycine (TTR-CysGly, $M_{r,\text{theo}} = 13937.4541$), the peptide glutathione (TTR-Glutathione, $M_{r,\text{theo}} = 14066.5696$) and the peptide cysteinyl-glutamic acid (TTR-CysGlu, $M_{r,\text{theo}} = 14009.5177$). As can be observed in Tables 1 and 2, the accuracy of the M_r determination with this oa-TOF mass spectrometer did not allow differentiation between TTR-Phosphorylated and TTR-Sulfonated PTMs, which were probably not separated by CE-MS or CapLC–MS. Similarly, we could not differentiate between TTR-Dehydroxylated or TTR-Sulfinic, which were isobars. It is worth mentioning that the E_r values were significantly lower for all the detected forms using CapLC–MS (with values ranging between 1 and 5 ppm, see Tables 1 and 2). As all the mass spectrometer parameters were the same in both cases and mass accuracy depended on the external mass calibration performed at the beginning of the day, the accuracy differences meant that calibration in CE-MS was less stable, probably due to the specificities of the ionization and the interface. We will show below that this fact did not affect repeatability of peak areas in CE-MS.

With regard to the quality parameters, Tables 1 and 2 summarize the RSD% obtained for the repeatability of the peak areas and migration or retention times (t_m or t_r , $n = 10$), LODs and calibration ranges for TTR proteoforms using CE-MS and CapLC–MS, respectively. The repeatability for peak areas was good with both techniques, ranging between 0.2% and 7.1% for CE-MS, and between 0.3% and 2.9% for CapLC–MS. Similarly, the repeatability for t_m or t_r was also acceptable and ranged between 0.1%–0.9% and 0.1%–0.3% for CE-MS and CapLC–MS, respectively. The average values of t_m or t_r show that there was only a slight separation between the TTR forms, because all of them presented similar charge-to-size ratios and hydrophobicities, which hindered their separation using capillary zone electrophoresis (CZE) and reversed phase C_{18} CapLC, respectively. Linear correlation between migration and retention times of the different proteoforms was very low ($R^2 = 0.36$), confirming the orthogonality of the separation mechanisms. With regard to the peak efficiency, the number of theoretical plates

($N = 16 \cdot (t/w)^2$, w = peak width) was approximately 4-fold higher in CE-MS (i.e. 616 vs 158 for the most abundant proteoform, TTR-Cys). Area versus concentration regression lines were evaluated in the range between 10 and 1000 $\mu\text{g mL}^{-1}$. Calibration ranges for all the TTR forms (except for TTR-CysGly, TTR-Glutathione and TTR-CysGlu, which were the proteoforms with the lowest abundances, see Tables 1 and 2) were observed between 50 and 1000 $\mu\text{g mL}^{-1}$ using CE-MS and between 100 and 1000 $\mu\text{g mL}^{-1}$ using CapLC-MS. In the case of TTR-CysGly, TTR-Glutathione and TTR-CysGlu, calibration ranges were observed between 100 and 1000 $\mu\text{g mL}^{-1}$ and between 250 and 1000 $\mu\text{g mL}^{-1}$ with CE-MS and CapLC-MS, respectively. Finally, as it was expected, the LODs were found to decrease according to the abundance of the different proteoforms using both techniques (see Tables 1 and 2). It is worth mentioning that the LODs achieved by CE-MS were slightly lower than by CapLC-MS (2–2.5-fold lower).

3.2. Purification of TTR from serum samples using Protein A Ultrarapid Agarose magnetic beads

In a previous work, we proposed a conventional IP method in solution to purify TTR from serum samples before CE-MS [20]. As the antibody altered the electrophoretic current and stability, making the method not reproducible, it was necessary to separate the TTR from the antibody using 100,000 M_r cut-off centrifugal filters. In order to avoid this filtration step, which could be detrimental for protein recovery because monomeric TTR is recovered from the filtrate, we propose now IP with UAPA MBs. In UAPA MBs the antibody orientation is optimum for the interaction with TTR because the antibody is coupled to the MBs through the fragment crystallizable (Fc) region. Furthermore, the agarose support, due to its hydrophilic nature, enables minimal unspecific binding of other proteins. However, under the optimized conditions, as there is not a covalent bond between the antibody and the MBs, both the antibody and TTR are eluted from the MBs before CE-MS or CapLC-MS analysis.

The IP procedure with UAPA MBs was applied to serum samples from two healthy controls and four patients suffering from FAP-I (one asymptomatic, one symptomatic, a liver transplanted patient with the specific mutation, and a iatrogenic patient originally without the mutation who received a liver transplant from an FAP-I patient). The eluted antibody and TTR mixture was passed through 10,000 M_r cut-off cellulose acetate filters (Amicon-Ultra 0.5, Millipore) in order to exchange the elution buffer, which contained glycine that is not recommended for ESI-MS, to HAc 1.0M before the analysis by CE-MS and CapLC-MS. Now, in contrast to our previous work [20], it was not necessary to separate the TTR from the antibody using 100,000 M_r centrifugal filters, probably because the antibody concentration in solution was lower than before. It is widely accepted that the recovery of the retentate in M_r cut-off centrifugal filters is higher than the recovery in the filtrate. The recovery of TTR in the retentate using 10,000 M_r filters was $\geq 95\%$, while the recovery of TTR in the filtrate using 100,000 M_r filters was 76% (in both cases the filters were passivated with PEG in order to ensure optimum recoveries [20,32]). We estimated that with the novel IP method based on MBs, TTR recovery was approximately twice higher than by the conventional IP method in solution.

3.3. Analysis of TTR in serum samples by CE-MS and CapLC-MS

The immunoprecipitated serum samples were subsequently analyzed in triplicate by CE-MS and CapLC-MS. Fig. 2A and B show the TIE and TIC and the TTR deconvoluted mass spectra obtained for a serum sample from a healthy control (i) and an FAP-I patient (symptomatic) (ii) using CE-MS and CapLC-MS, respectively. As it is shown, the main peaks corresponded to the antibody and the

monomeric TTR (in this case no DI was observed by CapLC-MS probably due to the IP conditions, data not shown). As can be observed, while the antibody migrated before TTR and separation was not complete by CE-MS, by CapLC-MS the antibody was highly retained and eluted baseline separated from TTR at a mobile phase composition of 100% (v/v) MeCN. The TTR deconvoluted mass spectra for the healthy control (Fig. 2A and B (i)) show that the main forms detected using both techniques corresponded (in order of abundance) to TTR-Cys (1), Free-TTR (2), TTR-Phosphorylated or TTR-Sulfonated (3), TTR-Dehydroxylated or TTR-Sulfinic (4), (10) C-G (5) and TTR-CysGly (6). Therefore, the IP with MBs allowed detecting TTR forms found at lower concentrations that were not detected with conventional IP in solution in our previous work (only TTR-Cys and Free-TTR were detected [20]). In the case of the FAP-I patient (symptomatic) (Fig. 2A and B (ii)), the main proteoforms corresponded to the normal forms detected for a healthy control, and the TTR(Met30)-Cys (1') and the Free-TTR(Met30) (2') for the abnormal (or mutant) TTR, with a M_r 32.07 units higher than the normal forms of TTR ($\Delta M_r = +32.07$). Unfortunately, the accuracy of the M_r determination did not allow differentiation between the normal TTR-Dehydroxylated (or TTR-Sulfinic) and the mutant Free-TTR(Met30) ($M_{r,theo}$ are very similar: 13793.2628 and 13793.3301, respectively). The detection of this mutant form was indirectly assessed taking into account that the peak area for this mixture of proteoforms increased 52.5% for the patients with regard to the controls, where Free-TTR(Met30) was not present.

Repeatability in the TTR analysis after the serum sample pre-treatment was good for peak areas and t_m or t_r using both techniques. As an example, for the symptomatic FAP-I patient, repeatability values ($n=3$, three independent samples) for peak areas and t_m or t_r ranged between 3.6%–10.4% and 0.1%–1.0% using CE-MS, and between 2.7%–11.4% and 0.2%–0.7% using CapLC-MS, respectively.

Under the optimized conditions, the total concentration of TTR in our samples was estimated to range between 47 and 160 $\mu\text{g mL}^{-1}$ using CE-MS, and between 40 and 130 $\mu\text{g mL}^{-1}$ using CapLC-MS. It is well known that the total concentration of TTR in serum and cerebrospinal fluid is variable and depends on the individual [20,31]. These values were similar to those reported in our previous work, where a conventional IP method in solution was used prior CE-MS (50 and 130 $\mu\text{g mL}^{-1}$) [20].

Some authors have pointed out the importance of TTR oxidized forms (e.g. TTR-Cys and TTR-CysGly), as biomarkers of the oxidative stress involved in amyloid deposit formation in FAP-I and other amyloid diseases [20,33,34]. In our previous study [20], the level of TTR(Met30)-Cys in serum samples from FAP-I symptomatic patients was found especially elevated when compared to the mutant Free-TTR(Met30) proteoform. Fig. 3A and B(i) show the average TTR-Cys/Free-TTR (grey bars) and TTR(Met30)-Cys/Free-TTR(Met30) (black bars) ratios for the healthy controls and the FAP-I patients using CE-MS and CapLC-MS, respectively. As can be observed, both techniques were providing similar information and the mutant proteoforms were only found in the asymptomatic, symptomatic and iatrogenic FAP-I patients (black bars), whereas the liver-transplanted patient behaved as the controls because the treatment was effective. In the untreated patients the abundance of TTR(Met30)-Cys was especially higher than the abundance of Free-TTR(Met30), as in our previous work with CE-MS [20]. With regard to the oxidized normal proteoforms (i.e. TTR-Cys, TTR-CysGly, TTR-Sulfonated or TTR-Sulfinic), only the abundance of TTR-CysGly was slightly increased compared to the Free-TTR proteoform in the untreated patients (see the black bars of Fig. 3A and B (ii)). As it is shown, the general profile obtained for the different average proteoform ratios was quite similar using both techniques, hence

suggesting that CE-MS and CapLC-MS are both useful to reliably gain insight into FAP-I.

4. Concluding remarks

In this work, CE-MS and CapLC-MS were compared for the analysis of serum TTR. A novel sample pretreatment based on IP using UAPA MBs was applied to purify TTR from serum samples. This novel IP procedure enhanced our previously method based on conventional IP in solution and allowed the detection of monomeric TTR proteoforms found at lower concentrations. CE-MS and CapLC-MS were applied to analyze TTR standard solutions and serum samples from healthy controls and FAP-I patients. The optimized methods with both techniques showed a similar performance for the analysis of TTR. However, despite the lower separation resolution between the antibody and TTR, we propose the CE-MS method as the best alternative, mainly because of the slightly lower LODs and total analysis times, the use of an organic solvent-free BGE, the smaller sample volumes injected and the lower cost of the separation columns.

Acknowledgements

Laura Pont acknowledges the Spanish Ministry of Economy and Competitiveness for a FPI fellowship. This study was supported by a grant from the Spanish Ministry of Education and Science (CTQ2011-27130/CTQ2014-56777-R). Kader Poturcu acknowledges Prof. Dr. Ismail Ozmen and Prof. Dr. A. Guleren Alsancak for their unique scientific support in the bilateral agreement for Staff Mobility and Student Exchange/Socrates/Erasmus Programme. We also thank Dr. C. Casasnovas and Dr. M. A. Alberti (Hospital Universitari de Bellvitge, HUB, Hospitalet de Llobregat, Spain) for providing the blood samples. Protein A Ultrarapid Agarose™ magnetic beads were kindly supplied by Lab on a Bead (Uppsala, Sweden).

References

- [1] N.L. Anderson, The clinical plasma proteome: a survey of clinical assays for proteins in plasma and serum, *Clin. Chem.* 56 (2010) 177–185.
- [2] E.S. Boja, T.E. Fehniger, M.S. Baker, G. Marko-Varga, H. Rodriguez, Analytical validation considerations of multiplex mass-spectrometry-based proteomic platforms for measuring protein biomarkers, *J. Proteome Res.* 13 (2014) 5325–5332.
- [3] A. Shimizu, T. Nakanishi, M. Kishikawa, A. Miyazaki, Detection and identification of protein variants and adducts in blood and tissues: an application of soft ionization mass spectrometry to clinical diagnosis, *J. Chromatogr. B* 776 (2002) 15–30.
- [4] A. Shimizu, T. Nakanishi, A. Miyazaki, Detection and characterization of variant and modified structures of proteins in blood and tissues by mass spectrometry, *Mass Spectrom. Rev.* 25 (2006) 686–712.
- [5] D. Guan, Z. Chen, Challenges and recent advances in affinity purification of tag-free proteins, *Biotechnol. Lett.* 36 (2014) 1391–1406.
- [6] N.E. Labrou, Protein purification: an overview, *Methods Mol. Biol.* 1129 (2014) 3–10.
- [7] H. Johnson, C.E. Eyers, Analysis of post-translational modifications by LC-MS/MS, *Methods Mol. Biol.* 658 (2010) 93–108.
- [8] I.A. Popov, N.L. Starodubtseva, M.I. Indeikina, Y.I. Kostyukovich, A.S. Kononikhin, M.I. Nikolaeva, Mass spectrometric identification of posttranslational modifications in transthyretin from human blood, *Mol. Biol.* 47 (2013) 885–893.
- [9] J. Hernández-Borges, C. Neusüss, A. Cifuentes, M. Pelzing, On-line capillary electrophoresis-mass spectrometry for the analysis of biomolecules, *Electrophoresis* 25 (2004) 2257–2281.
- [10] R. Haselberg, G.J. de Jong, G.W. Somsen, CE-MS for the analysis of intact proteins 2010–2012, *Electrophoresis* 34 (2013) 99–112.
- [11] H. Zhou, Z. Ning, A.E. Starr, M. Abu-Farha, D. Figeys, Advancements in top-down proteomics, *Anal. Chem.* 84 (2012) 720–734.
- [12] I.C. Forstenlehner, J. Holzmann, H. Toll, C.G. Huber, Site-specific characterization and absolute quantification of pegfilgrastim oxidation by top-down high-performance liquid chromatography-mass spectrometry, *Anal. Chem.* 87 (2015) 9336–9343.
- [13] 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.
- [14] K. Ahrer, A. Jungbauer, Chromatographic and electrophoretic characterization of protein variants, *J. Chromatogr. B* 841 (2006) 110–122.
- [15] Y. Li, M.M. Champion, L. Sun, P.A. Champion, R. Wojcik, N.J. Dovichi, Capillary zone electrophoresis-electrospray ionization-tandem mass spectrometry as an alternative proteomics platform to ultra-performance liquid chromatography-electrospray ionization-tandem mass spectrometry for samples of intermediate complexity, *Anal. Chem.* 84 (2012) 1617–1622.
- [16] M.V. Bimanpalli, P.S. Ghaswala, Isolation, purification and partial characterisation of prealbumin from cerebrospinal fluid, *J. Biosci.* 13 (1988) 159–169.
- [17] P. Raghu, P. Ravinder, B. Sivakumar, A new method for purification of human plasma retinol-binding protein and transthyretin, *Biotechnol. Appl. Biochem.* 38 (2003) 19–24.
- [18] A. Mahn, E. Lienqueo, C. Quilodrán, A. Olivera-Nappa, Purification of transthyretin as nutritional biomarker of selenium status, *J. Sep. Sci.* 35 (2012) 3184–3189.
- [19] N.H. Heegaard, M.Z. Hansen, J.W. Sen, M. Christiansen, P. Westermark, Immunoaffinity chromatographic and immunoprecipitation methods combined with mass spectrometry for characterization of circulating transthyretin, *J. Sep. Sci.* 29 (2006) 371–377.
- [20] 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.
- [21] M. Vilà-Rico, N. Colomé-Calls, L. Martín-Castel, M. Gay, S. Azorín, M. Vilaseca, Quantitative analysis of post-translational modifications in human serum transthyretin associated with familial amyloidotic polyneuropathy by targeted LC-MS and intact protein MS, *J. Proteomics* 127 (2015) 234–246.
- [22] H.X. Chen, J.M. Busnel, A.L. Gassner, G. Peltre, X.X. Zhang, H.H. Girault, Capillary electrophoresis immunoassay using magnetic beads, *Electrophoresis* 29 (2008) 3414–3421.
- [23] H.C. Tekin, M.A. Gijs, Ultrasensitive protein detection: a case for microfluidic magnetic bead-based assays, *Lab Chip* 13 (2013) 4711–4739.
- [24] B.D. Plouffe, S.K. Murthy, L.H. Lewis, Fundamentals and application of magnetic particles in cell isolation and enrichment: a review, *Reports Prog. Phys. Phys. Soc.* 78 (2015) 016601.
- [25] H. Terazaki, Y. Ando, O. Suhr, P.I. Ohlsson, K. Obayashi, T. Yamashita, Post-translational modification of transthyretin in plasma, *Biochem. Biophys. Res. Commun.* 249 (1998) 26–30.
- [26] W. Zheng, Y.M. Lu, G.Y. Lu, Q. Zhao, O. Cheung, W.S. Blaner, Transthyretin, thyroxine, and retinol-binding protein in human cerebrospinal fluid: effect of lead exposure, *Toxicol. Sci.* 61 (2001) 107–114.
- [27] K. Poulsen, J.M. Bahl, J.T. Tanassi, A.H. Simonsen, N.H. Heegaard, Characterization and stability of transthyretin isoforms in cerebrospinal fluid examined by immunoprecipitation and high-resolution mass spectrometry of intact protein, *Methods* 56 (2012) 284–292.
- [28] R.H. Falk, R.L. Comenzo, M. Skinner, The systemic amyloidoses, *N. Engl. J. Med.* 337 (1997) 898–909.
- [29] S.M. Johnson, S. Connelly, C. Fearn, E.T. Powers, J.W. Kelly, The transthyretin amyloidoses: from delineating the molecular mechanism of aggregation linked to pathology to a regulatory-agency-approved drug, *J. Mol. Biol.* 421 (2012) 185–203.
- [30] Y. Ando, P. Ohlsson, O. Suhr, N. Nyhlin, T. Yamashita, G. Holmgren, A new simple and rapid screening method for variant transthyretin-related amyloidosis, *Biochem. Biophys. Res. Commun.* 228 (1996) 480–483.
- [31] Y. Ando, O. Suhr, T. Yamashita, P.I. Ohlsson, G. Holmgren, K. Obayashi, Detection of different forms of variant transthyretin (Met30) in cerebrospinal fluid, *Neurosci. Lett.* 238 (1997) 123–126.
- [32] L. Pont, F. Benavente, J. Barbosa, V. Sanz-Nebot, An update for human blood plasma pretreatment for optimized recovery of low-molecular-mass peptides prior to CE-MS and SPE-CE-MS, *J. Sep. Sci.* 36 (2013) 3896–3902.
- [33] O.B. Suhr, Y. Ando, P.I. Ohlsson, A. Olofsson, K. Andersson, E. Lundgren, Investigation into thiol conjugation of transthyretin in hereditary transthyretin amyloidosis, *Eur. J. Clin. Invest.* 28 (1998) 687–692.
- [34] O.B. Suhr, I.H. Svendsen, P.I. Ohlsson, J. Lendoire, P. Trigo, K. Tashima, Impact of age and amyloidosis on thiol conjugation of transthyretin in hereditary transthyretin amyloidosis, *Amyloid Int. J. Exp. Clin. Invest.* 6 (1999) 187–191.



Contents lists available at ScienceDirect

Talanta

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

Characterisation of serum transthyretin by electrospray ionisation-ion mobility mass spectrometry: Application to familial amyloidotic polyneuropathy type I (FAP-I)



Laura Pont^a, Fernando Benavente^{a,*}, Marta Vilaseca^b, Estela Giménez^a, Victoria Sanz-Nebot^a

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

^b Institute for Research in Biomedicine (IRB Barcelona), Barcelona, Spain

ARTICLE INFO

Article history:

Received 16 April 2015

Received in revised form

20 July 2015

Accepted 28 July 2015

Available online 1 August 2015

Keywords:

Familial amyloidotic polyneuropathy

Ion mobility

Mass spectrometry

Native immunoprecipitation

Neurodegeneration

Protein–protein

ABSTRACT

Transthyretin (TTR) is a homotetrameric protein which is known to misfold and aggregate causing different types of amyloidosis, such as familial amyloidotic polyneuropathy type I (FAP-I). FAP-I is associated with a specific TTR mutant variant (TTR (Met30)) that can be easily detected analysing the monomeric forms of the mutant protein. Meanwhile, the mechanism of protein aggregation onset, which could be triggered by structural changes on the native tetrameric protein complex, remains uncertain. We developed and described herein a new sample pretreatment based on immunoprecipitation (IP) to purify TTR from serum under non-denaturing conditions. Later, a nano-electrospray ionization-ion mobility mass spectrometry (nano-ESI-IM-MS or IM-MS) method was optimised to analyse the protein complexes in serum samples from healthy controls and FAP-I patients. IM-MS allowed separation and characterisation of tetrameric, trimeric and dimeric TTR gas ions due to their differential drift time, which is related to ion size and charge. The tetramer-to-dimer abundance ratio was differential between healthy controls and FAP-I patients (asymptomatic, symptomatic and an iatrogenic patient originally without the mutation who received a liver transplant from an FAP-I patient), and was also indicative of the effectiveness of liver transplantation as a treatment for FAP-I.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Amyloidoses (e.g. familial amyloidotic polyneuropathy (FAP), amyotrophic lateral sclerosis (ALS), prion disease, Huntington's disease, Alzheimer's disease and Parkinson's disease) are diseases in which some organ functions are destroyed as a result of protein deposits of insoluble amyloid aggregates [1–4]. Transthyretin (TTR) is known to precipitate as amyloid fibrils in FAP and senile systemic amyloidosis (SSA) [5,6]. TTR is predominantly expressed in the liver and in the choroid plexus of the brain, binds to thyroid hormones and transports thyroxine and retinol. It is highly abundant in serum (200–400 $\mu\text{g TTR mL}^{-1}$) and cerebrospinal fluid (CSF) (10–40 $\mu\text{g TTR mL}^{-1}$). Its structure is a homotetramer

composed of four non-covalently associated subunits [7–10]. Each monomer consists of 127 amino acid residues (approximately 14 kDa) with different post-translational modifications (PTMs) [10–13]. While SSA is a non-hereditary cardiomyopathy affecting 25% of the male population over the age of 80, in which deposits are formed from wild-type TTR, FAP is associated with around 73 of the about 100 point mutations known in the TTR-gene [5,6]. FAP type I (FAP-I), which is the most common genetically inherited FAP, is related with a genetic variant of TTR that presents a single amino acid substitution of valine for methionine at position 30 (TTR (Met30), $\Delta M_r = +32.07$) [14,15]. FAP is a fatal disease and liver transplantation is the most promising therapy today [16,17].

The mechanism involving aggregation in FAP, as in other amyloidoses, is not well-defined. For example, in prion disease, misfolding or dimerisation of cellular prion protein (PrP^C) has been reported to trigger aggregation [18–20]. Similarly, ALS could involve a monomeric misfolded intermediate of dimeric superoxide dismutase 1 (SOD-1) [21–23], and we recently found an increased amount of SOD-1 monomers, with regard to healthy controls, in erythrocytes from patients with ALS [23]. In FAP, it has been

Abbreviations: FAP-I, familial amyloidotic polyneuropathy type I; IM-MS, ion mobility mass spectrometry; IP, immunoprecipitation; M_r , relative molecular mass; TTR, transthyretin

* Correspondence to: Department of Analytical Chemistry, UB, Martí I Franqués 1-11, 3rd floor, 08028 Barcelona, Spain. Fax: +34 934021233.

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

<http://dx.doi.org/10.1016/j.talanta.2015.07.079>

0039-9140/© 2015 Elsevier B.V. All rights reserved.

hypothesised that TTR tetramer dissociates into misfolded monomers to trigger oligomerization and aggregation into a variety of structures including amyloid fibrils [24,25]. Therefore, clarification of the mechanisms behind protein aggregation is crucial to further our understanding of the onset of these neurodegenerative diseases.

Several techniques are currently used for the analysis of TTR in serum samples, including matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF-MS), electrospray ionization mass spectrometry (ESI-MS), high-performance liquid chromatography coupled to electrospray ionization mass spectrometry (HPLC-MS) and capillary electrophoresis electrospray mass spectrometry (CE-MS) [26–29]. In general, acidic conditions are preferred to ensure the best sensitivity in positive ion mode MS to characterise the TTR peptide sequence and the PTMs. However, neutral pH conditions are recommended for preservation of the native tetrameric form of TTR and the potential intermediates that could be involved in triggering aggregation in FAP. Anyway, these MS techniques have the drawback that they are unable to distinguish between ions with identical mass-to-charge (m/z) ratios, thus hindering the interpretation of mass spectra from coeluting or comigrating mixtures of monomers, dimers or other oligomers that are usually highly overlapped. IM spectrometry is a technique that separates gaseous ions in function of their size and charge (even those with the same m/z ratios) [23,30–33]. Specifically, IM spectrometry measures the time (drift time) taken for an ion to cross a region containing a background inert gas (usually N_2 and He) at a controlled pressure under the influence of a weak electric field. Drift time depends on ion-gas collisions; hence ions are separated on the basis of their ion-neutral collision cross-section (Ω), which is related to the overall shape and topology of the ion. Small ions cross first as a result of their smaller Ω . 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. Consequently, IM is often described as being proportional to collision cross-section-to-charge ratio (Ω/z) [34] and can be coupled on-line with MS (ion mobility mass spectrometry, IM-MS) for reliable analyte identification from molecular mass spectra [23,30–34].

Here, we evaluate the capacity of IM-MS to characterise standard human TTR and human TTR purified from serum under near native conditions to preserve TTR complexes, while obtaining molecular mass information from the different proteoforms (isoforms and PTMs) [35]. A nano-electrospray (nano-ESI) interface is used for the ionization because the low flow rates allow excellent ionization efficiency at mild conditions (low ionic strength aqueous volatile buffers at neutral pH) [23,30,31,34]. For the first time to our knowledge, we established an immunoprecipitation (IP) method to purify TTR from serum samples that does not use acidic solutions to dissociate the antigen–antibody complex [14,15,26,27], hence avoiding complete tetramer disruption. With the aim of solving this issue, some authors have described alternative but labour-intensive methods such as affinity chromatography, size exclusion chromatography or ion exchange chromatography [36–39].

The IP followed by IM-MS method was applied to the analysis of TTR in serum samples from healthy controls and FAP-I patients. The results permitted us to propose a simple, rapid and reliable method to screen, diagnose and evaluate patients with suspected TTR amyloidosis and to gain insight into TTR microheterogeneity to understand the mechanism triggering aggregation and disease.

2. Materials and methods

2.1. Chemicals and reagents

All the chemicals used in the preparation of buffers and solutions were of analytical reagent grade. Ammonia (25%) and sodium chloride ($\geq 99.5\%$) were purchased from Merck (Darmstadt, Germany). Ammonium acetate (NH_4Ac) ($\geq 99.9\%$), TTR ($\geq 95.0\%$), β -lactoglobulin A ($\geq 90.0\%$), avidin ($\geq 98.0\%$), bovine serum albumin ($\geq 99.0\%$) and concanavalin A ($\geq 99.0\%$) were provided by Sigma (St. Louis, MO, USA). Rabbit antihuman TTR polyclonal antibody was supplied by Dako (Glostrup, Denmark). Polyethylene glycol (PEG) 8000M_r ($\sim 50\%$ in water) was purchased from Fluka (Buchs, Switzerland). Water with a conductivity value lower than $0.05 \mu S cm^{-1}$ was obtained using a Milli-Q water purification system (Millipore, Molsheim, France).

2.2. Standard TTR samples

A TTR standard ($2000 \mu g mL^{-1}$) was prepared in 10 mM NH_4Ac (pH=7.00) and stored in a freezer at $-20^\circ C$ when not in use. Excipients of low M_r were removed from the TTR standard by passage through 10,000M_r cut-off cellulose acetate filters (Amicon Ultra-0.5, Millipore). The sample was centrifuged at $25^\circ C$ for 10 min at $11,000 \times g$, and the residue was washed three times for 10 min in the same way, with an appropriate volume of the neutral solution. The final residue was recovered by inverting the upper reservoir in a vial and spinning once more at a reduced centrifugal force (2 min at $300 \times g$). Sufficient neutral solution was added to adjust the TTR concentration to $2000 \mu g mL^{-1}$. The 10,000M_r filters were passivated before filtration with 5% v/v PEG solution to ensure optimum protein recoveries [40]. In order to passivate the filters, 500 μL of the passivation solution were pipetted into the reservoir of the centrifugal device, capped and allowed to stand 24 h at room temperature. The reservoirs were then uncapped and rinsed thoroughly with tap water. A volume of 500 μL of purified water was added to remove any residue and the centrifuge was then spun to dead stop (30 min at $11,000 \times g$). The remaining water was eliminated by inverting the reservoir in a vial and spinning once more at a reduced centrifugal force (2 min at $300 \times g$).

2.3. Immunoprecipitation of serum samples

Human blood samples from three healthy controls and four FAP-I patients were kindly supplied by the Hospital Universitari de Bellvitge (HUB, Hospitalet de Llobregat, Spain). The four FAP-I samples corresponded to an asymptomatic patient, a symptomatic patient (symptom onset was 4 years before blood collection), a liver-transplanted patient with the specific mutation (transplanted in 2013), and a iatrogenic patient originally without the mutation who was transplanted a liver from an FAP-I patient (transplanted 14 years before blood collection and symptom onset was 6 years before collection). The assay was approved by the Ethics Committee of the HUB and written informed consent was obtained from all participants in the study. 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. After that, the blood was kept at $4^\circ C$ for 12–16 h in order to improve the clot retraction. Then, the supernatant serum was separated from the clot with a Pasteur pipette and centrifuged at $1200 \times g$ for 20 min at $4^\circ C$. Clear serum was separated and aliquoted to store in a freezer at $-20^\circ C$ when not in use [27].

Under the optimised conditions, 50 μL of serum were incubated with 20 μL of rabbit anti-TTR antibody overnight at $4^\circ C$ in a

shaker. After centrifugation of the sample at $9000 \times g$ for 5 min at 4°C , the precipitate was washed twice with $100\ \mu\text{L}$ of solution of 0.9% m/v sodium chloride in water (saline) and twice with $100\ \mu\text{L}$ of purified water (by centrifugation as before). After the last centrifugation step, the supernatant was removed and the precipitate was dissolved by the addition of $50\ \mu\text{L}$ of ammonia $0.1\ \text{M}$. This sample solution was desalted by centrifugal filtration through $10,000\text{M}_r$ cut-off cellulose acetate filters. The sample was centrifuged at 25°C for 10 min at $13,000 \times g$, and the residue was washed at $14,000 \times g$ twice for 15 min and once for 20 min with $50\ \mu\text{L}$ of $10\ \text{mM}$ NH_4Ac ($\text{pH}=7.00$). The final residue was recovered by inverting the upper reservoir in a vial and spinning at a reduced centrifugal force (2 min at $300 \times g$). Sufficient NH_4Ac solution was added until the final volume was adjusted to $25\ \mu\text{L}$. These protein samples were kept in the freezer and thawed immediately before the analysis to prevent precipitation, spray instability and current breakdown during the IM-MS run. They were discarded after each injection.

2.4. Apparatus and procedures

Travelling wave (T-Wave) IM-MS experiments were performed on a Synapt G1 HDMS mass spectrometer (Waters, Manchester, UK). Samples were placed on a 384-well plate refrigerated at 15°C and were introduced by automated chip-base nano-ESI using a Triversa NanoMate (Advion BioSciences) in positive mode. Under the optimised conditions, the spray voltage was set to 2 kV and delivery pressure to 0.5 psi. Unless otherwise indicated, IM-MS (i.e. nano-ESI-IM-MS) was operated under the following conditions. The source pumping speed in the backing region of the mass spectrometer was reduced until achieving a pressure of 5.50 mbar to ensure optimal transmission of high mass non-covalent ions. Cone voltage, extraction cone and source temperature were set to 80 V, 5 V and 80°C , respectively. Ions passed through a quadrupole mass filter to the IM section of the instrument. This section comprised three T-Wave devices. The first device (Trap T-Wave) accumulated ions and released them in packets into the IM T-Wave, in which the mobility separation was performed. The final device (Transfer T-Wave) was used to transport the separated ions into the oa-TOF analyser. Trap and transfer collision energies were set to 10 V. The pressure in the Trap and Transfer T-Wave regions were $6.63 \cdot 10^{-2}$ mbar of Ar, and the pressure in the IM T-Wave was 0.5 mbar of N_2 . Trap gas and IM gas flows were 5 and $24\ \text{mL min}^{-1}$, respectively. The T-wave used in the IM T-Wave for mobility separation was operated at a velocity of $300\ \text{m s}^{-1}$. The wave amplitude was set at 9 V. The bias voltage for entering in the T-Wave cell was 20 V. The instrument was calibrated over the m/z range between 500 and 8000 Da using a solution of caesium iodide. MassLynx 4.1 and Driftscope 2.1 packages were used for MS and IM processing.

pH measurements were performed with a Crison 2002 potentiometer and a Crison electrode 52-03 (Crison Instruments, Barcelona, Spain). Centrifugal filtration was carried out in a cooled Rotanta 460 centrifuge (Hettich Zentrifugen, Tuttlingen, Germany) for centrifugation at controlled temperature (4 or 25°C). Shaking during sample incubation overnight was performed in a cold chamber with a Vortex Genius 3 (Ika®, Staufen, Germany).

2.4.1. Calibration of ion-neutral collision cross-section (Ω)

Ω values for ions transported by travelling waves are typically determined by calibrating drift times using ions of known Ω values. Calibrant solutions (β -lactoglobulin A, avidin, TTR, bovine serum albumin and concanavalin A) were prepared at a concentration of $50\ \mu\text{M}$ as described elsewhere [34]. IM-MS was operated under the optimised conditions described above. Calibrant drift times were corrected for mass-dependent flight time using

the equation $t'_D = t_D - [C\sqrt{\frac{m/z}{1000}}]$, where t'_D is the corrected drift time in ms, t_D is the experimental drift time in ms, m/z is the mass-to-charge ratio of the observed ion in Da/e and C is a constant. The constant C can be found within the control software of the Synapt G1 HDMS mass spectrometer and it is designated as the 'EDC (Enhanced Duty Cycle) delay coefficient' (1.57, in our case). Literature Ω values [34] were corrected for both ion charge state (z) and reduced mass (μ) $\left(\mu = \frac{M_{\text{ion}} \times M_{\text{N}_2}}{M_{\text{ion}} + M_{\text{N}_2}}\right)$ to calculate $\Omega' = \Omega/[z \times (1/\mu)^{1/2}]$. The plot was fitted to a potential relationship of the form $\Omega' = A \times (t'_D)^N$, where A and N are constants (in our case $A=7.98$ and $N=0.26$, $R^2 > 0.98$). This calibration plot was used to measure Ω' and later to calculate Ω for the TTR ions of interest in standard and serum samples.

3. Results and discussion

3.1. Development of a method for the analysis of TTR by IM-MS

In previous studies, we demonstrated the suitability of CE-MS at acidic and neutral pH to detect monomeric and dimeric forms of TTR in standard solutions [27]. We observed that tetramer was disrupted because of the acidic separation electrolyte, the acidic sheath liquid and the high vacuum inside the mass spectrometer. Furthermore, at neutral pH, the monomer (MO) and the dimer (DI) were not baseline resolved by CE, and their ESI mass spectra were highly overlapped because only the DI 'odd ions' (ions with odd charge) were specific for the dimeric form. In this regard, the mild experimental conditions required in IM-MS to separate gas-phase ions on the basis of their differential drift times (even those with the same m/z ratios), are preferable for studying such oligomers. Furthermore, by IM-MS, the ions are formed with minimal internal energy and consequently undergo little structural dissociation or reorganisation upon transfer into the gas phase. It should be stated, however, that careful control over experimental conditions, such as the use of aqueous volatile buffers (typically ammonium acetate), low ionic strength, physiological pH, low vacuum inside the mass spectrometer and mild gas phase protocols are necessary to maintain native-like structures.

First, we studied the influence of NH_4Ac concentration on the quality of the mass spectrum for the TTR standard solutions. Protein solutions in 10, 20 and 100 mM of NH_4Ac ($\text{pH}=7.00$) were tested. A neutral aqueous solution was required to avoid disruption of the tetrameric native structure. Furthermore, we found that the best spray stability and signal-to-noise ratio were obtained with the lowest ionic strength solution, 10 mM NH_4Ac ($\text{pH}=7.00$), that was chosen for further analysis.

Fig. 1A shows the mass spectrum of a $2000\ \mu\text{g mL}^{-1}$ TTR standard solution in 10 mM of NH_4Ac ($\text{pH}=7.00$) using a cone voltage of 80 V. As can be observed, in accordance with the m/z values shown in Table 1A, the mass spectrum presented the multiple-charge distributions caused by the MO (+7 and +6 ions), the tetramer (TE) (from +16 to +13 ions) and the octamer (OCT) (from +23 to +20 ions). Resolution of the mass spectrometer in the scanned m/z range (2000–7500) was not enough to separate the different proteoforms and only an average M_r could be calculated for the MO, TE and OCT. As can be observed, the MO experimental m/z values (Fig. 1A) agreed with the theoretical average m/z values taking into account the most abundant proteoforms (Table 1A). These proteoforms were wild-type Free-TTR and wild-type Free-TTR forming a mixed disulphide with the amino acid cysteine at position 10 (TTR-Cys) (see the footnote in Table 1) [27]. In a similar way, from comparison of the m/z values of Fig. 1A and Table 1A, the TE and OCT ions were easily explained

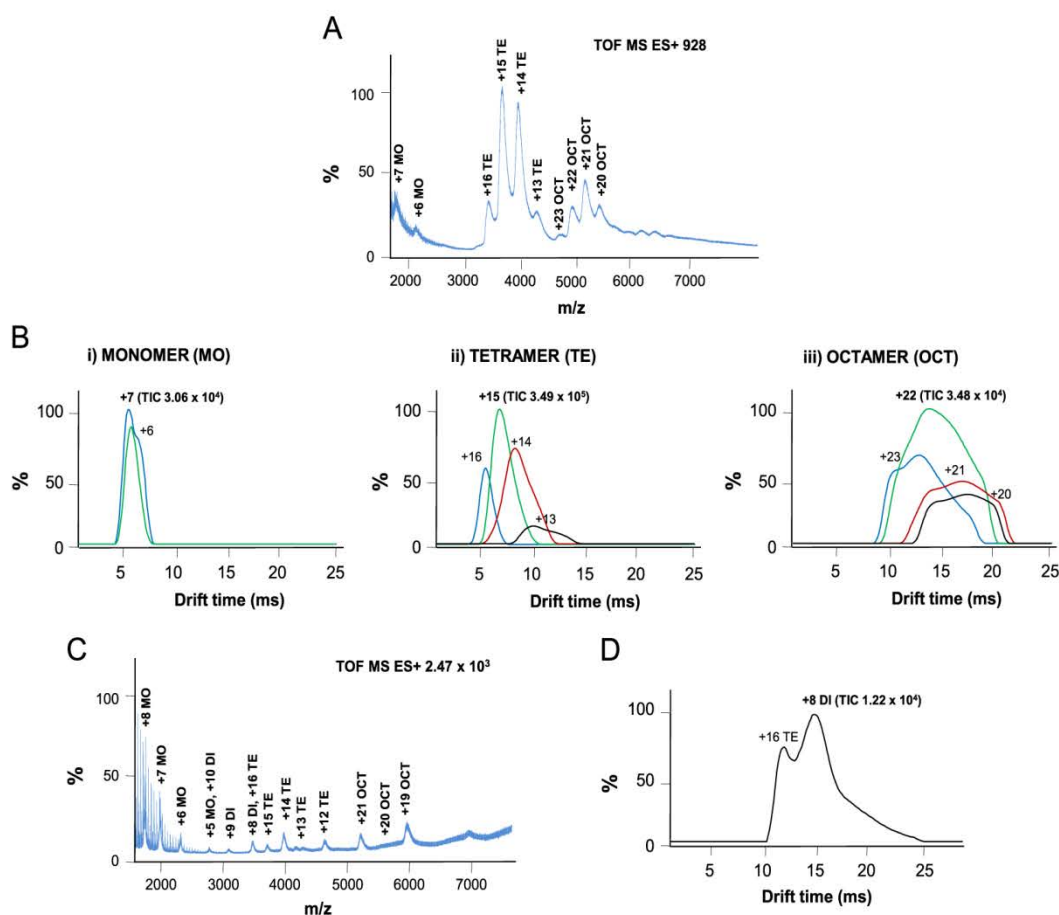


Fig. 1. Mass spectra (A and C) and extracted ion mobility profiles (EIMs) (B and D) obtained by IM-MS for a TTR standard solution at a concentration of $2000 \mu\text{g mL}^{-1}$ in $10 \text{ mM NH}_4\text{Ac}$ pH=7.00 at a cone voltage of 80 V (A and B) and 150 V (C and D).

considering homooligomers. For a better understanding of IM-MS mass spectra, ion drift time can be plotted against intensity to obtain extracted ion mobility (EIM) profiles (Fig. 1B). The distinct gas-phase species associated with each ion can be readily visualised from these profiles. Thus, an EIM profile for an ion m/z value showing a complex structure (two or more peaks or peaks with shoulders) indicates the presence of multiple species (MO, DI, etc.) or a single species with different non-interconverting conformations [23,30,31]. As it is shown in Fig. 1B, most of the EIM profiles presented one peak. Thus, single species were obtained for TTR MO, TE and OCT ions under these conditions. The presence of a small amount of MO was normal due to dissociation of the TE, and the OCT would be only stable in the gas-phase but not in solution, in which TTR is widely accepted to be a TE [7–10]. Focusing on MO, TE and OCT ions with different charges (e.g. +16TE, +15TE, +14TE and +13TE, Fig. 1B-ii), we observed that drift time decreased with charge. An increase in Ω with charge as a result of Coulomb repulsion was expected; however, Ω/z values are, in general, lower for higher order ions because drift times are proportional to Ω/z and charge increase predominates over Ω increase. At a cone voltage of 150 V (Fig. 1C), ions corresponding to the OCT and TE persisted, but in lower abundance, and DI and MO ions were also evident. Their presence was confirmed with the EIM profiles. Now, in some cases, no single peaks were obtained (i.e. +5MO and +10DI and +8DI and +16TE). As an example, Fig. 1D shows separation of +8DI and +16TE ions with the same m/z values but with different collision cross-section-to-charge ratios (Ω/z). For ions of a certain m/z value, TE ions travelled first because they showed a lower Ω/z value (i.e. $\Omega_{\text{TE}}/4z < \Omega_{\text{DI}}/2z$). In

general, the Ω for the TE (Ω_{TE}) will be less than twice the Ω of the DI (Ω_{DI}), because it is unlikely that DI ions unravel upon forming the TE, hence resulting in an extended structure. At a cone voltage of 150 V, the presence of DI ions and an increased amount of MO ions indicated that the TE and OCT ions were excessively activated, and hence underwent the initial stages of dissociation. We selected a cone voltage of 80 V as the best compromise between sensitivity and protein complex dissociation as a higher value may induce dissociation as well as conformational changes [23]. Other parameters which were very important to ensure optimal transmission of high mass non-covalent ions were the bias voltage and the trap collision energy, that were maintained low enough to prevent dissociation (20 V and 10 V, respectively) [23]. Furthermore the vacuum pressure in the backing region of the mass spectrometer was set at 5.50 mbar. At very low backing pressure (2.1 mbar), non-covalent interactions were broken because the gas pressure was not enough to promote enough collisions to allow the necessary loss of molecular internal energy.

Under the optimised conditions, the reproducibility values for drift times (%RSD, $n=3$, three independent analyses) of the most abundant molecular ions of the TTR standard solution (i.e. +7MO, +15TE and +22OCT, Fig. 1B) were 3.9%, 4.9% and 3.2%, respectively. Table 2 shows the measured Ω for these ions. As it is shown, %RSD of Ω was also good (1.8%, 1.3% and 0.8%, respectively), and both parameters allowed the unambiguous identification of the different ions. With regard to peak areas obtained from the EIM profiles, it was more reliable to evaluate reproducibility of peak area ratios instead of absolute peak areas to minimise the differences caused by sample matrix, the absolute protein

Table 1

Theoretical average m/z values of the detected TTR ions in (A) standard samples using 10 mM NH_4Ac (pH 7.00) and (B) healthy control and FAP-I sample (symptomatic) using the IP method followed by buffer exchange to 10 mM NH_4Ac (pH 7.00).

(A) Standard solutions			
		Theoretical average ^a <i>m/z</i>	
Monomer	+7MO	1975.7531	
	+6MO	2304.8773	
Tetramer	+16TE	3456.8119	
	+15TE	3687.1989	
	+14TE	3950.4982	
	+13TE	4254.3051	
Octamer	+23OCT	4809.0830	
	+22OCT	5027.6319	
	+21OCT	5266.9949	
	+20OCT	5530.2943	
B) Serum samples			
		Theoretical average <i>m/z</i>	
		Control sample ^a	FAP-I sample ^b (symptomatic)
Dimer and trimer	+13DI	2127.6565	2133.1404
	+12DI, +18TRI	2304.8773	2310.8181
	+17TRI	2440.3990	2446.6893
	+11DI	2514.3199	2520.8009
	+16TRI	2592.8609	2599.5455
	+10DI, +15TRI	2765.6512	2772.7802
	+14TRI	2963.1256	2970.7639
Tetramer	+16TE	3456.8119	3465.7232
	+15TE	3687.1989	3696.7042
	+14TE	3950.4982	3960.6826
	+13TE	4254.3051	4265.2729

^a Theoretical average values were calculated taking into account the TTR-Cys (52%) and Free-TTR (48%) abundances of the main proteoforms in wild-type TTR reported in our previous work by CE-MS for the standard solutions and control samples (Theoretical average M_r 13,823.2 (MO), 27,646.4 (DI), 41,469.6 (TRI), 55,292.9 (TE) and 110,585.7 (OCT), respectively) [27].

^b Theoretical average m/z values were calculated taking into account the TTR (Met30)-Cys (36%), TTR-Cys (33%), Free-TTR (19%) and Free-TTR(Met30) (12%) abundances of the main proteoforms in mutant TTR reported in our previous work by CE-MS for the symptomatic patient [27] (theoretical average M_r 27,717.4 (DI), 41,576.1 (TRI) and 55,434.8 (TE)).

concentrations and the possible precipitation (during analysis). Reproducibility of the peak area ratios of +22OCT/+7MO, +22OCT/+15TE and +15TE/+7MO were 5.3%, 8.1% and 2.8%, respectively. The limits of detection (LODs) for the analysis of TTR standard solutions were around $100 \mu\text{g mL}^{-1}$.

3.2. Analysis of TTR in serum by IM-MS

In previous studies, we optimised a rapid and simple sample pretreatment based on immunoprecipitation (IP) with a polyclonal anti-TTR rabbit antibody to analyse TTR in serum by CE-MS [27]. However, the use of an acidic solution of HAc 1.0 M to dissociate the antigen-antibody complex disrupted the tetrameric native structure and only TTR monomeric proteoforms were detected by CE-MS [27]. With the aim of solving this issue before IM-MS, several alternatives were tested. First, we verified whether the dissociation of the TTR tetrameric complex with acid was a reversible process or not. The mixture of TTR and antibody (resulting after IP with 50:50 μL serum:antibody volume ratio) was dissociated with HAc 1.0 M and, before analysis by IM-MS, centrifugal

Table 2

Measured Ω average (collision cross-section, nm^2) and %RSD ($n=3$) of the main detected TTR ions in a standard solution, in a healthy control and in a FAP-I sample (symptomatic) under the optimised IP IM-MS conditions.

Collision cross sections (Ω) average						
	Monomer		Tetramer		Octamer	
TTR standard	+7MO	%RSD ($n=3$)	+15TE	%RSD ($n=3$)	+22OCT	%RSD ($n=3$)
	16.5	1.8	36.7	1.3	63.0	0.8
	Dimer		Trimer		Tetramer	
	+11DI	%RSD ($n=3$)	+16TRI	%RSD ($n=3$)	+14TE	%RSD ($n=3$)
Control	26.3	0.3	38.2	0.05	37.7	0.2
FAP symptomatic	25.8	0.05	37.7	0.3	37.3	0.7

filtration (10,000 M_r cut-off cellulose acetate-filters) was used to exchange the acidic solution to 10 mM NH_4Ac , pH=7.00. Fig. 2A shows the mass spectrum obtained for a healthy control serum sample. As it is shown, the mass spectrum presented the multiple-charge distributions of the MO (from +9 to +7 ions), the TE (from +16 to +13 ions) and the OCT (from +22 to +19 ions), but TE and OCT ions were in very low abundance suggesting protein complex dissociation. The EIM profiles for the different ions of Fig. 2B indicate that all the ions were in general single species, as in the TTR standard sample (Fig. 1B), with the exception of +16TE and +15TE ions that were somehow splitted due to different conformations.

We tested as a novel alternative for dissociation of the antigen-antibody complex, NH_3 0.1 M followed by centrifugal filtration (10,000 M_r cut-off cellulose acetate-filters) to exchange the ammonia solution to 10 mM NH_4Ac , pH=7.00. However, the antibody had a negative influence on the current and spray stability. In order to prevent current breakdown during the analysis and improve reproducibility, we used 100,000 M_r cut-off centrifugal filters to separate TTR from the antibody prior to IM-MS. Unexpectedly, no TTR was recovered, probably because TTR was retained in the cellulose acetate membrane of the filters (TTR TE M_r is approximately 56,000). In order to avoid the 100,000 M_r cut-off filtration step to eliminate the antibody before exchanging the ammonia solution to 10 mM NH_4Ac , pH=7.00 (10,000 M_r cut-off), different serum:antibody volume ratios were tested, from 50:20 to 50:100 μL , and it was observed that a volume ratio of 50:20 allowed good TTR recoveries and stable current and spray during IM-MS analysis.

Fig. 3A shows the mass spectrum of TTR in 10 mM NH_4Ac (pH=7.00) after IP from a healthy control serum sample. As it is shown, the optimised IP procedure allowed preserving the TE to a large extent. The MO was not detected, but we did detect trimer (TRI) and DI ions. With regard to the OCT, a broad peak was observed, but sensitivity and resolution were not enough to assign the different ions. These differences observed in the mass spectrum with regard to the TTR standard solution (Fig. 1A) or the TTR control eluted with acid (Fig. 1B) were due to the matrix of the serum sample and the basic elution. In accordance with the m/z values shown in Table 1B (control sample), the mass spectrum presented the multiple-charge distributions characteristic of the DI (from +13 to 10 ions), the TRI (from +18 to +14 ions) and the TE (from +16 to +13 ions). Again, only an average M_r could be calculated for the DI, TRI and TE. As in standard solutions, the experimental m/z values (Fig. 3A) were close to the theoretical average m/z values taking into account the most abundant proteoforms of wild-type TTR (Free-TTR and TTR-Cys) (Table 1B, control sample, and footnote). The detection of the different ions

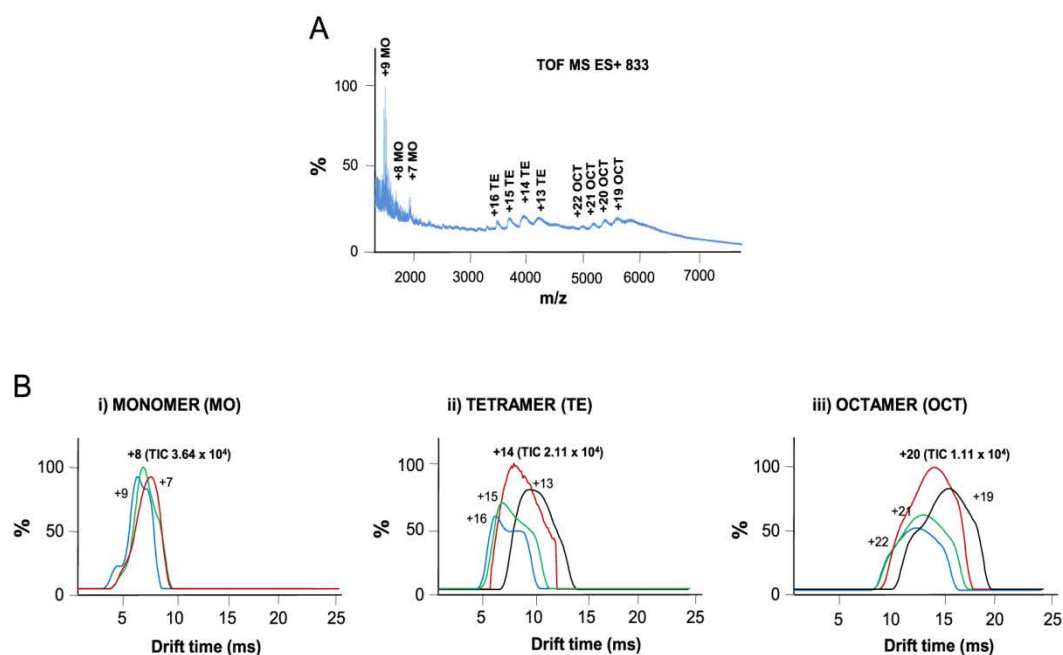


Fig. 2. Mass spectrum (A) and extracted ion mobility profiles (EIMs) (B) obtained by IM-MS for a control sample dissolving the antigen:antibody complex with HAc 1.0 M followed by acidic solution exchange to 10 mM NH_4Ac , pH=7.00 (IP serum:antibody volume ratio: 50:50 μL , non-optimised conditions).

was easily deduced from the EIM profiles of Fig. 3B. Again, in some cases, no single peaks were obtained because ions with the same m/z values but with different Ω/z were detected. A careful comparison of the m/z values of the MO, DI, TRI and TE ions shows that only the 'odd charge' DI ions are specific to the DI; all the TRI ions are specific to the TRI, with the exception of multiple of three ions; and only the 'odd charge' TE ions are specific to the TE. Thus, for example, the +18TRI ion and +12DI ion had the same m/z value (2304, Table 1B), but different drift time as can be observed in the EIM profile of Fig. 3B-i and ii (something similar happened with +15TRI and +10DI ions). As explained before, +18TRI ion migrates before than +12DI because it showed a lower Ω/z value (i.e. $\Omega_{\text{TRI}}/3z < \Omega_{\text{DI}}/2z$) (Fig. 3B). Fig. 3C shows the mass spectrum of TTR when the IM-MS method was applied to a FAP-I patient serum sample (FAP-I symptomatic). In accordance with the m/z values of the observed ions in Fig. 3C, which were close to the theoretical average m/z values of Table 1B (FAP-I symptomatic patient), the DI, TRI and TE were detected. Now, the experimental m/z values were close to the theoretical average m/z values taking into account the oxidised TTR-Cys proteoforms, mutant TTR (Met30)-Cys and wild-type TTR-Cys, which we found to be the most abundant proteoforms in patient samples in a previous work by CE-MS (see the footnote in Table 1) [27]. Several authors have also related oxidised mutant and normal forms as the trigger for the TTR misfolding and aggregation in FAP-I [27,41,42].

For the pretreated serum samples, reproducibility in drift times ($n=3$, three independent immunopurified samples) was similar to that obtained for the TTR standard solutions and, for example, values ranging between 0.8–1.2% and 1.6–4.9% were obtained for the most abundant molecular ions (i.e. +11DI, +16TRI and +14TE, Fig. 3B) of a healthy control and the symptomatic patient, respectively. Table 2 also shows the measured Ω for these TTR ions and the %RSD of Ω , which was again satisfactory (0.05–0.7%). The reproducibility in peak area ratios (+14TE/+11DI, +16TRI/+11DI and +14TE/+16TRI) was also good and, for example, values ranged between 3.3–7.6% and 3.1–12.3%, for the healthy control and the symptomatic patient, respectively.

To evaluate the potential of IM-MS to obtain a reliable fingerprint of human TTR structure that could be used to gain insight

into the mechanism underlying the disease, we performed a preliminary study with samples from a small group of healthy controls and FAP-I patient samples (i.e. asymptomatic, symptomatic, a liver transplanted patient with the specific mutation and a patient originally without the mutation who had the liver transplanted from an FAP-I patient (iatrogenic FAP-I)). All the blood samples were purified by IP and later analysed by IM-MS. The total concentration of TTR in serum and cerebrospinal fluid is variable and depends on the individual [27]. For these samples it was established in a previous study by CE-MS and ranged between 50 and 130 $\mu\text{g mL}^{-1}$ (see the insets on the x-axis of Fig. 4). In order to evaluate the relative abundances of DI, TRI and TE in control and patient samples, Fig. 4 shows a bar graph with the average ratios between the areas obtained from the EIMs for the peaks of +14TE/+11DI (black bars), +16TRI/+11DI (slashed bars) and +14TE/+16TRI (grey bars). As it is shown, TE/DI and TE/TRI ratios for the three control samples were lower than for the asymptomatic, symptomatic and iatrogenic FAP-I patients, while TRI/DI ratio was similar in all cases. TE/DI and TE/TRI ratios increased significantly for the asymptomatic, symptomatic and iatrogenic patients, whereas in the liver transplanted patient with the specific mutation these ratios were similar to the control ratios (measured m/z values were also close to the main wild-type proteoforms (Free-TTR and TTR-Cys)), hence confirming the effectiveness of the treatment. However, no clear trend was observed with regard to the symptom onset. The increase in TE relative abundance means that TTR in patients was less prone to dissociation. This could be related to the presence in patients of a higher amount of the oxidised TTR-Cys proteoforms. In a previous work, we found by CE-MS that especially the abundance of the mutant TTR(Met30)-Cys proteoform could be related with the symptom onset [27]. However, in this study resolution of the MS at these m/z values was not enough to differentiate between the mutant and wild-type oxidised TTR-Cys proteoforms. Some authors have reported that the natural ligand of TTR, thyroxine, provides a larger stability increase to the TE composed of mutant proteoforms than to the wild-type protein-ligand complex. Such small ligands bind to the hydrophobic interface of TTR and stabilise the TE assembly by raising the energetic barrier for

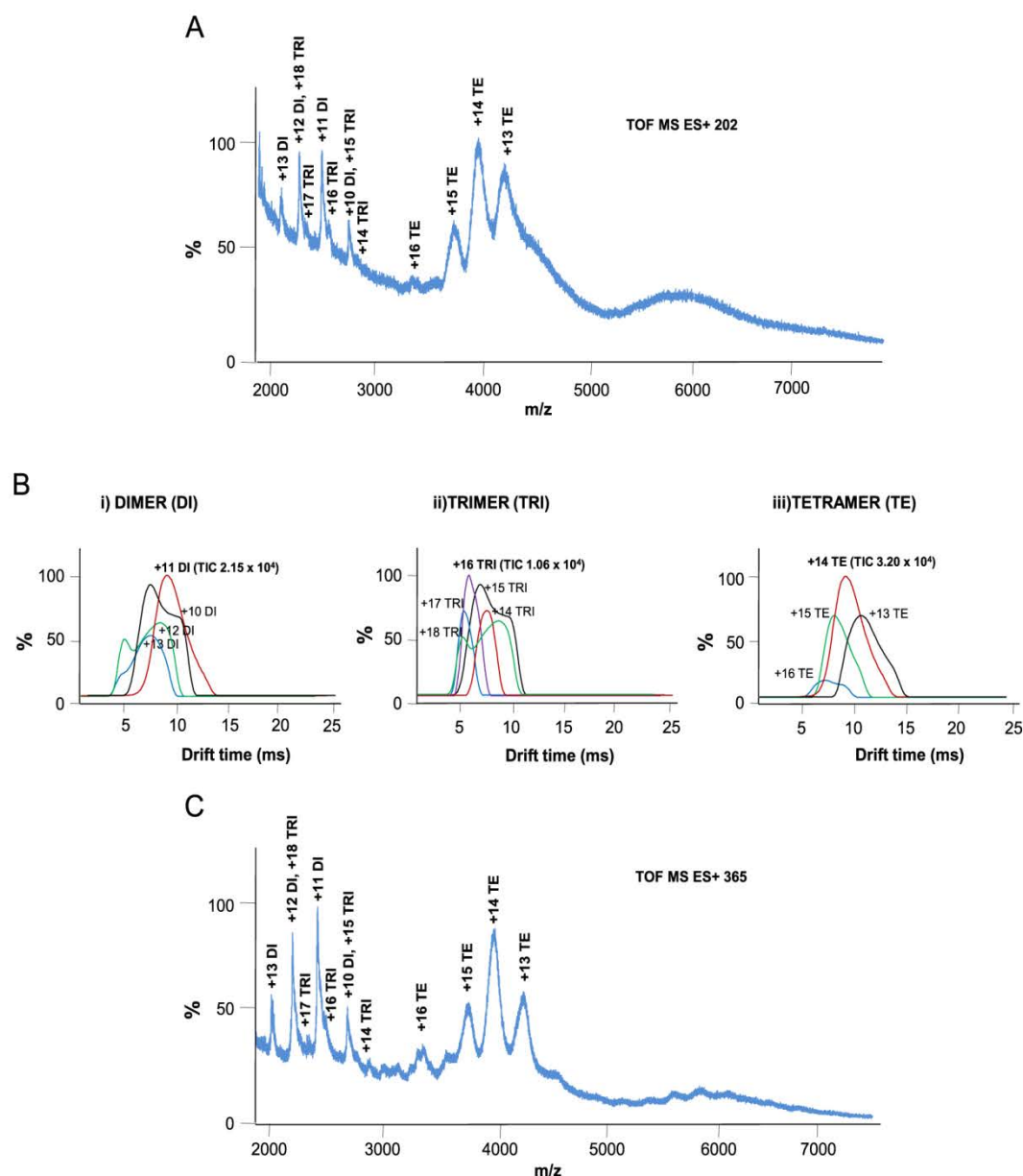


Fig. 3. Mass spectrum (A and C) and extracted ion mobility profiles (EIMs) (B) obtained by IM-MS for a control sample and a patient sample (symptomatic) using the optimised IP method (IP serum:antibody volume ratio: 50:20 μ L, antigen–antibody dissociation: NH_3 0.1 M, 10,000 M_r cut-off centrifugal filtration (exchange to 10 mM NH_4Ac pH=7.00)).

dissociation into monomers [32,43,44]. At this point, given the limited number of samples, it is difficult to give further explanations about the increased stability of TTR in patient samples but IM-MS shows a great potential to reliably gain insight into the different mechanisms triggering TTR aggregation in FAP-I. Our results highlight the importance of performing larger studies with carefully selected control and patient samples using higher resolution mass spectrometers.

4. Conclusions

We have established a novel IM-MS method that allows the detection and characterisation of wild-type and mutant TTR in standard and serum samples under non-denaturing conditions. First, we developed a sample pretreatment based on acid-free IP to purify TTR from serum samples before IM-MS analysis. Later, the most

appropriate instrumental conditions were selected in IM-MS to maintain native-like structures. Under optimised conditions, a comparison between TE/DI and TE/TRI abundance ratios in serum samples from healthy controls and FAP-I patients demonstrated the presence of a higher abundance and stability of TE in the asymptomatic, the symptomatic and the iatrogenic patients, from the main oxidised TTR forms (TTR(Met30)-Cys and TTR-Cys). In contrast, these ratios in liver transplanted patient were similar to those found in healthy controls from the main wild-type detected forms (Free-TTR and TTR-Cys), hence confirming the effectiveness of the treatment. Further studies with larger sets of samples using higher resolution mass spectrometers are required to evaluate the capacity of this technique to elucidate the molecular mechanism of aggregation that underlies FAP-I amyloidosis. This information may have implication for other neurodegenerative diseases, such Alzheimer's disease, Parkinson's disease and prion diseases, thus paving the way for early diagnosis, accurate prognosis or novel therapeutic targets.

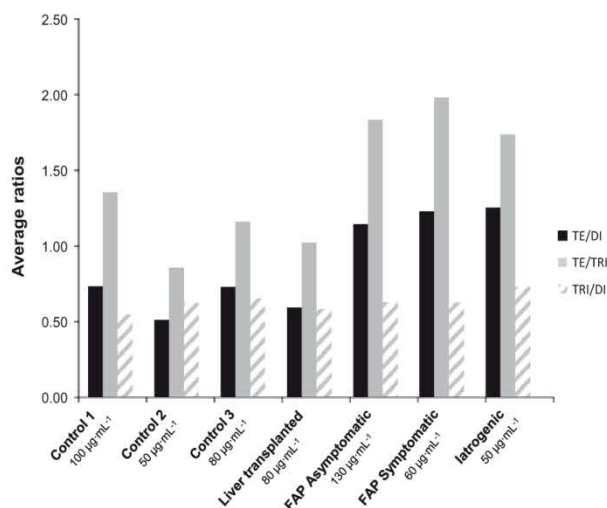


Fig. 4. Bar graph with the average ratios between the areas obtained for specific DI, TRI and TE ions by IM-MS of healthy controls and patient samples (TE/DI, TE/TRI and TRI/DI ratios). The area was measured from the extracted ion mobility profiles (EIMs) obtained for the ions +11DI, +16TRI and +14TE. The total concentration of TTR appears as insets (it was reported in a previous study [27]).

Acknowledgements

Laura Pont acknowledges the Spanish Ministry of Economy and Competitiveness for a FPI fellowship. This study was supported by a grant from the Spanish Ministry of Education and Science (CTQ2011-27130). The IRB Mass Spectrometry Core Facility is a Proteored group and participates in the European COST Action BM1403 (2014-2018); Native Mass Spectrometry and Related Methods for Structural Biology. We also thank Dr. C. Casasnovas and Dra. M.A. Alberti (Hospital Universitari de Bellvitge, HUB, Hospital de Llobregat, Spain) for providing the blood samples.

References

- [1] M.B. Pepys, J. Herbert, W.L. Hutchinson, G.A. Tennent, H.J. Lachmann, J. R. Gallimore, L.B. Lovat, T. Bartfai, A. Alanine, C. Hertel, T. Hoffmann, R. Jakob-Roetne, R.D. Norcross, J. a Kemp, K. Yamamura, M. Suzuki, G.W. Taylor, S. Murray, D. Thompson, A. Purvis, S. Kolstoe, S.P. Wood, P.N. Hawkins, Targeted pharmacological depletion of serum amyloid P component for treatment of human amyloidosis, *Nature* 417 (2002) 254–259.
- [2] C.M. Dobson, Protein folding and misfolding, *Nature* 426 (2003) 884–890.
- [3] S. Loizos, T. Shiakalli Chrysa, G.S. Christos, Amyloidosis: review and imaging findings, *Semin. Ultrasound CT MRI* 35 (2014) 225–239.
- [4] C. Soto, L.D. Estrada, Protein misfolding and neurodegeneration, *Neurol. Rev.* 65 (2014) 184–189.
- [5] R.H. Falk, R.L. Comenzo, M. Skinner, The systemic amyloidoses, *N. Engl. J. Med.* 337 (1997) 898–909.
- [6] A. Lim, T. Prokaveva, M.E. McComb, P.B. O'Connor, R. Théberge, L.H. Connors, M. Skinner, C.E. Costello, Characterization of transthyretin variants in familial transthyretin amyloidosis by mass spectrometric peptide mapping and DNA sequence analysis, *Anal. Chem.* 74 (2002) 741–751.
- [7] Y. Ingenbleek, V. Young, Transthyretin (prealbumin) in health and disease: nutritional implications, *Annu. Rev. Nutr.* 14 (1994) 495–533.
- [8] W. Zheng, Y.M. Lu, G.Y. Lu, Q. Zhao, O. Cheung, W.S. Blaner, Transthyretin, thyroxine, and retinol-binding protein in human cerebrospinal fluid: effect of lead exposure, *Toxicol. Sci.* 61 (2001) 107–114.
- [9] H.R. Bergen, S.R. Zeldenrust, S. Naylor, An on-line assay for clinical detection of amyloidogenic transthyretin isoforms in cerebrospinal fluid examined by immunoprecipitation and high-resolution mass spectrometry of intact protein, *Methods* 56 (2012) 284–292.
- [10] K. Poulsen, J.M.C. Bahl, J.T. Tanassi, A.H. Simonsen, N.H.H. Heegaard, Characterization and stability of transthyretin isoforms in cerebrospinal fluid examined by immunoprecipitation and high-resolution mass spectrometry of intact protein, *Methods* 56 (2012) 284–292.
- [11] B. Gericke, J. Raila, J. Sehoul, S. Haebel, D. Könsen, A. Mustea, F.J. Schweigert, Microheterogeneity of transthyretin in serum and ascitic fluid of ovarian cancer patients, *BMC Cancer* 5 (2005) 133–141.
- [12] T. Nakanishi, T. Sato, S. Sakoda, M. Yoshioka, A. Shimizu, Modification of cysteine residue in transthyretin and a synthetic peptide: analyses by electrospray ionization mass spectrometry, *Biochim. Biophys. Acta* 2004 (1698) 45–53.
- [13] H. Terazaki, Y. Ando, O. Suhr, P.I. Ohlsson, K. Obayashi, T. Yamashita, S. Yoshimatsu, M. Suga, M. Uchino, M. Ando, Post-translational modification of transthyretin in plasma, *Biochem. Biophys. Res. Commun.* 249 (1998) 26–30.
- [14] Y. Ando, O. Suhr, T. Yamashita, P.I. Ohlsson, G. Holmgren, K. Obayashi, H. Terazaki, C. Mambule, M. Uchino, M. Ando, Detection of different forms of variant transthyretin (Met30) in cerebrospinal fluid, *Neurosci. Lett.* 238 (1997) 123–126.
- [15] Y. Ando, P. Ohlsson, O. Suhr, N. Nyhlin, T. Yamashita, G. Holmgren, A. Danielsson, O. Sandgren, M. Uchino, M. Ando, A new simple and rapid screening method for variant transthyretin-related amyloidosis, *Biochem. Biophys. Res. Commun.* 483 (1996) 480–483.
- [16] S. Okamoto, J. Wixner, K. Obayashi, Y. Ando, M. Uchino, O.B. Suhr, Liver transplantation for familial amyloidotic polyneuropathy: impact on Swedish patients' survival, *Liver Transplant.* 15 (2009) 1229–1235.
- [17] E.A. Pomfret, W.D. Lewis, R.L. Jenkins, P. Bergeth, S.W. Dubrey, J. Reisinger, R. H. Falk, M. Skinner, Effect of orthotopic liver transplantation on the progression of familial amyloidotic polyneuropathy, *Transplantation*, 65, (1991) 918–925.
- [18] S.B. Prusiner, Prion diseases and the BSE crisis, *Science* 278 (1997) 245–251.
- [19] M. Borges-Alvarez, F. Benavente, M. Márquez, J. Barbosa, V. Sanz-Nebot, Evaluation of non-immunoaffinity methods for isolation of cellular prion protein from bovine brain, *Anal. Biochem.* 451 (2014) 10–17.
- [20] F. Moda, P. Gambetti, S. Notari, L. Concha-Marambio, M. Catania, K.-W. Park, E. Maderna, S. Suardi, S. Haik, J.-P. Brandel, J. Ironside, R. Knight, F. Tagliavini, C. Soto, Prions in the urine of patients with variant Creutzfeldt–Jakob disease, *N. Engl. J. Med.* 371 (2014) 530–539.
- [21] R. Rakhit, J.P. Crow, J.R. Lepock, L.H. Kondejewski, N.R. Cashman, A. Chakrabarty, Monomeric Cu,Zn-superoxide dismutase is a common misfolding intermediate in the oxidation models of sporadic and familial amyotrophic lateral sclerosis, *J. Biol. Chem.* 279 (2004) 15499–15504.
- [22] M. Borges-Alvarez, F. Benavente, J. Barbosa, V. Sanz-Nebot, Separation and characterization of superoxide dismutase 1 (SOD-1) from human erythrocytes by capillary electrophoresis time-of-flight mass spectrometry, *Electrophoresis* 33 (2012) 2561–2569.
- [23] M. Borges-Alvarez, F. Benavente, M. Vilaseca, J. Barbosa, V. Sanz-Nebot, Characterization of superoxide dismutase 1 (SOD-1) by electrospray ionization-ion mobility mass spectrometry, *J. Mass Spectrom.* 48 (2013) 60–67.
- [24] A. Shimizu, T. Nakanishi, M. Kishikawa, A. Miyazaki, Detection and identification of protein variants and adducts in blood and tissues: an application of soft ionization mass spectrometry to clinical diagnosis, *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* 776 (2002) 15–30.
- [25] A. Shimizu, T. Nakanishi, A. Miyazaki, Detection and characterization of variant and modified structures of proteins in blood and tissues by mass spectrometry, *Mass Spectrom. Rev.* 25 (2006) 686–712.
- [26] N.H.H. Heegaard, M.Z. Hansen, J.W. Sen, M. Christiansen, P. Westermarck, Immunoaffinity chromatographic and immunoprecipitation methods combined with mass spectrometry for characterization of circulating transthyretin, *J. Sep. Sci.* 29 (2006) 371–377.
- [27] 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, *Electrophoresis* (2015) 10.1002/elps.201400590, (in press).
- [28] R. Théberge, L.H. Connors, M. Skinner, C.E. Costello, Detection of transthyretin variants using immunoprecipitation and matrix-assisted laser desorption/ionization bioreactive probes: a clinical application of mass spectrometry, *J. Am. Soc. Mass Spectrom.* 11 (2000) 172–175.
- [29] R. Théberge, L. Connors, M. Skinner, J. Skare, C.E. Costello, Characterization of transthyretin mutants from serum using immunoprecipitation, HPLC/electrospray ionization and matrix-assisted laser desorption/ionization mass spectrometry, *Anal. Chem.* 71 (1999) 452–459.
- [30] B.T. Ruotolo, S.-J. Hyung, P.M. Robinson, K. Giles, R.H. Bateman, C.V. Robinson, Ion mobility-mass spectrometry reveals long-lived, unfolded intermediates in the dissociation of protein complexes, *Angew. Chem. Int. Ed. Engl.* 46 (2007) 8001–8004.
- [31] 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.
- [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] 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.
- [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] L.M. Smith, N.L. Kelleher, Proteoform: a single term describing protein complexity, *Nat. Methods* 10 (2013) 186–187.
- [36] M.V. Bimani, P.S. Ghaswala, Isolation, purification and partial characterization of prealbumin from cerebrospinal fluid, *J. Biosci.* 13 (1988) 159–169.
- [37] A. Mahn, E. Lienqueo, C. Quilodrán, A. Olivera-Nappa, Purification of transthyretin as nutritional biomarker of selenium status, *J. Sep. Sci.* 35 (2012) 3184–3189.
- [38] M. Navab, A.K. Mallia, Y. Kanda, D.S. Goodman, Rat plasma prealbumin.

- Isolation and partial characterization, *J. Biol. Chem.* 252 (1977) 5100–5106.
- [39] P. Raghu, P. Ravinder, B. Sivakumar, A new method for purification of human plasma retinol-binding protein and transthyretin, *Biotechnol. Appl. Biochem.* 38 (2003) 19–24.
- [40] L. Pont, F. Benavente, J. Barbosa, V. Sanz-Nebot, An up-date for human blood plasma pretreatment for optimized recovery of low-molecular-mass peptides prior to CE-MS and SPE-CE-MS, *J. Sep. Sci.* 36 (2013) 3896–3902.
- [41] O.B. Suhr, Y. Ando, P.I. Ohlsson, A. Olofsson, K. Andersson, E. Lundgren, M. Ando, G. Holmgren, Investigation into thiol conjugation of transthyretin in hereditary transthyretin amyloidosis, *Eur. J. Clin. Invest.* 28 (1998) 687–692.
- [42] O.B. Suhr, I.H. Svendsen, P.-I. Ohlsson, J. Lendoire, P. Trigo, K. Tashima, P. J. Ranlov, Y. Ando, Impact of age and amyloidosis on thiol conjugation of transthyretin in hereditary transthyretin amyloidosis, *Amyloid Int. J. Exp. Clin. Invest.* 6 (1999) 187–191.
- [43] P. Hammarström, R.L. Wiseman, E.T. Powers, J.W. Kelly, Prevention of transthyretin amyloid disease by changing protein misfolding energetics, *Science* 299 (2003) 713–716.
- [44] S.-J. Hyung, C.V. Robinson, B.T. Ruotolo, Gas-phase unfolding and disassembly reveals stability differences in ligand-bound multiprotein complexes, *Chem. Biol.* 16 (2009) 382–390.



Contents lists available at ScienceDirect

Talanta

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

On-line immunoaffinity solid-phase extraction capillary electrophoresis mass spectrometry using Fab' antibody fragments for the analysis of serum transthyretin



Laura Pont, Fernando Benavente, José Barbosa, Victoria Sanz-Nebot*

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

ARTICLE INFO

Keywords:

Capillary electrophoresis
Fab' fragments
Mass spectrometry
On-line immunopurification
Transthyretin

ABSTRACT

This paper describes an on-line immunoaffinity solid-phase extraction capillary electrophoresis mass spectrometry (IA-SPE-CE-MS) method using an immunoaffinity sorbent with Fab' antibody fragments (Fab'-IA) for the analysis of serum transthyretin (TTR), a homotetrameric protein ($M_r \sim 56,000$) involved in different types of amyloidosis. The IA sorbent was prepared by covalent attachment of Fab' fragments obtained from a polyclonal IgG antibody against TTR to succinimidyl silica particles. The Fab'-IA-SPE-CE-MS methodology was first established analyzing TTR standard solutions. Under optimized conditions, repeatability and reproducibility were acceptable, the method was linear between 1 and 25 $\mu\text{g mL}^{-1}$, limits of detection (LODs) were around 0.5 $\mu\text{g mL}^{-1}$ (50-fold lower than by CE-MS, $\sim 25 \mu\text{g mL}^{-1}$) and different TTR conformations were observed (folded and unfolded). The applicability of the developed method to screen for familial amyloidotic polyneuropathy type I (FAP-I), which is the most common hereditary systemic amyloidosis, was evaluated analyzing serum samples from healthy controls and FAP-I patients. For the analysis of sera, the most abundant proteins were precipitated with 5% (v/v) of phenol before Fab'-IA-SPE-CE-MS. The current method enhanced our previous results for the analysis of TTR using intact antibodies immobilized on magnetic beads. It allowed a slight improvement on LODs (2-fold), the detection of proteoforms found at lower concentrations and the preparation of microcartridges with extended durability.

1. Introduction

In recent years, increased emphasis has been placed on the detection, characterization and quantification of normal and variant forms (proteoforms) of different proteins as predictive indicators of an ongoing disease state [1–4]. For this purpose, some advanced analytical techniques are available, such as enzyme-linked immunosorbent assay (ELISA), immunoblotting, radioimmunoassay, polyacrylamide gel electrophoresis (PAGE), two dimensional gel electrophoresis (2D-PAGE), liquid chromatography mass spectrometry (LC-MS) and capillary electrophoresis mass spectrometry (CE-MS). Unfortunately, although the usefulness of these techniques for analyzing a wide range of protein biomarkers, those that are the most sensitive (i.e. immunological techniques) are not informative enough to confirm compound identity [5–7], while high performance separation techniques coupled to mass spectrometry are often not sensitive enough to detect some protein biomarkers found at very low concentrations in certain biological samples [8–10].

In order to overcome these major drawbacks, we propose on-line

immunoaffinity solid-phase extraction capillary electrophoresis (IA-SPE-CE) for the targeted enrichment and clean-up of different compounds from complex samples, including large biomolecules (e.g. proteins and glycoproteins). IA-SPE-CE is a state-of-the-art hyphenated technology that combines the high selectivity of immunocapture by IA-SPE with the high separation efficiency of CE, which can be also coupled on-line with MS detection for a reliable compound identification (IA-SPE-CE-MS) [7,11–16]. IA-SPE-CE is a variant of SPE-CE based on immunosorbents prepared by immobilization of antibodies or antibody fragments that are placed inside the separation capillary (or the channels of a microchip). In one of the typical configurations, an extraction microcartridge or analyte concentrator is inserted near the inlet of the separation capillary. This microcartridge contains the IA sorbent, which retains the antigenic analyte from a large volume of sample (~ 50 – $100 \mu\text{L}$). Then, the retained analyte is eluted in a small volume of an appropriate solution (~ 25 – 50 nL), resulting in sample clean-up and concentration enhancement before electrophoretic separation and detection [11,13–16]. IA-SPE-CE has been most often described with laser-induced-fluorescence detection (LIF) for preconcentration of small molecules, mostly peptides, from biological samples

* Corresponding author.

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

<http://dx.doi.org/10.1016/j.talanta.2017.03.104>

Received 19 December 2016; Received in revised form 29 March 2017; Accepted 31 March 2017

Available online 04 April 2017

0039-9140/ © 2017 Published by Elsevier B.V.

[7,12,17–19]. In contrast, IA-SPE-CE-MS has been demonstrated to a much lesser extent, probably due to the difficulties of making compatible the requirements of IA-SPE-CE with on-line MS detection [11,13–16]. Reports on the analysis of large biomolecules by “fully on-line” IA-SPE-CE-MS are especially scarce due to the additional issues related to low extraction efficiency, poor ionization efficiency and adsorption onto the inner capillary walls [15,16]. We have recently described two different IA-SPE-CE-MS methods based on immunosorbents prepared with intact antibodies immobilized on conventional silica particles or magnetic beads for the analysis of transferrin and transthyretin (TTR) from serum samples [15,16].

The major drawback of typical chromatographic sorbents used in SPE (e.g. C₈, C₁₈, polymeric with hydrophilic-lipophilic balance, etc.) is their limited selectivity, which hinders the analysis of complex samples such as biological fluids [20,21]. In this regard, IA sorbents are an interesting alternative with improved selectivity, which may provide excellent extraction efficiency if the immunoreactivity and orientation of the antibody and the active surface area are optimum and non-specific adsorption is minimized. In recent years, different activated supports for immunoextraction have become available at a reasonable price, with a wide range of surface properties to easily and reproducibly couple intact antibodies. We have recently shown the good performance of commercial magnetic beads derivatized with and intact antibody to analyze TTR from serum samples by off-line IA-SPE and CE-MS and IA-SPE-CE-MS [16]. However, preconcentration factor enhancement, clean-up efficiency and non-specific adsorption are issues that need to be further addressed with sorbents prepared with these commercial supports. Several strategies have been described for the preparation of immunosorbents [22,23]. Although direct adsorption of antibodies to activated surfaces has been used in affinity chromatography, such approaches are not sufficient for IA-SPE-CE because of the limited stability. The covalent attachment of the antibody to a chemically modified support surface is more effective, using in general a secondary molecule as a spacer arm to ensure the largest active surface area. The immunoreactive area can be further enlarged using antibody fragments instead of intact antibodies. In this regard, the supports activated to contain free sulfhydryl (thiols) groups are highly appreciated for optimum orientation of Fab' antibody fragments, which also contain free thiol groups to form a disulfide bridge [24–26]. Here, we show for the first time to the best of our knowledge, that an Fab'-IA sorbent can be used for the analysis of large biomolecules (i.e. TTR) by IA-SPE-CE-MS.

In this paper, an IA-SPE-CE-MS method using a lab-made IA sorbent with Fab' antibody fragments is described for the analysis of serum TTR. TTR is a homotetrameric protein composed of four identical monomers (MO) (M_r~14,000) with different proteoforms (isoforms and post-translational modifications (PTMs)) [27–29]. TTR is known to misfold and aggregate as stable insoluble amyloid fibrils, which are characteristic of different neurodegenerative diseases known as amyloidosis [30,31]. Familial amyloidotic polyneuropathy type I (FAP-I), which is the most common type of hereditary amyloidosis, is associated with a TTR variant that presents a single amino acid substitution of valine for methionine at position 30 (Met 30) [29,32,33]. The Fab'-IA-SPE-CE-MS method was developed using TTR standard solutions and, under optimized conditions, serum samples from healthy controls and FAP-I patients were analyzed to evaluate the potential of the proposed methodology to reliably screen for FAP-I. The results with the novel Fab'-IA sorbent (conventional silica particles) were also compared with those reported before with an intact antibody IA sorbent (magnetic beads) [16].

2. Materials and methods

2.1. Chemicals and reagents

All the chemicals used in the preparation of background electrolytes

(BGEs) and solutions were of analytical reagent grade. Propan-2-ol (≥99.9%), methanol (MeOH, ≥99.9%), acetonitrile (ACN, ≥99.9%), formic acid (HFor, 99.0%), ammonia (NH₃, 25.0%), ortho-phosphoric acid (85.0%), hydrochloric acid (HCl, 25.0%), sodium hydrogenphosphate (≥99.0%), sodium chloride (NaCl, ≥99.5%), sodium borate (98.0%), sodium hydroxide (NaOH, ≥99.0%, pellets), potassium dihydrogenphosphate (≥99.0%), potassium chloride (99.0%), phenol (≥99.5%), boric acid (99.8%), glycine (99.7%) and TTR (≥95.0%) were purchased from Merck (Darmstadt, Germany). Ammonium acetate (NH₄Ac, ≥99.9%), ethylenediaminetetraacetic acid disodium salt 2-hydrate (EDTA, 99.0–101.0%), 2-mercaptoethylamine-HCl (2-MEA, 98.0%), glycidol (96.0%) and PEG 8,000 M_r (~50% in water) were provided by Sigma (St. Louis, MO, USA). ImmunoPure Fab' micro preparation kit (Fab' kit) and sodium sulfosuccinimidyl 4-(N-maleimidomethyl) cyclohexane-1-carboxylate (SSMCC) were supplied by Fisher Scientific (Madrid, Spain). Rabbit antihuman TTR polyclonal antibody was purchased from Dako (Glostrup, Denmark). Water with a conductivity value lower than 0.05 μS cm⁻¹ was obtained using a Milli-Q water purification system (Millipore, Molsheim, France). Sep-Pak Classic NH₂ cartridges (55–105 μm particle diameters, 125 Å (1 Å=0.1 nm) pore size, 970 μmol of NH₂ per gram of particle) were provided by Waters (Milford, MA, USA).

2.2. Electrolyte solutions, sheath liquid, protein standards and serum samples

The BGE for CE-MS and IA-SPE-CE-MS was a solution of 10 mM of NH₄Ac adjusted to pH 7.00 with NH₃. The BGE was passed through a 0.45 μm nylon filter (MSI, Westboro, MA, USA). The 100 mM phosphate buffer solution (PBS) for antibody fragmentation was part of the Fab' kit. The PBS for the sample pretreatment was prepared with 0.011 M sodium hydrogenphosphate, 0.0015 M potassium dihydrogenphosphate, 0.14 M sodium chloride and 0.0027 M potassium chloride (pH 7.20). The 50 mM borate buffer solution for the preparation of the IA sorbent was adjusted to pH 7.60 with boric acid. A phosphate buffer solution for the storage of the IA sorbent was prepared from 100 mM of potassium dihydrogen phosphate and it was adjusted to pH 7.40 with NaOH (1.0 M). The sheath liquid solution consisted of a mixture of 60:40 (v/v) propan-2-ol: water with 0.25% (v/v) of HFor. The sheath liquid and the BGEs were degassed for 10 min by sonication before use.

An aqueous standard solution (1,000 μg mL⁻¹) of TTR was prepared and stored in a freezer at -20 °C when not in use. Excipients of low M_r were removed from the sample by passage through 10,000 M_r cut-off cellulose acetate filters (Amicon Ultra-0.5, Millipore). The sample was centrifuged at 25 °C for 10 min at 11,000×g, and the residue was washed three times for 10 min in the same way, with an appropriate volume of PBS. The final residue was recovered by inverting the upper reservoir in a vial and spinning once more at a reduced centrifugal force (2 min at 300×g). Sufficient PBS was added to adjust the concentration of TTR to 1,000 μg mL⁻¹. It is worth mentioning that it was necessary to passivate the centrifugal filters with an aqueous solution of 5% (v/v) of PEG to avoid TTR loss through adsorption [29].

Human blood samples from a healthy control and a symptomatic FAP-I patient were kindly supplied by the Hospital Universitari de Bellvitge (HUB, Hospitalet de Llobregat, Spain). The assay was approved by the Ethics Committee of the HUB and written informed consent was obtained from all participants in the study. Serum was prepared as described elsewhere [29]. Serum aliquots were stored in a freezer at -20 °C when not in use.

2.3. Apparatus and procedures

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

Centrifugal filtration was carried out in a cooled Rotanta 460 centrifuge (Hettich Zentrifugen, Tuttlingen, Germany) for centrifugation at controlled temperature (4 or 25 °C). Agitation was performed with a Vortex Genius 3 (Ika®, Staufen, Germany). Incubations were carried out in a TS-100 thermoshaker (Biosan, Riga, Latvian Republic).

Fused silica capillaries were supplied by Polymicro Technologies (Phoenix, AZ, USA). All CE-MS experiments were performed in an HP^{3D} CE system coupled with an orthogonal G1603A sheath-flow interface to a 6220 oa-TOF LC/MS spectrometer (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). ChemStation and MassHunter softwares (Agilent Technologies) were used for CE and mass spectrometer control, data acquisition, integration and *m/z* mass spectra deconvolution. The mass spectrometer was operated under optimum conditions in positive mode with the parameters established in previous works [16,29].

2.3.1. Preparation of the IA sorbent

The IA sorbent was prepared as described elsewhere [14] following the “maleimide method” to immobilize Fab’ fragments through their sulfhydryl groups on commercial aminopropyl silica particles with succinimidyl groups (see Supplementary Fig 1).

First, the aminopropyl silica particles were derivatized to obtain the activated succinimidyl silica particles by reaction with SSMCC (Supplementary Fig 1-A). 0.01 g of aminopropyl silica particles were washed twice with water and resuspended in 805 µL of a solution of 3 mg mL⁻¹ of SSMCC in 50 mM sodium borate buffer (pH 7.60). The mixture was then moderately shaken for 1 h at 30 °C. The resulting activated succinimidyl silica particles were washed twice with 50 mM of borate buffer. The particles were separated after each washing by centrifugation (2 min at 4000×*g*). Finally, they were stored in 250 µL of borate buffer at 4 °C. In order to prepare a sorbent for testing end-capping of the unreacted aminopropyl groups, 30 µL of glycidol were added to the activated succinimidyl silica particles (the storage borate buffer was removed first) and the mixture was shaken 24 h [34]. The end-capped silica particles were washed with 200 µL of water and methanol, dried under vacuum and washed with potassium dihydrogen phosphate buffer.

Second, the polyclonal IgG antibody against TTR was fragmented and the Fab’ fragments were purified using the Fab’ kit according to the manufacturer’s instructions (Supplementary Fig 1-B). 125 µL of a 200 mg mL⁻¹ IgG solution were digested using immobilized pepsin and the F(ab’)₂ fragments were purified with an immobilized Protein A spin column. Then, the solvent of the obtained F(ab’)₂ solution was changed to 100 mM of PBS containing 1 mM of EDTA by filtration through 50,000 M_r cut-off cellulose acetate filters (Amicon Ultra-0.5). Under optimum conditions, 125 µL of F(ab’)₂ solution were mixed with the same volume of a 100 mM 2-MEA solution, hence generating the Fab’ fragments. The reaction mixture was then moderately shaken for 30 min at 37 °C. Fab’ fragments were separated by filtration through 30,000 M_r cut-off cellulose acetate filters (Amicon Ultra-0.5). The purified Fab’ retentate was recovered by centrifugation upside down into a new vial (2 min at 1000×*g*) and borate buffer was added until the final volume was adjusted to 125 µL.

For the covalent binding of the Fab’ fragments to the activated succinimidyl silica particles through their sulfhydryl group (Supplementary Fig 1-C), 125 µL of the Fab’ solution and 0.01 g of the activated silica particles were mixed and allowed to react with constant stirring for 24 h at 4 °C. After the reaction, the sorbent was washed three times with 200 µL of 100 mM phosphate buffer (pH 7.40) and it was stored at 4 °C in the same buffer. For periods longer than one month, the Fab’-IA sorbent was stored with the phosphate buffer solution containing also 0.1% (m/v) of sodium azide.

2.3.2. IA-SPE-CE-MS

The construction of the IA microcartridge or analyte concentrator

was carried out as described elsewhere [13–16]. The microcartridge (0.7 cm L×250 µm id×365 µm od fused silica capillary), packed with the Fab’-IA sorbent, was inserted using two plastic sleeves at 7.5 cm from the inlet of a conditioned separation capillary (72 cm L_T×75 µm id×365 µm od fused silica capillary). The sorbent particles were retained between two frits obtained from the material of the original frits found in the Sep-Pak Classic NH₂ cartridges.

In all these experiments, a neutral BGE (10 mM of NH₄Ac adjusted to pH 7.00 with NH₃) was used to avoid extreme pH that would cause Fab’ denaturation and TTR elution. Capillaries were first conditioned flushing at 930 mbar for 3 min with BGE. TTR standards in PBS and serum samples were hydrodynamically introduced at 930 mbar for 10 min (~50 µL using the Hagen-Poiseuille equation [35]). A final rinse with BGE (2 min at 930 mbar) eliminated non-retained molecules and equilibrated the capillary before the electrophoretic separation. Under optimized conditions, an eluent of 100 mM glycine/100 mM NH₃ (pH=9.50) (standard solutions) or 200 mM glycine/200 mM NH₃ (pH=9.50) (serum samples) was injected at 50 mbar for 10 s (~50 nL). All these steps were performed by switching off the nebulizer gas and the ESI capillary voltage to prevent the entrance of non-volatile contaminants into the mass spectrometer. Then, both were switched on and separation was carried out at 25 °C and +25 kV (normal polarity, cathode in the outlet). Between runs, the capillary was rinsed with the eluent and water (2 min at 930 mbar) to avoid carry-over.

A simple off-line sample pretreatment was necessary to analyze TTR in serum samples, in order to prevent microcartridge saturation and capillary inner surface damage due to the presence of other high-abundance proteins, mainly albumin [16]. At 2 °C, 8 mg of NaCl were added to 100 µL of human serum and 100 µL of 5% (v/v) phenol. After sample centrifugation for 10 min at 11,000×*g*, the supernatant was collected and subjected to a second precipitation with 100 µL of 5% (v/v) phenol. The supernatant was again collected and diluted 1:1 (v/v) with PBS before analysis.

2.3.3. Quality parameters

All quality parameters were calculated from data obtained by measuring migration times (*t_m*) and peak areas from the extracted ion electropherograms (EIEs) of TTR MO proteoforms (considering the *m/z* of the most abundant molecular ions for each proteoform, i.e. +7, +8, +9, +13, +14 and +15 MO ions). In Fab’-IA-SPE-CE-MS the contribution of both conformers was considered together to get a unique integration for each TTR proteoform (in standard solutions and serum samples). Repeatability (*n*=6) and reproducibility (*n*=9, three replicates and three different microcartridges) were evaluated as the relative standard deviation (%RSD) of *t_m* and peak areas of a 25 µg mL⁻¹ TTR standard solution. Linearity range was studied by analyzing standard solutions of TTR at concentrations between 0.5 and 50 µg mL⁻¹. An estimation of the limits of detection (LODs) was obtained by analyzing low-concentration standard solutions of TTR (close to the LOD level, as determined from the approach based on *S/N*=3). The lifetime of the microcartridges was evaluated by repeatedly analyzing a 25 µg mL⁻¹ standard solution of TTR and pretreated serum samples.

3. Results and discussion

3.1. Analysis of TTR standards by Fab’-IA-SPE-CE-MS

The conditions established for IA-SPE-CE-MS in our previous work, where magnetic beads were used for the preparation of an intact antibody IA sorbent, were the starting point for the establishment of the method with the Fab’-IA sorbent [16]. Again, a neutral BGE (10 mM of NH₄Ac adjusted to pH 7.00 with NH₃) was necessary because the use of an acidic BGE with an IA sorbent would cause Fab’ denaturation and TTR elution during system conditioning and clean-up. As showed for the intact antibody sorbent [16], TTR standards

needed to be dissolved in PBS, because the protein was not retained when dissolved in water or neutral BGE. PBS probably benefited the proper interaction between TTR and the Fab' fragments, because it had a similar osmolarity and ion concentration to the human body fluids.

The method adaptation with regard to the eluent composition was not so straightforward. First, the basic volatile eluent consisting of 100 mM NH_3 (pH=11.30) optimized in our previous work with the intact antibody IA sorbent was used (10 s at 50 mbar) [16]. However, results were not reproducible and after a few injections, TTR peak broadened, peak area decreased and the microcartridge needed to be substituted. Organic modifiers were also added to the basic eluent, for example 10% (v/v) of MeOH or ACN, but still results did not improve. The poor performance with these eluents was probably due to the limited stability of the disulfide bridge between Fab' fragments and the succinimidyl silica particles at alkaline pHs [36]. As an alternative, several acidic eluents typically used in immunoextraction were also evaluated, such as 100 mM glycine/50 mM HCl (pH=2.50) [11] and 100 mM phosphoric acid (pH=1.50) [17]. Unfortunately, the current was unstable and breakdown occurred at the beginning of the electrophoretic runs, probably because of protein precipitation or Fab' degradation. Finally, a solution consisting of 100 mM glycine/100 mM NH_3 (pH=9.50) was tested. This moderately basic eluent allowed the reproducible elution of TTR and multiple analyses on a single microcartridge.

Using the optimized elution conditions, sample loading pressure was studied at 50 and 930 mbar, and sample loading time at 10, 20, 30 and 45 min with a $10 \mu\text{g mL}^{-1}$ standard of TTR (Fig. 1). In addition, Fig. 1 also shows the amount of protein loaded at each pressure and time (in μL of sample and μg of protein). The extraction efficiency at 50 mbar was better because the amount of TTR loaded was much lower than at 930 mbar and the peak area did not proportionally decrease. However, at 930 mbar peak areas were larger in all cases. At this loading pressure, the largest TTR peak area was detected loading the sample for 20 min, but the difference with 10 min was very small. Furthermore, broader electrophoretic peaks were observed at loading times higher than 10 min, leading to current instability during the electrophoretic runs. For this reason, a sample loading time of 10 min at 930 mbar was finally selected for further analysis.

Fig. 2-A and B show the extracted ion electropherograms (EIEs) (i) and the mass spectra (ii and iii) of standard solutions of TTR by CE-MS ($50 \mu\text{g mL}^{-1}$) and Fab'-IA-SPE-CE-MS ($25 \mu\text{g mL}^{-1}$), respectively. In Fig. 2-B-i the EIE of a blank sample is also shown. As can be observed from comparison of the EIEs of TTR and the blank under the optimized elution conditions for Fab'-IA-SPE-CE-MS (Fig. 2-B-i), TTR eluted as

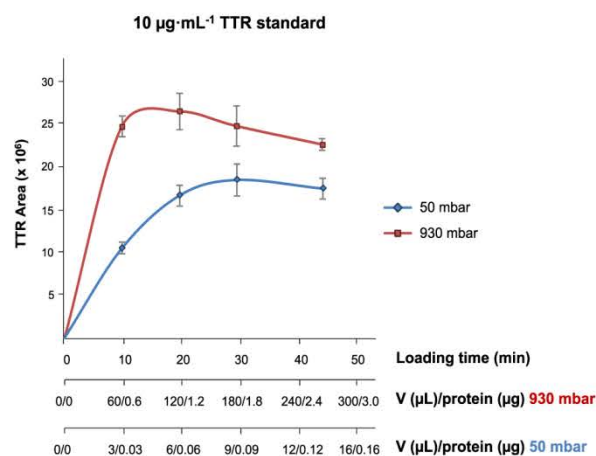


Fig. 1. Study of sample loading time (10, 20, 30 and 45 min) and pressure (50 mbar and 930 mbar) on the peak area of the eluted TTR (TTR standard of $10 \mu\text{g mL}^{-1}$). Peak area values were an average of three analyses ($n=3$).

two peaks with quite different shapes and some differences in their molecular mass spectra (Figs. 2-B-ii and B-iii). In our previous work with the intact antibody IA sorbent [16], we showed that peak shape could be improved if we ensured a rapid and quantitative protein elution pushing the elution plug with pressure until it passed the microcartridge before applying the voltage for the separation. This was not the case with the Fab'-IA sorbent and the results were similar following this strategy, probably because this characteristic elution/migration pattern was promoted by the larger active surface area and the better extraction capacity of the sorbent, as well as by the moderately basic eluent necessary for a reproducible elution. The two peaks were probably due to the presence of two different conformers of TTR, which has a natural tendency to misfold and aggregate [30,31]. Both peaks were not observed by CE-MS (Fig. 2-A-i), but were detected by Fab'-IA-SPE-CE-MS because they differently interacted with the IA sorbent and/or the eluent (Fig. 2-B-i). The first peak would correspond to a folded conformer that eluted faster from the Fab'-IA sorbent because of the smaller affinity. In contrast, the second peak would be due to an unfolded conformer and elution would be hindered because of the strong retention on the Fab'-IA sorbent. Several authors have shown that folded and unfolded protein conformers can be separated by CE, and that the affinity of these conformers to certain ligands can be different [37–40]. The findings on the molecular mass spectra of both peaks supported the presence of two molecular entities (Figs. 2-B-ii and B-iii). Small charge MO ions (+7, +8 and +9) were predominantly detected for the first conformer (folded) while high charge MO ions (e.g. +13, +14 and +15) appeared only for the second conformer (unfolded). In addition, a conscientious analysis of the cluster of multiply charged ions obtained in the m/z spectrum of the unfolded conformer showed the presence of a small amount of dimer (DI). As can be observed in Fig. 2-B-iii, ions with nominal m/z values of 1984, 2314 and 2777 were the summed contribution of the MO (+7, +6 and +5 ions, respectively) and DI (+14, +12 and +10 ions, respectively) because these DI even ions had the same m/z values as these MO ions. In contrast, 2136 and 2525 ions were exclusively detected for the DI (+13 and +11 odd ions, respectively). Notice that +16 DI ion was certainly not observed because +15 DI ion was not detected. It is well known that folded and unfolded proteins produce molecular mass spectra with different charge distributions in positive ESI. The folded conformers tend to have a narrow distribution with a smaller net charge, whereas the unfolded conformers produce a broad distribution centered on a much higher charge. These differences are believed to be related to the accessibility of the basic ionizable groups of the amino acid sequence, which would be easier to charge in the unfolded conformer [41].

Fig. 2-A and B (as insets) and Table 1 show the experimental M_r of the detected MO proteoforms obtained after deconvolution of the molecular mass spectra corresponding to TTR electrophoretic peaks (the contribution of the two conformers was summed together). As can be observed, three low abundant proteoforms (proteoforms $n=6-8$, Table 1) were additionally detected by Fab'-IA-SPE-CE-MS due to the better sensitivity. The detected TTR proteoforms corresponded to TTR showing a mixed disulfide with the amino acid cysteine at position 10 of the sequence (TTR-Cys, $n=1$), the free TTR (Free-TTR, $n=2$), the phosphorylated or the sulfonated TTR (TTR-Phosphorylated or TTR-Sulfonated, $n=3$), the dehydroxylated or the conjugated cysteine sulfinic acid TTR (TTR-Dehydroxylated or TTR-Sulfinic, $n=4$), the isoform presenting a single amino acid substitution of cysteine for glycine at position 10 ((10) C-G, $n=5$), and the proteoforms resulting when the cysteine residue on position 10 makes a mixed disulfide with the peptide cysteinyl-glycine (TTR-CysGly, $n=6$), the peptide glutathione (TTR-Glutathione, $n=7$) and the peptide cysteinyl-glutamic acid (TTR-CysGlu, $n=8$). As it is shown in Table 1, the calculated relative abundances (%A) for all the MO TTR proteoforms, which were calculated taking into account the peak areas of each EIE, were slightly higher with preconcentration, probably because recovery of TTR-Cys,

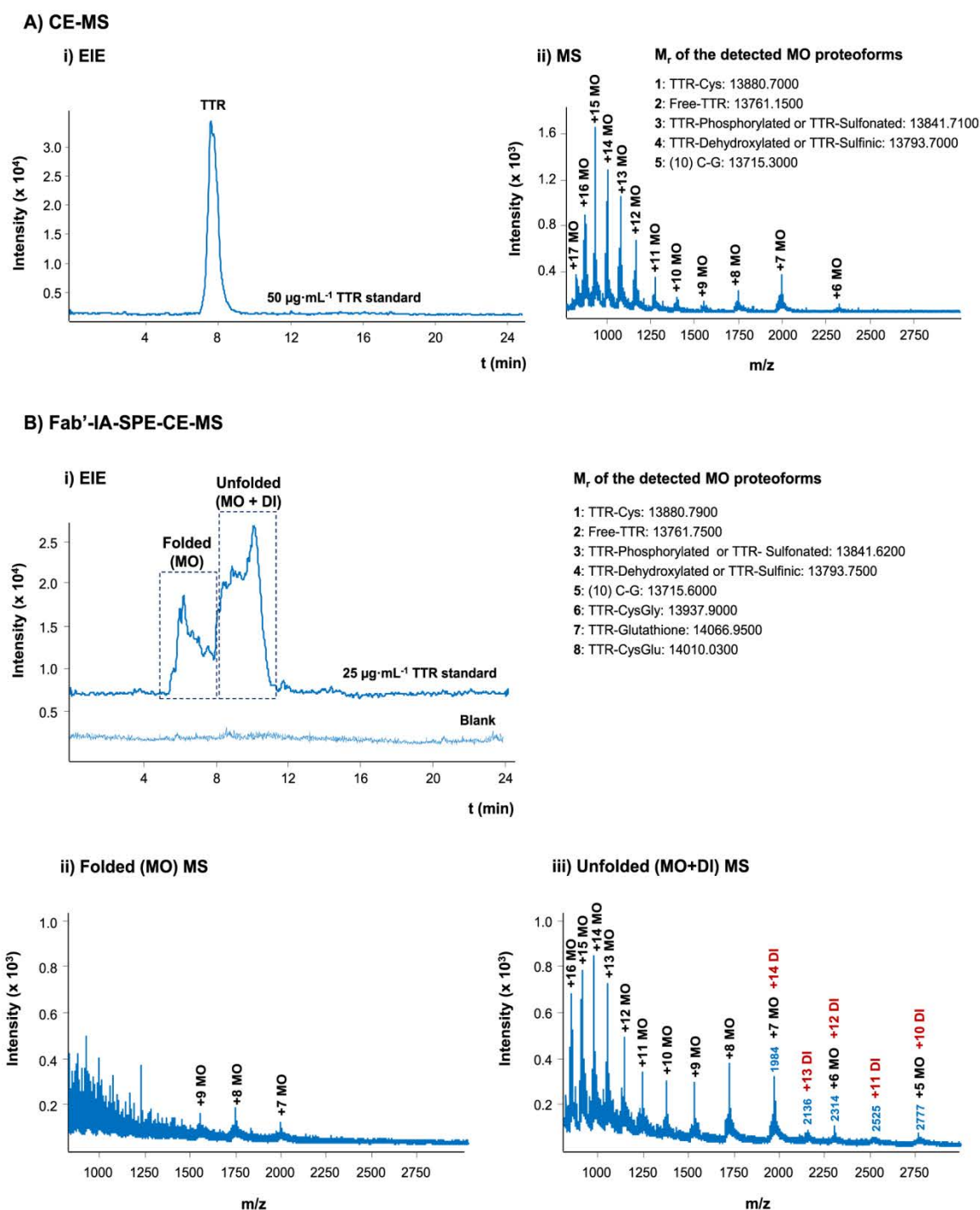


Fig. 2. (A) CE-MS for a 50 $\mu\text{g}\cdot\text{mL}^{-1}$ TTR standard and (B) Fab'-IA-SPE-CE-MS for a 25 $\mu\text{g}\cdot\text{mL}^{-1}$ TTR standard using 10 mM NH_4Ac (pH 7.00) as BGE. (MO=monomer, DI=dimer) (i: Extracted ion electropherograms (all proteoforms), ii and iii: TTR mass spectra).

which was taken as a reference, was lower than for the rest of proteoforms.

As summarized in Table 1-B, the optimized Fab'-IA-SPE-CE-MS method was repeatable ($n=6$) in terms of t_m and peak areas for all the proteoforms (ranging between 7–13% and 8–13% for t_m and peak areas, respectively). Similarly, the reproducibility ($n=9$, three replicates and three different microcartridges) was also acceptable and ranged between 6–12% and 9–12% for t_m and peak areas, respectively.

Moreover, the method was linear ($r^2 > 0.99$) between 1 and 25 $\mu\text{g}\cdot\text{mL}^{-1}$ and LODs were around 0.5 $\mu\text{g}\cdot\text{mL}^{-1}$, which meant an improvement of 50 and 2-fold with regard to the LODs obtained using CE-MS ($\sim 25\text{ }\mu\text{g}\cdot\text{mL}^{-1}$) and the intact antibody IA sorbent method developed in our previous work ($\sim 1\text{ }\mu\text{g}\cdot\text{mL}^{-1}$) [16]. The lifetime of the microcartridges was superior to 20 analyses without problems of extraction performance, current instability or breakage (similar to the 20 runs achievable for the analysis of peptides and small molecules

Table 1

Theoretical and deconvoluted average M_r , relative abundance, repeatability and reproducibility for the detected MO proteoforms of a 50 $\mu\text{g mL}^{-1}$ TTR standard by CE-MS and 25 $\mu\text{g mL}^{-1}$ TTR standard by Fab'-IA-SPE-CE-MS. (BGE: 10 mM NH_4Ac , pH 7.00).

N	Detected MO TTR proteoforms	Theoretical Average M _r	A) CE-MS				B) Fab'-IA-SPE-CE-MS							
			Deconvoluted Average M _r (n=6)		%A ^b (n=6)		Deconvoluted Average M _r (n=6)		%A ^b (n=6)		Repeatability (% RSD) (25 µg mL ⁻¹ , n=6)		Reproducibility (% RSD) (25 µg mL ⁻¹ , n=9)	
			Experimental	E _r ^a (ppm)			Experimental	E _r ^a (ppm)			Peak area	Migration time	Peak area	Migration time
1	TTR-Cys	13,880.4022	13,880.7000	21	100	13,880.7900	28	100	13	7	12	6		
2	Free-TTR	13,761.2640	13,761.1500	8	40	13,761.7500	35	44	10	9	10	8		
3	TTR-Phosphorylated ^c	13,841.2439	13,841.7100	34	39	13,841.6200	27	63	10	9	11	8		
4	TTR-Sulfonated ^c	13,841.3283		28			21							
	TTR-Dehydroxylated/ TTR-Sulfinic ^c	13,793.2628	13,793.7000	32	39	13,793.7500	35	45	9	11	11	10		
5	(10) C-G	13,715.1713	13,715.3000	9	33	13,715.6000	31	41	10	9	9	8		
6	TTR-CysGly	13,937.4590	(not detected) n. d.	–	–	13,937.9000	32	59	12	12	11	11		
7	TTR-Glutathione	14,066.5732	n. d.	–	–	14,066.9500	27	42	8	13	9	12		
8	TTR-CysGlu	14,009.5218	n. d.	–	–	14,010.0300	36	36	8	10	9	9		

^a Relative error (E_r) was calculated in ppm as: $|M_r \text{ exp} - M_r \text{ theo}| / M_r \text{ theo} \times 10^6$ (exp=experimental and theo=theoretical).

^b The relative abundance (%A) was calculated normalizing to the area value of the most abundant form (the contribution of both conformers was summed together to get a unique integration for each MO TTR proteoform).

^c The mass accuracy and resolution of the mass spectrometer were not enough to differentiate between proteoforms with very close M_r , which were neither confirmed by MS/MS.

with C_{18} microcartridges in C_{18} -SPE-CE-MS [20,21]). Furthermore, in contrast to the intact antibody microcartridges prepared in our previous work [16], which had to be removed every day because they lost their immunoextraction efficiency, the present Fab' microcartridges could be used several consecutive days without loss of performance if filled overnight with water or BGE.

3.2. Analysis of TTR serum samples by Fab'-IA-SPE-CE-MS

For the purpose of evaluating the selectivity of the Fab' sorbent and the potential of the established method for the analysis of TTR in biological fluids, serum samples from healthy controls and FAP-I patients (symptomatic patient) were analyzed by Fab'-IA-SPE-CE-MS. Loading of serum samples without any pretreatment (or just diluting) was not possible because of current instability and breakage. As in our previous work with the intact antibody IA sorbent [16], a simple off-line pretreatment based on a double precipitation with 5% (v/v) of phenol was used to prevent saturation of the microcartridges and capillary inner surface damage due to the loading of salts and other high-abundance proteins, such as albumin. In addition, increasing the ionic strength of the eluent up to 200 mM glycine/200 mM NH_3 (pH=9.50) was necessary in order to reproducibly elute the TTR from the pretreated serum samples. Under these conditions, repeatability in terms of t_m and peak areas for all the proteoforms was similar to the values obtained with standards (between 12–14% and 15–19% RSD, respectively). However, the lifetime of the Fab' microcartridges was slightly lower than the 20 runs achievable with the TTR standards (around 15 analyses), probably because of the faster saturation of the microcartridges due to non-specific interactions with matrix components of the pretreated biological samples. In this regard, an end-capping procedure with glycidol was tested in order to prevent non-specific binding with the unreacted aminopropyl groups. However, results did not improve and the end-capping was discarded after some preliminary experiments. Without end-capping, the durability of the Fab' microcartridges was higher than the durability of the intact antibody microcartridges developed in our previous work for the analysis of TTR in serum samples (~15 vs. ~10 analyses, respectively) [16].

Fig. 3-A shows the EIE (i) and the mass spectra (ii and iii) of TTR

for a serum sample from an FAP-I symptomatic patient. As can be observed (Fig. 3-A-i), TTR eluted as three peaks that would correspond to different conformers, but differences in their molecular mass spectra were only evident between the first and the last two peaks (Figs. 3-A-ii and A-iii). The presence in the mass spectra of a cluster of multiply charged ions with a broad distribution centered on a high charge, as observed before for the TTR standard (Fig. 2-B-iii), confirmed the presence of only unfolded TTR conformers, each interacting with the Fab'-IA sorbent in a different manner. The absence of a folded conformer and migration time shifts with regard to the TTR standard solution (Fig. 2-B) were probably due to the matrix of the serum sample, because we discarded the influence of the reoptimized eluent composition by separated experiments with this eluent and the standard. Since protein unfolding studies usually adopt a two-state model system, only two conformations are ideally addressed (folded and unfolded states). However, it is widely accepted that once unfolded, the protein can reversibly aggregate, thus leading to new unfolded conformational changes and oligomerization [37,39]. More specifically, the first peak corresponded to an unfolded conformer presenting MO and DI ions with a similar migration time to the one detected in the TTR standard (Fig. 2-B-i). As can be observed in Fig. 3-A-ii, ions with nominal m/z values of 2314 and 2777 would be the summed contribution of the MO and DI, while the 2525 ion would be only detected for the DI (+11 ion). In contrast, the last two peaks would correspond to a further unfolded conformer presenting MO, DI and trimer (TRI) ions. The ion with nominal m/z value of 2314 would be the summed contribution of the MO, DI and TRI, while 2083, 2193 and 2450 ions would be only detected for the TRI (+20, +19 and +17 ions, respectively). These differences observed now in the mass spectra with regard to the TTR standard solution (Fig. 2-B) were probably due again to the sample matrix. Furthermore, some differences were evident between the electrophoretic profiles obtained for the healthy control (Fig. 3-B-i) and the FAP-I patient (Fig. 3-A-i). In the case of the healthy control, the abundance of the third conformer was much smaller than for the FAP-I patient. Several other control and FAP-I patient samples were analyzed with similar results, suggesting that these particular electrophoretic profiles were related to the differences between TTR proteoforms in control and patient samples and the higher ability to misfold and oligomerize of TTR in FAP-I patients. Tables 2-A and -B

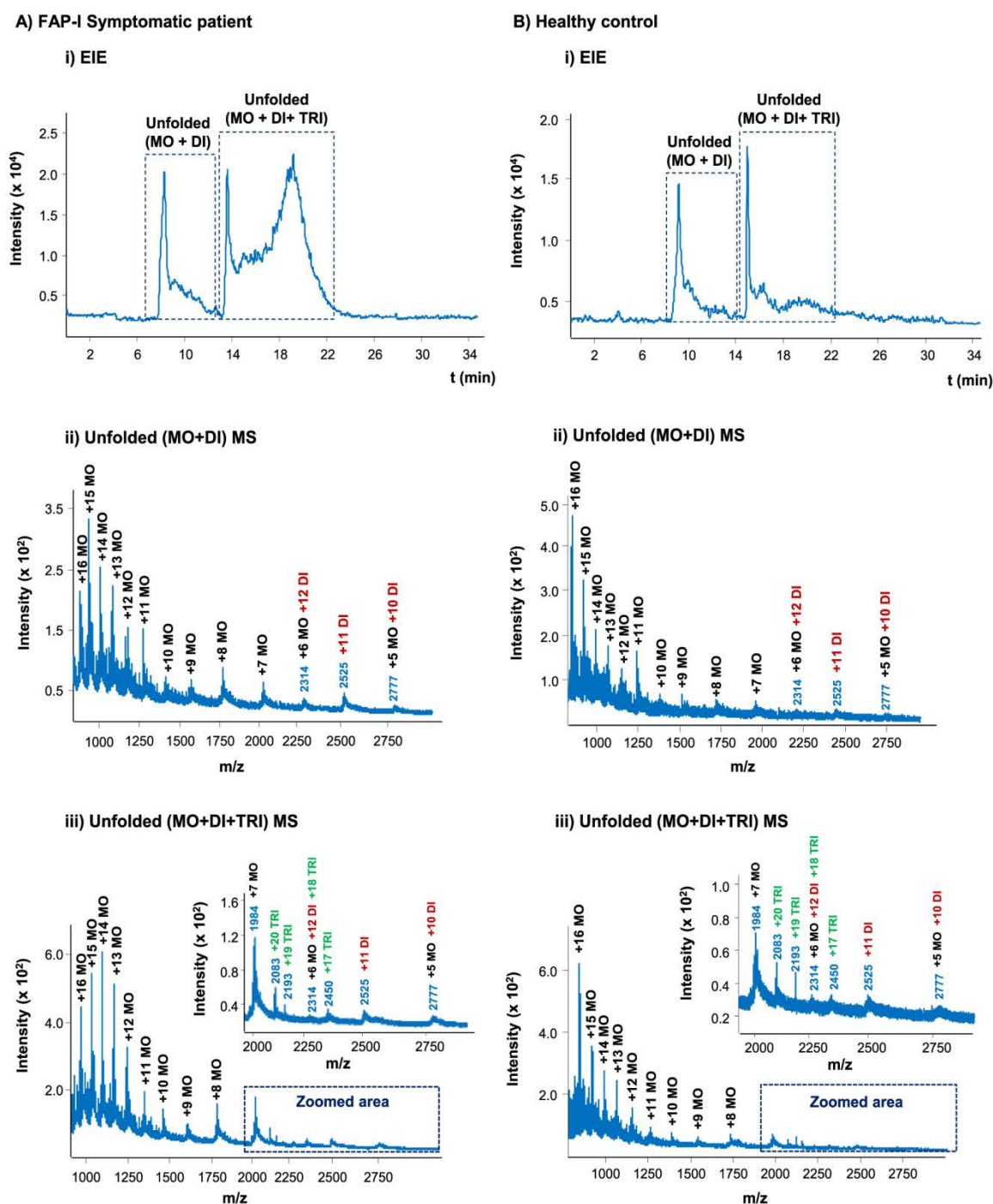


Fig. 3. Fab-IA-SPE-CE-MS for a serum sample from a symptomatic FAP-I patient (A) and a healthy control (B) after sample pretreatment with 5% (v/v) of phenol. (MO=monomer, DI=dimer, TRI=trimer) (i: Extracted ion electropherograms (all proteoforms), ii and iii: TTR mass spectra). The detected MO proteoforms are summarized in Table 2.

show the theoretical and deconvoluted average M_r of the detected MO proteoforms in serum samples from the healthy control and the symptomatic FAP-I patient, respectively. Again, the contribution of all the conformers was summed together to get a unique integration for each TTR proteoform. In contrast to the method developed in our previous work with the intact antibody [16], where only five of the most abundant normal TTR proteoforms were detected in serum samples (TTR-Cys ($n=1$), Free-TTR ($n=2$), TTR-Phosphorylated or TTR-Sulfonated ($n=3$), TTR-Dehydroxylated or TTR-Sulfinic ($n=4$) and

(10) C-G ($n=5$), Table 2), the present method allowed the detection of three extra normal proteoforms (TTR-CysGly ($n=6$), TTR-Glutathione ($n=7$) and TTR-CysGlu ($n=8$)), with similar figures of merit (Table 2) compared to the TTR standard (Table 1-B). In addition, the main mutant TTR proteoform (TTR-Cys(Met30) ($n=1'$)) was detected for the symptomatic FAP-I patient, allowing diagnosis of FAP-I. However, the mass spectrometer mass accuracy and resolution were not enough to differentiate between the minor mutant proteoform Free-TTR(Met30) ($n=2'$, Table 2) and TTR-Dehydroxylated or TTR-

Table 2

Theoretical and deconvoluted average M_r and relative abundance for the detected MO TTR proteoforms in serum samples pretreated with 5% (v/v) phenol solution (dilution 1:1 (v/v)) by Fab'-IA-SPE-CE-MS. (BGE: 10 mM NH_4Ac , pH 7.00).

N	Detected MO TTR proteoforms	Theoretical average M_r	A) Healthy control			A) FAP-I patient		
			Deconvoluted Average M_r (n=6)		%A ^b (n=6)	Deconvoluted Average M_r (n=6)		%A ^b (n=6)
			Experimental	E_r^a (ppm)		Experimental	E_r^a (ppm)	
1	TTR-Cys	13,880.4022	13,880.6600	19	100	13,880.8800	34	100
1'	TTR(Met30)-Cys	13,912.4683	(not detected) n.d.	–	–	13,912.6500	13	74
2	Free-TTR	13,761.2640	13,760.8300	32	36	13,760.9300	24	70
3	TTR-Phosphorylated ^c	13,841.2439	13,841.8000	40	42	13,841.3500	8	74
	TTR-Sulfonated ^c	13,841.3283		34			2	
4	TTR-Dehydroxylated/ TTR-Sulfenic ^c	13,793.2628	13,794.0100	54	32	13,793.7500	35	48
2'	Free-TTR(Met30) ^c	13,793.3301	n.d.	–	–		30	
5	(10) C-G	13,715.1713	13,714.7800	29	37	13,715.2000	2	50
6	TTR-CysGly	13,937.4590	13,936.9500	37	52	13,937.6000	10	65
7	TTR-Glutathione	14,066.5732	14,067.1000	37	40	14,067.0500	34	56
8	TTR-CysGlu	14,009.5218	14,010.0500	38	35	14,010.0100	35	49

^a Relative error (E_r) was calculated in ppm as: $|M_r \text{ exp} - M_r \text{ theo}|/M_r \text{ theo} \times 10^6$ (exp=experimental and theo=theoretical).

^b The relative abundance (%A) was calculated normalizing to the area value of the most abundant form (the contribution of both conformers was summed together to get a unique integration for each MO TTR proteoform). Repeatability (n=6) of t_m and peak areas for all the proteoforms ranged between 12–14% and 15–19%, respectively.

^c The mass accuracy and resolution of the mass spectrometer were not enough to differentiate between proteoforms with very close M_r , which were neither confirmed by MS/MS.

Sulfenic (n=4, Table 2). Therefore, although the %A corresponding to these three proteoforms in the FAP-I patient was higher than the summed contribution of the two normal proteoforms in the healthy control (48% vs. 32%, Table 2), only detection of the TTR(Met30)-Cys proteoform would unambiguously confirm the TTR amyloidosis. The applicability of the method for diagnosis of FAP-I is evident, but the analysis of a larger set of samples (using a mass spectrometer with a higher mass accuracy and resolution) would be necessary to determine whether the information related to the detection by Fab'-IA-SPE-CE-MS of different proteoform abundance profiles and unfolded TTR conformers in patients is useful to gain a novel insight into the onset and prognosis of FAP-I.

4. Concluding remarks

We have developed a method for purification, separation and characterization of TTR from serum samples by IA-SPE-CE-MS using an IA sorbent with Fab' antibody fragments. Due to the limited stability of the disulfide bridge between Fab' fragments and the succinimide silica particles at alkaline pHs, elution was only possible at certain conditions where TTR was detected as different conformers. Under the optimized conditions with TTR standard solutions, migration times and peak areas were repeatable and reproducible, microcartridges lifetime was good (> 20 analyses in consecutive days) and LODs were 50 and 2-fold lower compared to the LODs obtained using CE-MS (~25 $\mu\text{g mL}^{-1}$) and the intact antibody IA sorbent method developed in our previous work (~1 $\mu\text{g mL}^{-1}$) [16]. The reliability and selectivity of the proposed methodology was evaluated analyzing FAP-I serum samples after applying a simple off-line pretreatment based on protein precipitation with 5% (v/v) of phenol. The current Fab'-IA-SPE-CE-MS method enhanced our previous method for the analysis of TTR based on intact antibodies and allowed the detection of proteoforms found at lower concentrations and the preparation of microcartridges with extended durability. Furthermore, TTR from FAP-I patients and control sera showed a differential conformer profile that may be related to differences in TTR proteoforms. The potential of the method for diagnosis of FAP-I is demonstrated, but further experiments would be necessary using mass spectrometers with improved capabilities (e.g. mass accuracy and resolution, sensitivity and MS/MS mode) to show if it is useful to gain a novel insight into FAP-I. Furthermore, results with TTR could be regarded as a starting point for the analysis of other large biomolecules (e.g. proteins and glycoproteins) by Fab'-IA-SPE-CE-MS.

Acknowledgements

Laura Pont acknowledges the Spanish Ministry of Economy and Competitiveness (BES-2012-061127) for a FPI fellowship. This study was supported by a grant from the Spanish Ministry of Education and Science (CTQ2014-56777-R). We also thank Dr. C. Casasnovas and Dr. M.A. Alberti (Hospital Universitari de Bellvitge, HUB, Hospitalet de Llobregat, Spain) for providing the blood samples.

Appendix A. Supporting information

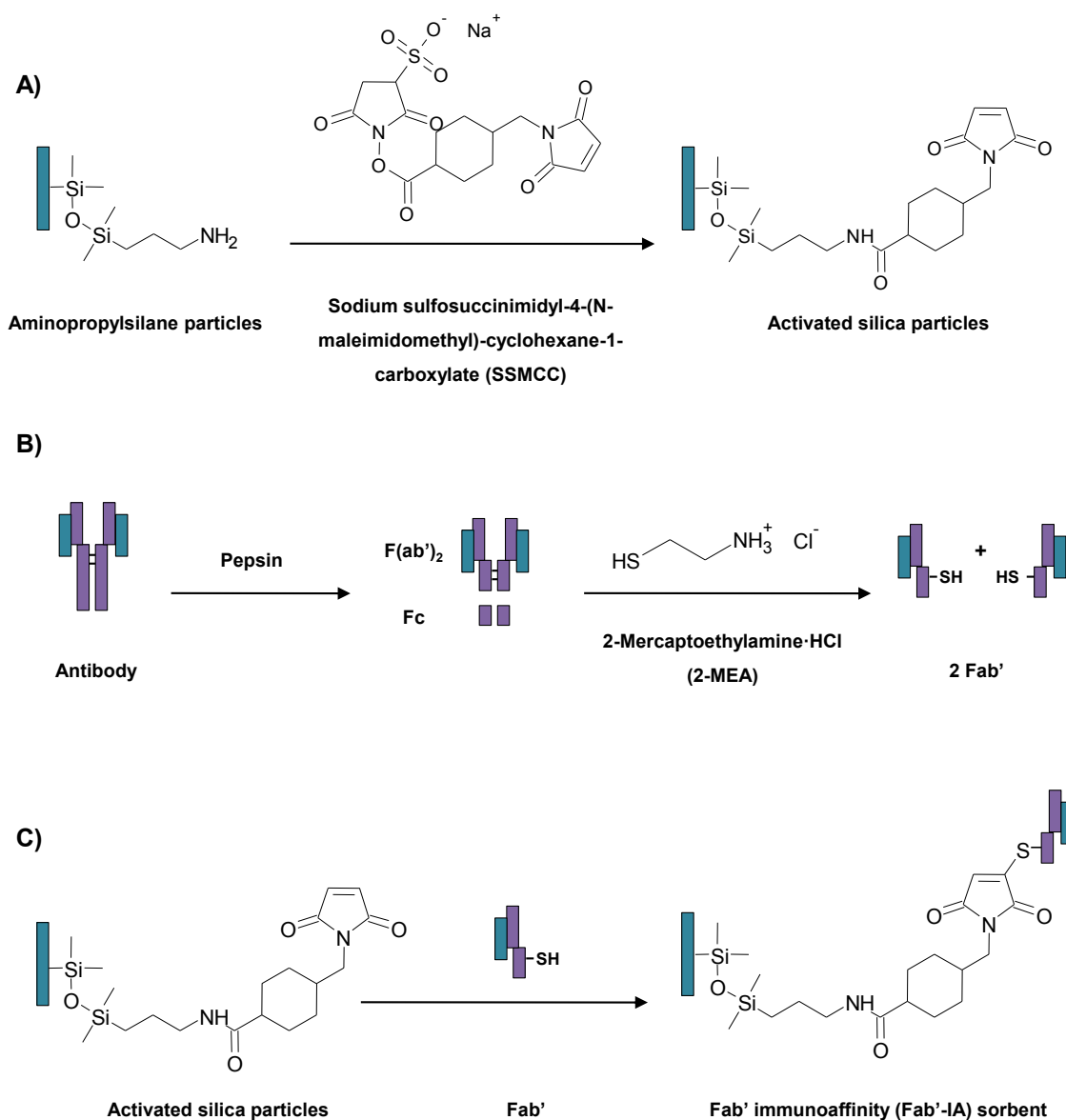
Supplementary data associated with this article can be found in the online version at doi:10.1016/j.talanta.2017.03.104.

References

- [1] N.L. Anderson, The clinical plasma proteome: a survey of clinical assays for proteins in plasma and serum, *Clin. Chem.* 56 (2010) 177–185.
- [2] E.S. Boja, T.E. Fehniger, M.S. Baker, G. Marko-Varga, H. Rodriguez, Analytical validation considerations of multiplex mass-spectrometry-based proteomic platforms for measuring protein biomarkers, *J. Proteome Res.* 13 (2014) 5325–5332.
- [3] T.K. Toby, L. Fornelli, N.L. Kelleher, Progress in top-down proteomics and the analysis of proteoforms, *Annu. Rev. Anal. Chem.* 9 (2016) 499–519.
- [4] A. Shimizu, T. Nakanishi, A. Miyazaki, Detection and characterization of variant and modified structures of proteins in blood and tissues by mass spectrometry, *Mass Spectrom. Rev.* 25 (2006) 686–712.
- [5] T.S. Collier, D.C. Muddiman, Analytical strategies for the global quantification of intact proteins, *Amino Acids* 43 (2012) 1109–1117.
- [6] P.J. Tighe, R.R. Ryder, I. Todd, L.C. Fairclough, ELISA in the multiplex era: potentials and pitfalls, *Proteom. Clin. Appl.* 9 (2015) 406–422.
- [7] N.A. Guzman, D.E. Guzman, An emerging micro-scale immuno-analytical diagnostic tool to see the unseen. Holding promise for precision medicine and P4 medicine, *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* 1021 (2016) 14–29.
- [8] R. Aebersold, M. Mann, Mass spectrometry-based proteomics, *Nature* 422 (2003).
- [9] H. Zhou, Z. Ning, A.E. Starr, M. Abu-Farha, D. Figeys, Advancements in top-down proteomics, *Anal. Chem.* 84 (2012) 720–734.
- [10] R. Haselberg, G.J. de Jong, G.W. Somsen, CE-MS for the analysis of intact proteins 2010–2012, *Electrophoresis* 34 (2013) 99–112.
- [11] N.A. Guzman, Determination of immunoreactive gonadotropin-releasing hormone in serum and urine by on-line immunoaffinity capillary electrophoresis coupled to mass spectrometry, *J. Chromatogr. B Biomed. Sci. Appl.* 749 (2000) 197–213.
- [12] N. a. Guzman, T.M. Phillips, Immunoaffinity capillary electrophoresis: a new versatile tool for determining protein biomarkers in inflammatory processes, *Electrophoresis* 32 (2011) 1565–1578.
- [13] 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* 717 (2012) 134–142.
- [14] S. Medina-Casanellas, F. Benavente, J. Barbosa, V. Sanz-Nebot, Preparation and evaluation of an immunoaffinity sorbent with Fab' antibody fragments for the

- analysis of opioid peptides by on-line immunoaffinity solid-phase extraction capillary electrophoresis-mass spectrometry, *Anal. Chim. Acta* 789 (2013) 91–99.
- [15] S. Medina-Casanellas, F. Benavente, E. Giménez, J. Barbosa, V. Sanz-Nebot, On-line immunoaffinity solid-phase extraction capillary electrophoresis mass spectrometry for the analysis of large biomolecules: a preliminary report, *Electrophoresis* 35 (2014) 2130–2136.
 - [16] R. Peró-Gascón, L. Pont, F. Benavente, J. Barbosa, V. Sanz-Nebot, Analysis of serum transthyretin by on-line immunoaffinity solid-phase extraction capillary electrophoresis mass spectrometry using magnetic beads, *Electrophoresis* 37 (2016) 1220–1231.
 - [17] T.M. Phillips, E. Wellner, Measurement of neuropeptides in clinical samples using chip-based immunoaffinity capillary electrophoresis, *J. Chromatogr. A* 1111 (2006) 106–111.
 - [18] H. Kalish, T.M. Phillips, Analysis of neurotrophins in human serum by immunoaffinity capillary electrophoresis (ICE) following traumatic head injury, *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* 878 (2010) 194–200.
 - [19] H. Kalish, T.M. Phillips, Assessment of chemokine profiles in human skin biopsies by an immunoaffinity capillary electrophoresis chip, *Methods* 56 (2012) 198–203.
 - [20] E. Hernández, F. Benavente, V. Sanz-Nebot, J. Barbosa, Evaluation of on-line solid phase extraction-capillary electrophoresis-electrospray-mass spectrometry for the analysis of neuropeptides in human plasma, *Electrophoresis* 29 (2008) 3366–3376.
 - [21] F. Benavente, S. Medina-Casanellas, J. Barbosa, V. Sanz-Nebot, Investigation of commercial sorbents for the analysis of opioid peptides in human plasma by on-line SPE-CE, *J. Sep. Sci.* 33 (2010) 1294–1304.
 - [22] T.M. Phillips, B.F. Dickens, Affinity and immunoaffinity purification, *Techniques* (2000) 1–18.
 - [23] M. Nisnevitch, M.A. Firer, The solid phase in affinity chromatography: strategies for antibody attachment, *J. Biochem. Biophys. Methods* 49 (2001) 467–480.
 - [24] R. Mallik, C. Wa, D.S. Hage, Development of sulfhydryl-reactive silica for protein immobilization in high-performance affinity chromatography, *Anal. Chem.* 79 (2007) 1411–1424.
 - [25] T.M. Phillips, E. Wellner, Detection of cerebral spinal fluid-associated chemokines in birth traumatized premature babies by chip-based immunoaffinity CE, *Electrophoresis* 34 (2013) 1530–1538.
 - [26] V. Crivianu-Gaita, M. Thompson, Immobilization of Fab' fragments onto substrate surfaces: a survey of methods and applications, *Biosens. Bioelectron.* 70 (2015) 167–180.
 - [27] H. Terazaki, Y. Ando, O. Suhr, P.I. Ohlsson, K. Obayashi, T. Yamashita, S. Yoshimatsu, M. Suga, M. Uchino, M. Ando, Post-translational modification of transthyretin in plasma, *Biochem. Biophys. Res. Commun.* 249 (1998) 26–30.
 - [28] K. Poulsen, J.M.C. Bahl, J.T. Tanassi, A.H. Simonsen, N.H.H. Heegaard, Characterization and stability of transthyretin isoforms in cerebrospinal fluid examined by immunoprecipitation and high-resolution mass spectrometry of intact protein, *Methods* 56 (2012) 284–292.
 - [29] 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.
 - [30] R.H. Falk, R.L. Comenzo, M. Skinner, The systemic amyloidoses, *N. Engl. J. Med.* 337 (1997) 898–909.
 - [31] S.M. Johnson, S. Connelly, C. Fearn, E.T. Powers, J.W. Kelly, The transthyretin amyloidoses: from delineating the molecular mechanism of aggregation linked to pathology to a regulatory-agency-approved drug, *J. Mol. Biol.* 421 (2012) 185–203.
 - [32] Y. Ando, P. Ohlsson, O. Suhr, N. Nyhlin, T. Yamashita, G. Holmgren, A. Danielsson, O. Sandgren, M. Uchino, M. Ando, A new simple and rapid screening method for variant transthyretin-related amyloidosis, *Biochem. Biophys. Res. Commun.* 228 (1996) 480–483.
 - [33] Y. Ando, O. Suhr, T. Yamashita, P.I. Ohlsson, G. Holmgren, K. Obayashi, H. Terazaki, C. Mambule, M. Uchino, M. Ando, Detection of different forms of variant transthyretin (Met30) in cerebrospinal fluid, *Neurosci. Lett.* 238 (1997) 123–126.
 - [34] R.J. Markovich, X.X. Qiu, D.E. Nichols, C. Pidgeon, B. Invergo, F.M. Alvarez, Silica subsurface amine effect on the chemical stability and chromatographic properties of end-capped immobilized artificial membrane surfaces, *Anal. Chem.* 63 (1991) 1851–1860.
 - [35] D. Heiger, High performance capillary electrophoresis: a primer, *Agil Technol.* (2010) 84–85.
 - [36] P.J. Hogg, Disulfide bonds as switches for protein function, *Trends Biochem. Sci.* 28 (2003) 210–214.
 - [37] C.S. Jørgensen, L.R. Ryder, A. Steino, P. Højrup, J. Hansen, N.H. Beyer, N.H.H. Heegaard, G. Houen, Dimerization and oligomerization of the chaperone calreticulin, *Eur. J. Biochem.* 270 (2003) 4140–4148.
 - [38] N.H.H. Heegaard, T.J.D. Jørgensen, N. Rozlosnik, D.B. Corlin, J.S. Pedersen, A.G. Tempesta, P. Roepstorff, R. Bauer, M.H. Nissen, Unfolding, aggregation, and seeded amyloid formation of lysine-58-cleaved β 2-microglobulin, *Biochemistry* 44 (2005) 4397–4407.
 - [39] J.M. a. Gavina, P. Britz-McKibbin, Protein unfolding and conformational studies by capillary electrophoresis, *Curr. Anal. Chem.* 3 (2007) 17–31.
 - [40] G.G. Mironov, C.M. Clouthier, A. Akbar, J.W. Keillor, M.V. Berezovski, Simultaneous analysis of enzyme structure and activity by kinetic capillary electrophoresis-MS, *Nat. Chem. Biol.* 12 (2016) 918–922.
 - [41] I.N. Serdyuk, N.R. Zaccai, J. Zaccai, *Methods in molecular, Biophysics* (2007) 1–1120.

Supplementary Figures



Supplementary Figure S-1. Preparation procedure of the Fab'-IA sorbent: (A) derivatization of aminopropyl silica particles with SSMCC, (B) generation of F(ab')₂ fragments from IgG by digestion with pepsin followed by reduction with 2-MEA to obtain Fab' fragments and (C) immobilization of Fab' to the activated silica particles.

Chapter 3. Targeted analysis of protein biomarkers. Bottom-up proteomics

Biomarker discovery in cancer is a scientific challenge that is currently requiring a great deal of effort. Cancer is a leading cause of death, being colorectal cancer (CRC) the second most commonly diagnosed cancer among females and third among males worldwide. Furthermore, in patients with CRC, the liver is the most commonly involved site of distant metastasis, with over half of all patients with CRC developing liver metastatic disease. Recent research in the glycomics field has given insight into the biological significance of the *N*-glycome in different diseases, particularly in cancer. More than half of the cancer biomarkers are glycosylated proteins, and specific glycoforms may serve as biomarkers for the early diagnosis, prognosis or the evaluation of therapeutic efficacy for treatment diseases. In this regard, human carcinoembryonic antigen (CEA) is an interesting model to study the modifications of *N*-glycosylation patterns induced during cancer biogenesis. CEA is a highly *N*-glycosylated protein (60% m/m of glycans) found in normal human colonic epithelial cells, as well as in tumor forming and colonic adenocarcinogenic cell lines. CEA has been identified in several human cancer tissues and it still remains the only tumor marker of recognized efficacy in monitoring CRC and liver metastatic disease patients' therapy. However, further investigation has revealed that the quantitation of the serum levels of this glycoprotein is not sufficiently specific for CRC diagnosis or prognosis.

Despite the usefulness of top-down strategies for the analysis of intact proteins, the technical difficulty of proteome analysis at the intact level has caused top-down proteomics to lag behind bottom-up proteomics, in which protein samples are digested into peptides prior to mass spectrometry (MS) analysis. Targeted bottom-up proteomics provides, therefore, an indirect measurement of a certain protein through peptides derived from proteolytic digestion.

In this chapter, we describe a multienzyme sheathless capillary electrophoresis tandem mass spectrometry (CE-MS/MS) **bottom-up** strategy for the *N*-glycosylation analysis of CEA samples purified from human colon carcinoma and human liver metastases. The obtained results permit us to propose novel potential *N*-glycopeptide markers, which could provide important

biologic information on how *N*-glycosylation may influence CEA structure in CRC and liver metastatic disease.

This chapter includes the following publication:

- **Publication 3.1.** *N*-glycosylation analysis of human carcinoembryonic antigen using glycopeptide maps obtained by sheathless capillary electrophoresis tandem mass spectrometry. L. Pont, G. Kammeijer, F. Benavente, V. Sanz-Nebot, O. Mayboroda, M. Wuhrer. Analytical Chemistry 2017 (Submitted for publication).

***N*-glycosylation analysis of human carcinoembryonic antigen using glycopeptide maps obtained by sheathless capillary electrophoresis tandem mass spectrometry**

Laura Pont^{a*}, Guinevere S.M. Kammeijer^{b*}, Fernando Benavente^{a#}, Victoria Sanz-Nebot^a, Oleg A.

Mayboroda^b, Manfred Wuhrer^b

^aDepartment of Chemical Engineering and Analytical Chemistry, University of Barcelona, Barcelona, Spain

^bCenter for Proteomics and Metabolomics, Leiden University Medical Center, Leiden, The Netherlands

*Shared first author

#Corresponding author:

F. Benavente, PhD, fbenavente@ub.edu, Tel: (+34) 934039778, Fax: (+34) 934021233

Abstract

Human carcinoembryonic antigen (CEA) is a highly *N*-glycosylated protein (approx. 60% m/m glycans, with 28 potential *N*-linked glycosylation sites) found in normal human colonic epithelial cells, as well as in tumor forming and colon adenocarcinogenic cell lines. CEA displays a considerable heterogeneity in the sugar content, but detailed carbohydrate structures for the different *N*-glycosylation sites have so far not been reported. In the present study, CEA samples purified from human colon carcinoma and human liver metastases were digested and the enzymatic hydrolysates subsequently analyzed by sheathless capillary electrophoresis tandem mass spectrometry (sheathless CE-MS/MS). The information obtained using specific enzymes, such as trypsin and endoproteinase Glu-C, and a non-specific enzyme, namely pronase, allowed the characterization of 20 out of the 28 potential *N*-glycosylation sites with respect to their occupancy. In total, 219 different *N*-glycopeptide glycoforms were identified. The relative abundances of the CEA glycopeptides, obtained after integration of the extracted ion electropherograms, were investigated following a targeted multivariate data analysis approach. Partial least squares discriminant analysis (PLS-DA) was applied to easily differentiate the complex and characteristic glycopeptide maps of CEA samples purified from human colon carcinoma and human liver metastases. A hundred and twenty one *N*-glycopeptide glycoforms (out of the total of 219) were found to be the most significant to discriminate between the two CEA types. Overall, it seems that CEA samples purified from human liver metastases have more fucosylation and sialylation.

1. Introduction

Protein glycosylation is by far the most common and complex post-translational modification (PTM) with more than half of all secretory and cellular human proteins being glycosylated. Glycoproteins play a crucial role in recognition, signaling and adhesion processes on the cell surfaces. For instance, *N*-glycosylation patterns govern and regulate critical immune responses and play a major role in recognition of tumors and anti-tumor responses¹⁻³. An assembly error such as a misplacement of a single monosaccharide on a core glycan or branches may change the function of a glycoprotein and, in turn, the phenotype of a cell^{4,5}. Furthermore, dysregulation of the expression (over- or under expression) of naturally occurring *N*-linked glycans have been associated with cancer metastases⁶.

Human carcinoembryonic antigen (CEA) is an interesting model to study the modifications of *N*-glycosylation patterns induced during cancer biogenesis. CEA is a highly *N*-glycosylated protein (60% m/m glycans) with a relative molecular mass (M_r) of approximately 180,000 that can be found in normal human colonic epithelial cells, as well as in tumor forming and colonic adenocarcinogenic cell lines^{7,8}. It has

been identified in human cancer tissues, and its presence is known to be related to the progression of several tumors^{9,10}. However, further investigations revealed that the quantitation of the serum levels of this glycoprotein is not sufficiently specific for tumor diagnosis or prognosis⁹. Therefore, improvement in the specificity of CEA detection remains a challenge in cancer research. CEA contains 28 potential *N*-linked glycosylation sites^{11,12}, and the *N*-glycosylation of CEA is known to display a considerable heterogeneity¹², but detailed carbohydrate structures of these *N*-glycosylation sites have so far not been reported. This is probably due to the diverse macro- and microheterogeneity of the glycoprotein, hence providing a wide variety in concentrations and a large number of different glycoforms that complicate the analysis.

Different techniques are currently used for the analysis of glycoproteins, including matrix assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF-MS)¹³, or electrospray ionization mass spectrometry (ESI-MS)^{14,15}, coupled or not, with reversed phase high-performance liquid chromatography (HPLC)¹⁴⁻¹⁷. Capillary electrophoresis (CE) is another miniaturized high-performance

separation technique with the applicability to analyze glycoproteins in top-down¹⁸ and bottom-up^{19,20} approaches. While sheath liquid interfaces are more commonly used, as they are readily available and have historically provided high robustness²¹, sheathless CE-MS with a novel interface has recently been reported to feature significantly improved limits of detection (LODs) and increased analyte coverage when compared to sheath liquid CE-MS and LC-MS^{22,23}. This sheathless CE-MS interface, which follows the design by Moini²⁴, has recently been commercialized. In this interface, the nanospray emitter is the porous end of the 30 μm ID separation capillary, which is inserted into a grounded needle filled with a conductive liquid, providing the necessary electrical contact through the pores.

In this paper, we used sheathless capillary electrophoresis tandem mass spectrometry (sheathless CE-MS/MS) for the analysis of CEA *N*-glycopeptides obtained after enzymatic digestion. CEA samples retrieved from two different sources (purified from human colon carcinoma and human liver metastases) were digested with trypsin, endoproteinase Glu-C and pronase and the enzymatic hydrolysates subsequently analyzed by sheathless CE-MS/MS. The information obtained using these

enzymes allowed the detection of 20 out of the 28 potential *N*-glycosylation sites, as well as their site-specific dominant *N*-glycan subtype. In addition, partial least squares discriminant analysis (PLS-DA) was applied to easily differentiate the complex and characteristic glycopeptide maps of CEA samples purified from human colon carcinoma and human liver metastases. The comparison between the different *N*-glycosylation profiles allowed us to propose novel potential *N*-glycopeptide markers to distinguish between both types of CEA.

2. Materials and methods

2.1. Chemicals, reagents and samples

All the chemicals used in the preparation of solutions were of analytical reagent grade or higher. Isopropanol (iPrOH), methanol (MeOH), ammonium bicarbonate (ABC) and sodium hydroxide (NaOH) were purchased from Merck (Darmstadt, Germany). Water (LC-MS grade), acetonitrile (MeCN) (LC-MS grade), glacial acetic acid (HAc), hydrochloric acid (HCl), DL-dithiothreitol (DTT), iodoacetamide (IAA), ammonium acetate (AAC), formic acid (FA), TPCK-treated trypsin from bovine pancreas and pronase from *Streptomyces griseus* were provided by Sigma-Aldrich (Steinheim, Germany). Endoproteinase Glu-C from

Staphylococcus aureus V8 was supplied by Promega Corporation (Madison, WI, USA). CEA purified from human colon carcinoma was purchased from MyBioSource, Inc. (CEA1; San Diego, CA, USA) and Fitzgerald Industries International (CEA2; Acton, MA, USA). CEA purified from human liver metastases was obtained from Lee BioSolutions, Inc. (CEA3; Maryland Heights, MO, USA).

2.2. Digestion of CEA

2.2.1. Tryptic digestion

Firstly, 10 µg of CEA were diluted with 25 mM ABC to reach a concentration of 1 µg/µL. Following the manufacturer's instructions, one µL of 22 mM DTT was added for reduction of the disulfide bridges (30 min at 60°C) and cooled down at room temperature (rt). One µL of 72 mM IAA was then added to the samples for alkylation via incubation in the dark for 30 min. One µL of 78 mM DTT was added to end the alkylation (30 min at rt). Overnight tryptic cleavage was carried out with 2 µL of 0.5 µg/µL TPCK-treated trypsin (enzyme:substrate 1:10 m/m) at 37°C. Before CE-MS/MS analysis, CEA tryptic digests were diluted with 250 mM AAC at pH 4.0 (digest solution:leading electrolyte 3:2, v/v) to reach a digested protein concentration of 0.40 µg/µL.

2.2.2. Digestion with Endoproteinase Glu-C or Pronase

Ten µg of CEA were first diluted with 25 mM ABC to reach a concentration of 1 µg/µL. Following the manufacturer's instructions, one µL of 55 mM DTT was added (30 min at 60°C) and cooled down at rt. One µL of 180 mM IAA was then added to the samples for alkylation and incubated in the dark for 30 min. CEA digestions were carried out with 2.5 µL of 0.2 µg/µL of Glu-C or pronase (enzyme:substrate 1:20 m/m) at 37°C. Before CE-MS/MS analysis, CEA digests were diluted with 250 mM AAC at pH 4.0 (digest solution:leading electrolyte 3:2, v/v) to reach a digested glycoprotein concentration of 0.40 µg/µL.

2.3. Sheathless CE-MS/MS

CE experiments were carried out on a Sciex/Beckman Coulter CESI 8000 system (Sciex, Framingham, USA) equipped with a temperature controlled sample tray and a power supply able to deliver up to 30 kV. A 90 cm long (LT) × 30 µm ID × 150 µm OD bare fused-silica capillary with the high sensitivity porous sprayer (HSPS; Sciex/Beckman Coulter) in the outlet tip was used for all the separations. The background electrolyte (BGE) was composed of

a solution of 10% HAc v/v (pH 2.3). Prior to each sample injection, the capillary was rinsed with 0.1 M NaOH (2.5 min), water (4 min), 0.1 M HCl (2.5 min), water (4 min), and BGE (4 min). The 250 mM AAC at pH 4.0 added to the samples (3:2 v/v), acted as the leading electrolyte (LE) for the on-line preconcentration by transient isotachopheresis (t-ITP)²⁵. Samples were hydrodynamically injected at 8 psi for 60 s (corresponding to 70 nL). In all the experiments, sample injection was followed by a post-plug of BGE by applying a pressure of 0.5 psi for 25 s (corresponding to 2 nL). A separation voltage of 20 kV (normal polarity, anode at the injection capillary end) was applied for the electrophoretic separations at 25°C. Multiple injections ($n=6$) were carried out for every digested CEA sample (S1-6 for CEA1 samples, S7-12 for CEA2 samples, and S13-18 for CEA3 samples).

The CE apparatus was hyphenated to a UHR-QqTOF mass spectrometer (maXisImpact HD; Bruker Daltonics, Bremen, Germany) via the sheathless CE-MS interface, which was hold on a custom-made platform (Sciex/Beckman Coulter) to allow an optimal position of the capillary porous tip in front of the mass spectrometer nanospray shield. All the experiments were carried out using dopant

enriched nitrogen gas (DEN-gas), as it showed improved sensitivities for the analysis of glycopeptides when compared to the conventional nanospray sheathless CE-MS setup²⁶. For this purpose, an in-house made polymer cone was slid onto the housing of the porous tip, allowing for a coaxial sheath flow of the DEN-gas around the nanoESI emitter, using MeCN as dopant. Under optimized conditions, CE-MS/MS experiments were carried out in ESI positive mode using the following parameters: capillary voltage 1200 V, drying gas temperature 150°C, drying gas flow rate 1.2 L/min, nebulizer gas pressure 0.2 bar, quadrupole ion energy 3.0 eV and collision cell energy 7.0 eV. MS data were acquired between m/z 200 and 2000 with a spectral acquisition rate of 1 Hz. MS/MS data were acquired in data dependent mode with an absolute threshold of 4548 counts and active exclusion. Specific m/z values that were already acquired three times were excluded and released after 0.8 min, unless the precursor had a 5 times higher intensity than the previous observed intensity.

2.4. Data Analysis

CE-MS/MS data were analyzed with DataAnalysis 4.2 (Build 387, Bruker Daltonics). Mass spectra were recalibrated using sodium

acetate clusters detected at the beginning of the electrophoretic runs. From the automatically acquired MS/MS spectra of the precursor ions, the glycan moieties of the *N*-glycopeptides were deduced by manual assignment on the basis of general glycan fragmentation rules. Mass errors (E_r) were below 10 ppm throughout (see Supplementary Tables S-1, S-2 and S-3 for digestion of CEA samples with trypsin, Glu-C and pronase, respectively). Glycoforms were labeled taking into account the glycan compositions in terms of number of hexoses (H), N-acetylglucosamines (N), fucoses (F) and sialic acids (S) and these abbreviations will be further used throughout this manuscript.

Once CEA *N*-glycopeptide glycoforms were identified, extracted ion electropherograms (EIEs, smoothed with a Gaussian fit) were generated for the first three isotopic peaks of the doubly, triply and quadruply charged analytes with a width of $m/z \pm 0.02$ units. *N*-glycopeptide peak areas were then measured from the different EIEs and normalized with the peak area of a CEA reference peptide (S_{127} DLVNEEATGQFR₁₃₉, S_{40} TPFNVAE₄₇ and A_{89} TPGP₉₃/T₉₀PGPA₉₄ for digestion with trypsin, Glu-C and pronase, respectively). The reference peptides showed peak areas and migration times within the ranges of the

detected *N*-glycopeptides. Relative peak areas of the *N*-glycopeptides were autoscaled (mean centered and scaled to unit standard deviation) before multivariate data analysis was applied to the data obtained for CEA samples from colon carcinoma and liver metastases with the three different enzymes. First, principal component analysis (PCA) was applied to explore the different data classes and the presence of outliers^{17,27}. Second, partial least squares discriminant analysis (PLS-DA) was performed to maximize classes separation and rapidly differentiate the colon carcinoma and liver metastases CEA samples, as well as to identify which *N*-glycopeptides were the most significant to discriminate between the two types of CEA^{17,28}. In this regard, there are numerous methods for feature selection when considering PLS-DA models. The variable importance in the projection (VIP) method was used^{17,29}, because it is one of the preferred methods due to its ability for handling multicollinear data³⁰. For each model, VIP scores estimate the importance of each feature in the projection. Only features with a VIP score over a particular threshold (usually VIP=1) are considered important and selected for further analysis. In all cases, leave-one-out cross-validation was used to assess the performance of the built models^{17,31}. SOLO (Version 8.1,

student edition, Eigenvector, Research Inc., Wenatchee, WA, USA) was used for PCA and PLS-DA analysis.

3. Results and discussion

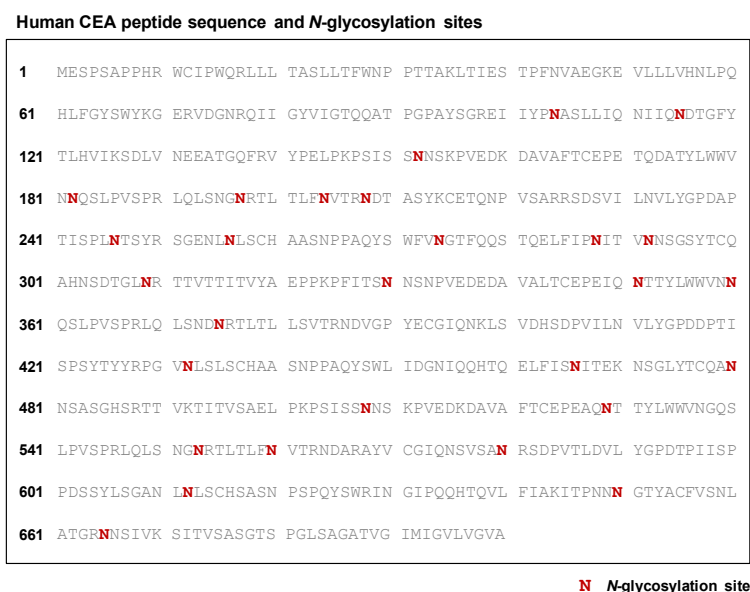
3.1. Analysis of CEA *N*-glycopeptides by sheathless CE-MS/MS

Figure 1 shows the amino acid sequence of CEA, its potential 28 *N*-glycosylation sites (marked in red), and the percentage of glycosylation (60% m/m glycans). In addition to the numerous *N*-glycosylation sites, CEA displays a considerable heterogeneity in the sugar content of the protein, therefore the analysis and structure elucidation of glycopeptides from CEA enzymatic digests is not straightforward. In this study, CEA samples purified from human colon carcinoma (CEA1 and CEA2) and human liver metastases (CEA3) were digested with trypsin, Glu-C and pronase and the enzymatic hydrolysates subsequently analyzed by sheathless CE-MS/MS. The analysis of individual *N*-glycopeptides allowed a detailed characterization of CEA carbohydrate structures by collision induced dissociation (CID) MS/MS. Moreover, data of the three different proteolytic enzymes were combined in order to maximize the coverage of *N*-glycosylation sites and to obtain a more

comprehensive characterization. Several authors have pointed out that using several proteases or mixtures of proteases could benefit the sequence coverage and improve the characterization of PTMs^{32–34}. Trypsin and Glu-C are specific serine proteases that selectively cleave peptide bonds C-terminal to lysine (K) and arginine (R) residues (in the case of trypsin) and aspartic (D) and glutamic acid (E) residues (in the case of Glu-C). Although MS/MS analysis of trypsin and Glu-C digests often provides enough information for an appropriate characterization of carbohydrate moieties in glycoproteins³⁴, in some cases these proteases may fail to cleave in highly PTM regions of proteins. Moreover, in many MS analyses, large non-glycosylated peptides can suppress the ionization of glycopeptides, which often show worse detection sensitivity³⁴. Alternatively, pronase is a non-specific mixture of proteolytic enzymes that can be employed in order to overcome some of the limitations found with trypsin and Glu-C. In this regard, pronase digests result in a final glycopeptide hydrolysate with significantly smaller average size of peptide moieties (often from 1 to 8 amino acids)^{35,36}. In addition, non-glycosylated peptide stretches are usually cleaved to the level of single amino acids, hence avoiding interferences on the MS analysis of *N*-glycopeptides^{35,36}. However, the main

disadvantage of using pronase is that the interpretation of peptides and *N*-glycopeptides found in the hydrolysates is more complex than when specific enzymes are used.

of these electropherograms indicated that glycopeptides migrated between 25-32 min (trypsin), 25-49 min (Glu-C) and 28-45 min (pronase) (see Figure 2 (ii)).



$M_{r, \text{glycoprotein}} = 180,000$ $M_{r, \text{peptide}} = 71,328$ $M_{r, \text{carbohydrate}} = 108,671$ (60% m/m of glycans)

Figure 1. Amino acid sequence of CEA and its 28 potential *N*-linked glycosylation sites marked in red. $M_{r, \text{glycoprotein}}$ and $M_{r, \text{peptide}}$ were found using the ExPASy tool. Percentage of *N*-glycosylation was calculated by difference between the $M_{r, \text{glycoprotein}}$ and the $M_{r, \text{peptide}}$.

Figures 2 (i) show the base peak electropherograms (BPEs) obtained for a CEA2 sample after enzymatic digestion with trypsin (A), Glu-C (B) and pronase (C) and analysis by sheathless CE-MS/MS. In all cases, the obtained MS/MS data were screened for the presence of *N*-glycopeptides by inspecting the EIE (MS/MS) of 366.1401 *m/z* (corresponding to the oxonium ion of a single-charged HN unit fragmented from *N*-glycopeptides). The pattern

The MS/MS data were then inspected for neutral losses of one H (162.0528 M_r), one N (203.0794 M_r), one F (146.0579 M_r) or combination of them (i.e. a neutral loss of 365.1322 M_r would indicate the HN unit). Furthermore, the presence of sialic acid fragment ions at 292.1033 *m/z* (corresponding to the single-charged S unit) or sialylated structures at 657.2355 *m/z* (corresponding to the single-charged HNS unit) were also screened. In

the case of pronase, the list of deduced peptide M_r was used for matching to theoretical M_r of randomly cleaved peptides from CEA using the FindPept tool with $E_r \leq 10$ ppm³⁷. Peptide sequences, which were indicated by this tool, were then compared with the peptide sequence information obtained from the MS/MS spectra of the *N*-glycopeptide species. As an example, the interpretation of the fragmentation spectra of some of the *N*-glycopeptides that, as it will be discussed later on, allowed differentiating between CEA samples from human colon carcinoma and human liver metastases, are shown in Supplementary Figure S-1 for digestion of CEA samples with trypsin, Glu-C and pronase.

Tables 1-A-B-C show the *N*-glycosylation sites (and their corresponding number of *N*-glycopeptides) detected by sheathless CE-MS/MS after digestion with trypsin (A), Glu-C (B) and pronase (C) for the two samples purified from human colon carcinoma (CEA1 and CEA2) and the sample purified from human liver metastases (CEA3). Supplementary Tables S-1, S-2 and S-3 also show the observed average m/z ($n=6$), the relative m/z error and the %relative abundances (normalizing the integrated peak areas by the total integrated peak areas for the *N*-glycosylation site) of the

detected *N*-glycopeptides for the analyzed CEA samples using the three enzymes. As can be observed in Table 1-A, the tryptic digestion allowed the identification of a total number of 128 different *N*-glycopeptides (considering the three CEA sample types) corresponding to 9 *N*-glycosylation sites (*N*-197/*N*-553, *N*-204/*N*-560, *N*-375, *N*-580, *N*-612, *N*-650 and *N*-665). However, it was not possible to differentiate between *N*-glycopeptides belonging to *N*-197 and *N*-553 or *N*-204 and *N*-560, as they presented the same peptide sequences after digestion with trypsin (LQLSNG_{N197/553}R and TLTLFN_{204/560}VTR, respectively). With regard to the non-detected *N*-glycosylation sites, this was probably due to the fact that, for most of the peptides, tryptic digestion resulted in large peptide moieties (number of amino acids (aa) ≥ 20). Furthermore, certain peptide sequences carried several potential *N*-glycosylation sites. If all of these sites were occupied, the complexity of the resulting glycopeptides would increase, hence complicating the detection and identification of these sites (see Table 1 for the peptide sequence length in the *N*-glycopeptides). In general, the same glycan compositions could be identified on the same *N*-glycosylation site for CEA1 and CEA2 purified from colon carcinoma, whereas some differences were found with CEA3 from

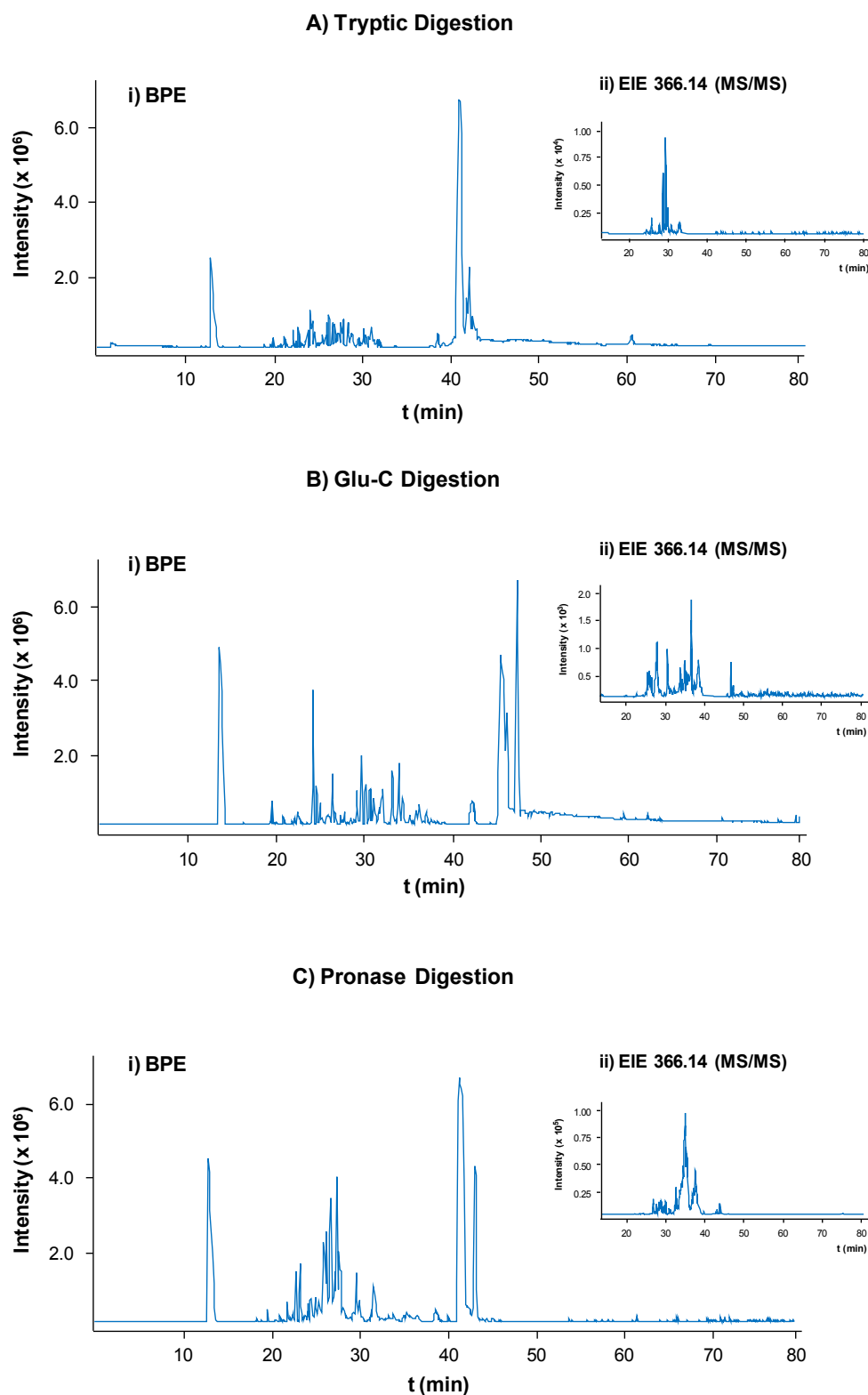


Figure 2. (i) Base peak electropherograms (BPEs) and (ii) extracted ion electropherograms (EIEs) (MS/MS) for the nominal m/z 366 obtained for a colon carcinoma CEA sample (CEA2) after enzymatic digestion with (A) trypsin, (B) Glu-C and (C) pronase and analysis by sheathless CE-MS/MS.

liver metastases, thus suggesting that the two sources might reveal a different *N*-glycopeptide profile.

The higher number of *N*-glycopeptides was observed for the *N*-197/*N*-553 (52 identified *N*-glycopeptides if we consider the *N*-197/*N*-553 deamidated) followed by the *N*-204/*N*-560 (40 identified *N*-glycopeptides) and the *N*-665 (18 identified *N*-glycopeptides). For the *N*-650, *N*-375, *N*-612 and *N*-580, a smaller number of *N*-glycopeptides could be identified namely, 11, 4, 2 and 1, respectively. Throughout the identified *N*-glycosylation sites, different types of *N*-glycans could be observed on the same site, including some high mannose and hybrid types but mainly complex (bi-, tri- and tetraantennary structures) were detected (Supplementary Table S-1). Furthermore, highly fucosylated (including terminal and core fucosylation) and sialylated *N*-glycopeptides were predominantly observed for the *N*-204/*N*-560, *N*-580, *N*-612, *N*-650 and *N*-665, especially for the CEA3 sample purified from human liver metastases (Supplementary Table S-1).

The digestion with Glu-C allowed the identification of a total number of 50 *N*-glycopeptides corresponding to 4 *N*-glycosylation sites (*N*-152/*N*-508, *N*-466 and *N*-

580) (Table 1-B). Most of these *N*-glycosylation sites were exclusively detected after digestion with Glu-C, with the exception of the *N*-580, which was also observed after digestion with trypsin. However, a higher degree of micro-heterogeneity was found with Glu-C (9 vs. 1 *N*-glycopeptides were detected), hence suggesting the importance of the selection of the enzyme and its ability to generate different peptide moieties. As Glu-C generated identical peptide sequences for the potential *N*-152 and *N*-508 sites (LPKPSISSN_{152/508}NSKPVE), the identified *N*-glycopeptides could not be differentiated. Similar to the tryptic digestion, the same glycan compositions could be identified on the same *N*-glycosylation site for the two colon carcinoma samples CEA1 and CEA2, whereas some differences were found for the liver metastasis sample CEA3. In general, as can be deduced from our results, tryptic digestion was found to provide a higher degree of *N*-glycopeptide coverage compared to the Glu-C digestion. The observed differences in the effect of both proteolytic enzymes on the CEA samples could be explained by their different specificity and stability. On the basis of the manufacturer's specifications, the trypsin used in this study was chemically modified, yielding a highly active and stable enzyme that was extremely resistant to autolytic digestion in

solution. Such modifications were not reported for Glu-C and up to three missed cleavage sites (MCs) in the neighborhood of the *N*-glycosylation sites were observed after digestion, hence suggesting a lower efficiency of the enzyme. It is worth mentioning that zero MCs were observed for the identified *N*-glycopeptides of *N*-466 and *N*-580 sites, whereas between zero and three MCs were found for the glycosylation sites *N*-152/*N*-508. In this regard, all the MCs (from zero up to three) were used for the structural interpretation (Supplementary Table S-2). In this case, the higher number of *N*-glycopeptides was observed for the *N*-466 site (21 *N*-glycopeptides) followed by the *N*-152/*N*-580 sites (20 *N*-glycopeptides) and the *N*-580 site (9 *N*-glycopeptides). Complex *N*-glycopeptides (bi-, tri- and tetraantennary structures) were identified for the *N*-466 and the *N*-580 sites (see Supplementary Table S-2), whereas only high mannose *N*-glycopeptides were observed for the *N*-152/*N*-508 sites (Supplementary Table S-2). In contrast to the results obtained with tryptic digestion, sialylated *N*-glycopeptides were not detected after digestion with Glu-C, probably due the larger peptide moieties bonded to the *N*-glycosylated amino acids (see Table 1 for the peptide sequence length in the *N*-glycopeptides).

As it is shown in Table 1-C, the digestion with pronase allowed the identification of a total number of 41 different *N*-glycopeptides corresponding to 11 *N*-glycosylation sites (*N*-104, *N*-208, *N*-256, *N*-274, *N*-330, *N*-351, *N*-375, *N*-432, *N*-529, *N*-612 and *N*-665). Most of these *N*-glycosylation sites were not yet identified with the previous enzymes, with the exception of the *N*-375, the *N*-612 and the *N*-665 sites, which were also observed after digestion with trypsin. Except for the *N*-612 site (6 (pronase) vs 2 (trypsin) *N*-glycopeptides were detected), a higher number of *N*-glycopeptides was found for these sites using trypsin. It is worth mentioning that more *N*-glycopeptides than those reported in Table 1-C were identified after digestion with pronase (see Supplementary Table S-3-B). However, these *N*-glycopeptides were discarded for further discussion, as multiple overlaps between *N*-glycosylation sites were observed, hence making the interpretation of the data extremely difficult. The digestion with pronase allowed the identification of a higher number of *N*-glycosylation sites, probably because digestion resulted in *N*-glycopeptides with smaller peptide backbones bonded to the *N*-glycosylated amino acids. However, predicting the peptide sequences that were produced after digestion was quite challenging, as non-specific enzymes cleave the

Table 1. *N*-glycosylation sites (and total number of *N*-glycopeptides) detected by sheathless CE-MS/MS ($n=6$) after digestion with trypsin (A), Glu-C (B) and pronase (C) for the CEA samples purified from human colon carcinoma (CEA1 and CEA2) and human liver metastases (CEA3). The structures of the detected *N*-glycopeptides are summarized in Supplementary Tables S-1, S-2 and S-3. (aa = aminoacid).

			Number of <i>N</i> -glycopeptides=128			
A) Trypsin	<i>N</i> -glycosylation sites	Peptide length (number of aa)	CEA1 Human Colon Carcinoma	CEA2 Human Colon Carcinoma	CEA3 Human Liver Metastases	Total
	<i>N</i> -104+ <i>N</i> -115 ^a	28	-	-	-	-
	<i>N</i> -152	21	-	-	-	-
	<i>N</i> -182	30	-	-	-	-
	<i>N</i> -197/ <i>N</i> -553 ^b	8	25	25	23	26
	<i>N</i> -197/ <i>N</i> -553-De ^c	8	26	26	21	26
	<i>N</i> -204/ <i>N</i> -560 ^b	9	35	34	16	40
	<i>N</i> -208	7	-	-	-	-
	<i>N</i> -246	25	-	-	-	-
	<i>N</i> -256+ <i>N</i> -274+ <i>N</i> -288+ <i>N</i> -292+ <i>N</i> -309 ^a	60	-	-	-	-
	<i>N</i> -330+ <i>N</i> -351+ <i>N</i> -360 ^a	58	-	-	-	-
	<i>N</i> -375	8	4	4	2	4
	<i>N</i> -432+ <i>N</i> -466 ^a	72	-	-	-	-
	<i>N</i> -480	18	-	-	-	-
	<i>N</i> -508	24	-	-	-	-
	<i>N</i> -529	30	-	-	-	-
	<i>N</i> -580	14	-	-	1	1
	<i>N</i> -612	47	-	-	2	2
	<i>N</i> -650	20	-	-	11	11
	<i>N</i> -665	6	11	4	18	18

			Number of <i>N</i> -glycopeptides=50			
B) Glu-C ^d	<i>N</i> -glycosylation sites	Peptide length (number of aa)	CEA1 Human Colon Carcinoma	CEA2 Human Colon Carcinoma	CEA3 Human Liver Metastases	Total
	<i>N</i> -104+ <i>N</i> -115 ^a	17	-	-	-	-
	<i>N</i> -152/ <i>N</i> -508 ^b	15	20	19	18	20
	<i>N</i> -182+ <i>N</i> -197+ <i>N</i> -204+ <i>N</i> -208 ^a	36	-	-	-	-
	<i>N</i> -246	15	-	-	-	-
	<i>N</i> -256+ <i>N</i> -274 ^a	30	-	-	-	-
	<i>N</i> -288+ <i>N</i> -292 ^a	22	-	-	-	-
	<i>N</i> -309+ <i>N</i> -330 ^a	31	-	-	-	-
	<i>N</i> -351+ <i>N</i> -360 ^a	26	-	-	-	-
	<i>N</i> -375	13	-	-	-	-
	<i>N</i> -432	36	-	-	-	-
	<i>N</i> -466	8	21	21	9	21
	<i>N</i> -480	30	-	-	-	-
	<i>N</i> -529+ <i>N</i> -553+ <i>N</i> -560 ^a	39	-	-	-	-
	<i>N</i> -580	23	9	9	-	9
	<i>N</i> -612/ <i>N</i> -650/ <i>N</i> -665 ^a	83	-	-	-	-
			Number of <i>N</i> -glycopeptides=41			
C) Pronase	<i>N</i> -glycosylation sites	Peptide length (number of aa)	CEA1 Human Colon Carcinoma	CEA2 Human Colon Carcinoma	CEA3 Human Liver Metastases	Total
	<i>N</i> -104	4	1	1	-	1
	<i>N</i> -115	-	-	-	-	-
	<i>N</i> -152	-	-	-	-	-
	<i>N</i> -182	-	-	-	-	-
	<i>N</i> -197	-	-	-	-	-
	<i>N</i> -204	-	-	-	-	-
	<i>N</i> -208	4	-	-	1	1
	<i>N</i> -246	-	-	-	-	-

			Number of <i>N</i> -glycopeptides=41			
C) Pronase	<i>N</i> -glycosylation sites	Peptide length (number of aa) ^e	CEA1 Human Colon Carcinoma	CEA2 Human Colon Carcinoma	CEA3 Human Liver Metastases	Total
	<i>N</i> -256	7	3	3	4	5
	<i>N</i> -274	2-4 ^f	5	5	7	7
	<i>N</i> -288	-	-	-	-	-
	<i>N</i> -292	-	-	-	-	-
	<i>N</i> -309	-	-	-	-	-
	<i>N</i> -330	5	1	1	1	1
	<i>N</i> -351	7	5	5	5	5
	<i>N</i> -360	-	-	-	-	-
	<i>N</i> -375	4	1	1	1	1
	<i>N</i> -432	5	1	1	-	1
	<i>N</i> -466	-	-	-	-	-
	<i>N</i> -480	-	-	-	-	-
	<i>N</i> -508	-	-	-	-	-
	<i>N</i> -529	4	7	7	8	10
	<i>N</i> -553	-	-	-	-	-
	<i>N</i> -560	-	-	-	-	-
	<i>N</i> -580	-	-	-	-	-
	<i>N</i> -612	6	3	3	4	6
	<i>N</i> -650	-	-	-	-	-
	<i>N</i> -665	3	2	2	3	3

^aMultiple *N*-glycosylation sites are present in the same sequence. ^bThey cannot be differentiated as they share the same sequence. ^c*N*-197/*N*-553 also exist as deamidated. ^d*N*-glycopeptides with missed cleavage sites (MCs) from zero to three were used for the structural interpretation. ^eAs pronase is a non-specific enzyme, peptide length can only be reported for the identified *N*-glycosylation sites. ^fTwo different peptide sequences were observed for this *N*-glycosylation site (NG and NGTF).

glycoprotein in an undefined manner to produce these shorter glycopeptides and peptides. Again, as it occurred with the rest of enzymes, the same glycan compositions could be identified on the same *N*-glycosylation site for the colon carcinoma samples CEA1 and CEA2, which were different to some extent from the liver metastasis sample CEA3. In this case, the higher number of *N*-glycopeptides was observed for the *N*-529 (10 identified *N*-glycopeptides) followed by the *N*-274 (7 identified *N*-glycopeptides) and the *N*-612 (6 identified *N*-glycopeptides). For the rest of *N*-glycosylation sites, a lower number could be identified, which varied between 1 and 5 *N*-glycopeptides (Table 1-C and Supplementary Table S-3-A). A variety of high mannose and complex *N*-glycopeptides (bi- and triantennary structures) were identified for many of the *N*-glycosylation sites (Supplementary Table S-3-A). Furthermore, highly fucosylated (including terminal and core fucosylation) and sialylated *N*-glycopeptides were predominantly observed for the *N*-274 (H6N5F3S1 and H5N4S2) and the *N*-612 (H6N5F4S1, H6N5F2S2 and H6N5F3S2), especially for the liver metastases sample CEA3 (Supplementary Table S-3-A).

In order to increase the reliability of the identification, the migration behavior of the

identified *N*-glycopeptides, which is related to their charge-to-size ratio, was also investigated. In this regard, while the number and type of glycans attached to the carbohydrate chains affect the size of the glycopeptides, and hence M_r , the specific presence of sialic acids strongly contributes to their charge and vastly influences their migration times³⁸. As an example, Figure 3 shows the EIEs of the eleven *N*-650 glycopeptides identified by sheathless CE-MS/MS after digestion with trypsin for the liver metastases CEA3 sample. As it is shown, *N*-glycopeptides without sialic acids (global charge was positive at pH~2.3 (BGE)) appeared at the shorter migration times (Figure 3). In contrast, *N*-glycopeptides with sialic acids were detected at longer migration times, because sialic acids carry a negative charge at pH~2.3 (BGE) and decreased the global positive molecular charge. Moreover, as illustrated in Figure 3, the higher the number of sialic acids, the lower the global molecular charge and, therefore, the sialylated *N*-glycopeptides migrated in two small time windows corresponding to the *N*-glycopeptides with one sialic acid and two sialic acids. Several authors have explored the use of different semiempirical approaches which relate the M_r , the charge (q) and the electrophoretic mobility (μ_e) in order to predict the electrophoretic migration behavior,

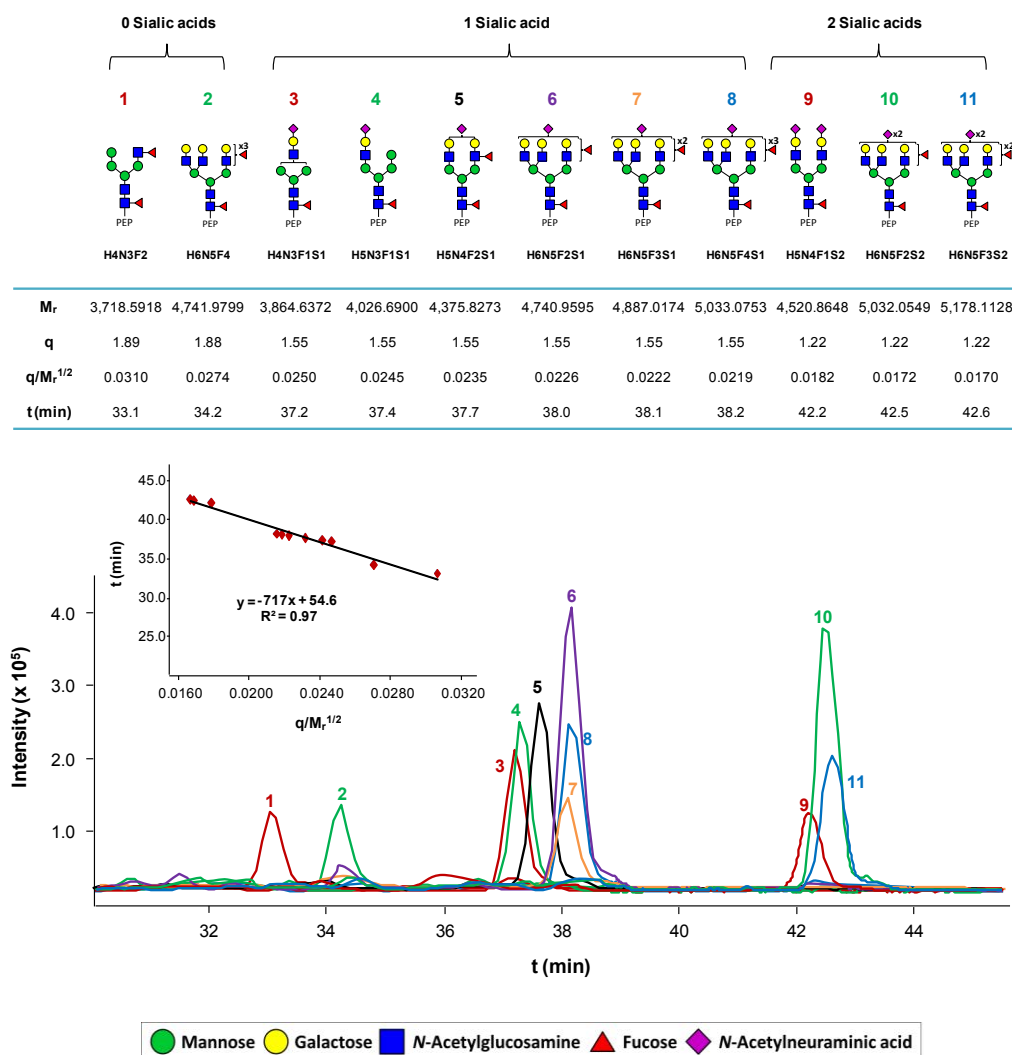


Figure 3. Extracted ion electropherograms (EIEs) of the identified N-650 glycopeptides of a liver metastases CEA sample (CEA3) obtained by sheathless CE-MS/MS after digestion with trypsin (1: H4N3F2, 2: H6N5F4, 3: H4N3F1S1, 4: H5N3F1S1, 5: H5N4F2S1, 6: H6N5F2S1, 7: H6N5F3S1, 8: H6N5F4S1, 9: H5N4F1S2, 10: H6N5F2S2, 11: H6N5F3S2). M_r , q and $q/M_r^{1/2}$ ($n=6$) values are also shown for the identified N-glycopeptides. Experimental migration times of the N-glycopeptides in the electropherogram were fitted to a linear regression line with the classical polymer model ($q/M_r^{1/2}$)³⁹.

study structural modifications, charge characteristics and conformations of peptides. Barroso et al. recently shown that the classical polymer model ($\mu \propto q/M_r^{1/2}$) can be applied to predict the electrophoretic migration behavior of

N-glycopeptides³⁹. Figure 3 shows as an example the theoretical M_r , q and $q/M_r^{1/2}$ for the eleven identified N-650 glycopeptides. As can be observed, the linear correlation between the experimental migration times of the N-

glycopeptides in the electropherogram and their theoretical $q/M_r^{1/2}$ was excellent, hence supporting the structural interpretation obtained from MS/MS spectra.

3.2. Multivariate data analysis

First, we used principal component analysis (PCA) for the unsupervised identification of trends and clustering of the data, as well as outliers²⁷. Three separate PCA models, for CEA samples digested with trypsin, Glu-C and pronase, and a unique PCA model, for CEA samples digested with the three enzymes were built. Results were similar in both cases, but the first strategy based on three separate PCA models was preferred because it allowed an easier interpretation of the data. The three separate PCA models (scores and loadings plots are not shown) allowed discrimination between colon carcinoma and liver metastases CEA samples. Two principal components (PCs) explained a total variance of 90%, 87% and 82% (by the sum of PC 1 and PC 2) for the samples digested with trypsin, Glu-C and pronase, respectively. In the case of the digestion with pronase, sample 9 (S9) was detected as an outlier (probably due to the lower signal-to-noise ratio in this specific run) and, therefore, it was discarded for further analysis

(see Supplementary Figure S-2). In all cases, the first component helped to differentiate between samples purified from colon carcinoma and liver metastases. However, PCA did not reveal the importance of the *N*-glycopeptides for differentiation between both types of CEA. In this regard, partial least squares discriminant analysis (PLS-DA)²⁸ was used to build a refined classification model with improved class separation that would also facilitate data interpretation and identification of potential *N*-glycopeptide markers to distinguish between both groups of samples.

PLS-DA models were built considering the two classes of samples observed by PCA and excluding the outlier sample (S9) for pronase digestion. Figures 4-A-B-C (i) show the scores plots for the CEA samples taking into account the three PLS-DA models. Two latent variables (LVs) explained 90, 84 and 68% of X-variance and 99, 89 and 90% of Y-variance for the samples digested with trypsin, Glu-C and pronase, respectively. All the PLS-DA models allowed perfect discrimination between colon carcinoma and liver metastases samples along the x-axis, which corresponds to the first LV. The loadings plots (Supplementary Figure S-3) showed the contribution of the *N*-glycopeptides (variables) to the different LVs. As can be seen

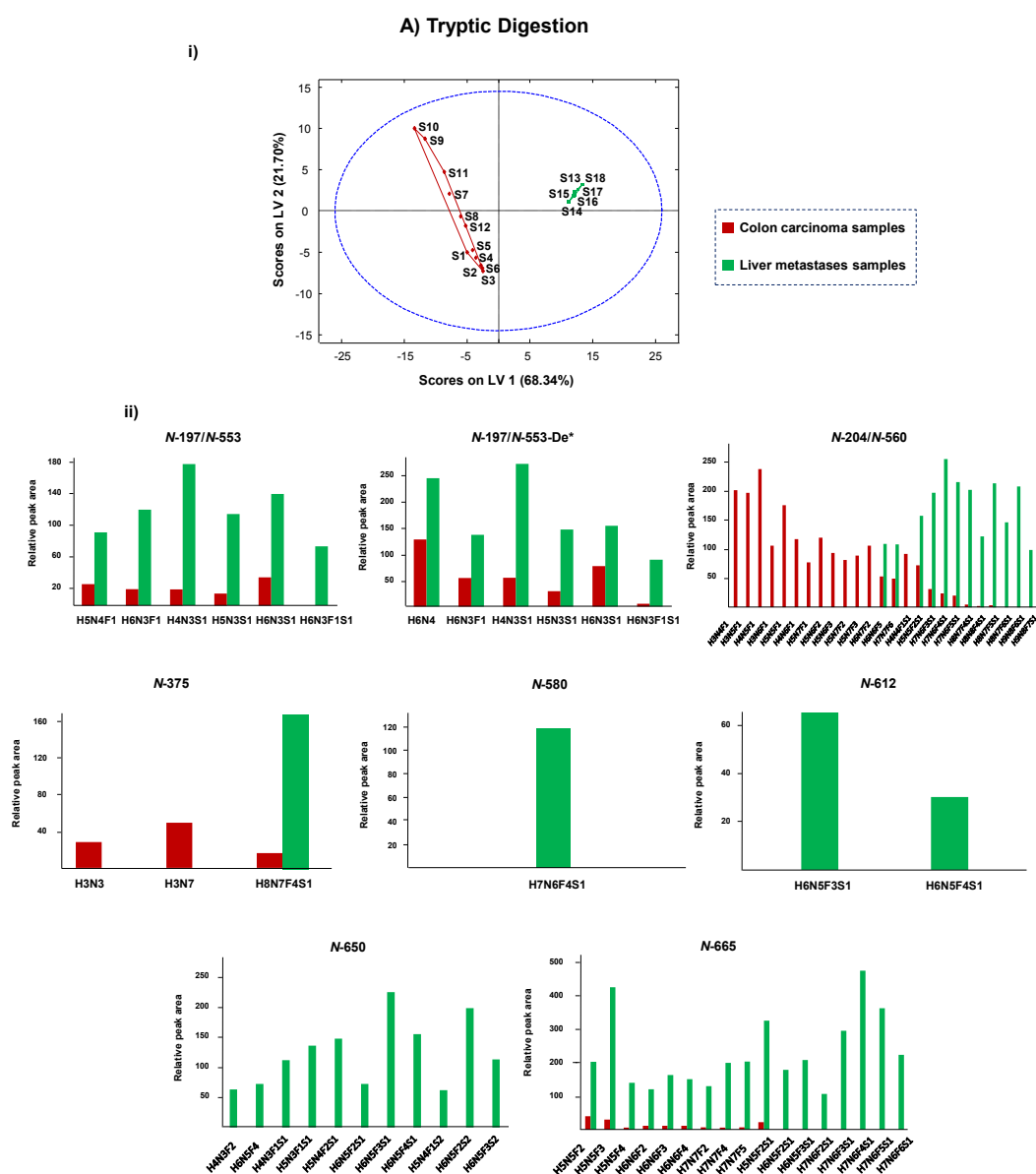
for the digestion with trypsin (Supplementary Figure S-3-A), *N*-glycopeptides with the highest content of sialic acids and fucoses, which were in the positive side of the x-axis (especially those corresponding to *N*-204/*N*-560, *N*-580, *N*-612, *N*-650 and *N*-665, Supplementary Table S-1), were increased in the liver metastases CEA sample when compared to colon carcinoma CEA samples, thus suggesting the important role of sialylation and fucosylation for differentiation of both types of CEA. In the case of the digestion with Glu-C (Supplementary Figure S-3-B), *N*-glycopeptides presenting fucoses (ranging from 1 to 3), which were mostly in the negative side of the x-axis (*N*-glycopeptides corresponding to *N*-466 and *N*-580, Supplementary Table S-2), were increased in CEA samples purified from human colon carcinoma. With regard to the results obtained after digestion with pronase (Supplementary Figure S-3-C), *N*-glycopeptides with the highest number of sialic acids and fucoses, which were in the positive side of the x-axis (especially those corresponding to *N*-274, *N*-612 and *N*-665, Supplementary Table S-3-B), were increased in the liver metastases CEA sample when compared to colon carcinoma CEA samples. In addition to this qualitative information, the VIP scores allowed now to quantify the influence of the different *N*-

glycopeptides on the separation between the two types of CEA. In this regard, VIP scores values higher than 1^{29,30} were used as a feature selection tool in order to choose only the most relevant candidate *N*-glycopeptides for each PLS-DA model (71 out of 128 for tryptic digestion, 26 out of 50 for Glu-C digestion and 24 out of 41 for pronase digestion). VIP scores values are shown in Supplementary Figures S-4-A-B-C for digestion with trypsin, Glu-C and pronase, respectively. The bar graphs in Figures 4-A-B-C (ii) show the relative peak areas for the identified *N*-glycopeptides (grouped by *N*-glycosylation sites) that were useful (VIP>1) to distinguish between the CEA1 and CEA2 samples purified from colon carcinoma (red bars) and the CEA3 sample purified from human liver metastases (green bars) after digestion with trypsin, Glu-C and pronase, respectively.

As can be observed in Figure 4-A (ii) for tryptic digestion, in general, relative peak areas of the *N*-glycopeptides presenting fucoses (ranging from 1 to 7) and sialic acids (ranging from 1 to 2), were found to increase in the CEA3 liver metastases sample when compared to colon carcinoma CEA1 and CEA2 samples. This trend was observed for all the *N*-glycosylation sites, with the exception of the *N*-204/*N*-560, where

some *N*-glycopeptides with the above mentioned characteristics were found to be upregulated in colon carcinoma CEA1 and CEA2 samples (see the red bars in the graphic for *N*-204 and *N*-560 in Figure 4-A (ii)). In contrast, as it is shown in Figure 4-B (ii) for the

digestion with Glu-C, relative peak areas of the *N*-glycopeptides presenting fucoses (ranging from 1 to 3) were found to decrease in the CEA3 liver metastases sample when compared to colon carcinoma CEA1 and CEA2 samples.



B) Glu-C Digestion

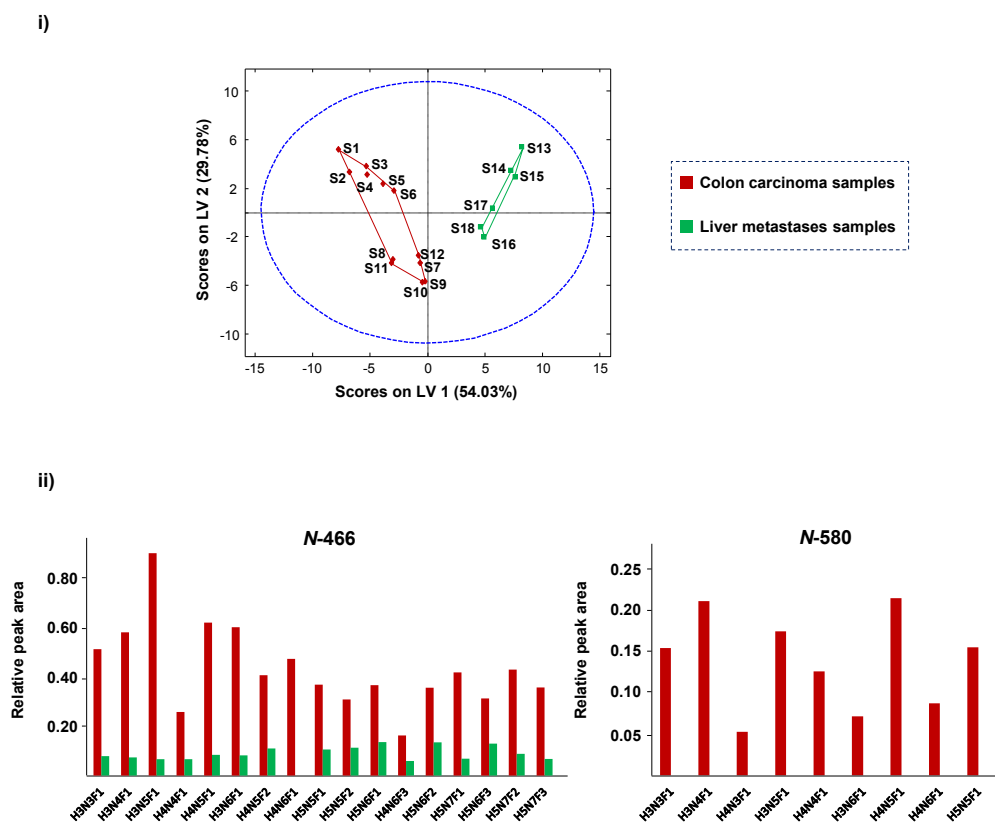


Figure 4 (continue). (i) PLS-DA scores plots for the model of the CEA samples and (ii) relative peak areas for the identified *N*-glycopeptides that are significant ($VIP > 1$) to distinguish between colon carcinoma CEA1 and CEA2 samples and the liver metastases CEA3 sample after digestion with (B) Glu-C.

With regard to the results obtained after digestion with pronase (Figure 4-C (ii)), different trends were observed for high mannose and complex *N*-glycopeptides. As an example, relative peak area of the H6N2 *N*-256 glycopeptide was found to increase in colon carcinoma CEA samples, whereas the opposite trend was observed for the H8N2 and H9N2 *N*-256 glycopeptides. With regard to the complex *N*-glycopeptides presenting only 1 fucose, relative peak areas of H3N2F1 (*N*-432) and

H3N3F1 (*N*-612) were found to decrease in the liver metastases CEA sample. In contrast, relative peak areas of the *N*-glycopeptides presenting fucoses (ranging from 2 to 4) and sialic acids (ranging from 1 to 2), were found to be upregulated in the liver metastases CEA3 sample when compared to colon carcinoma CEA1 and CEA2 samples. In this regard, several authors have shown that sialylated and fucosylated carbohydrates have an important role in a variety of biological

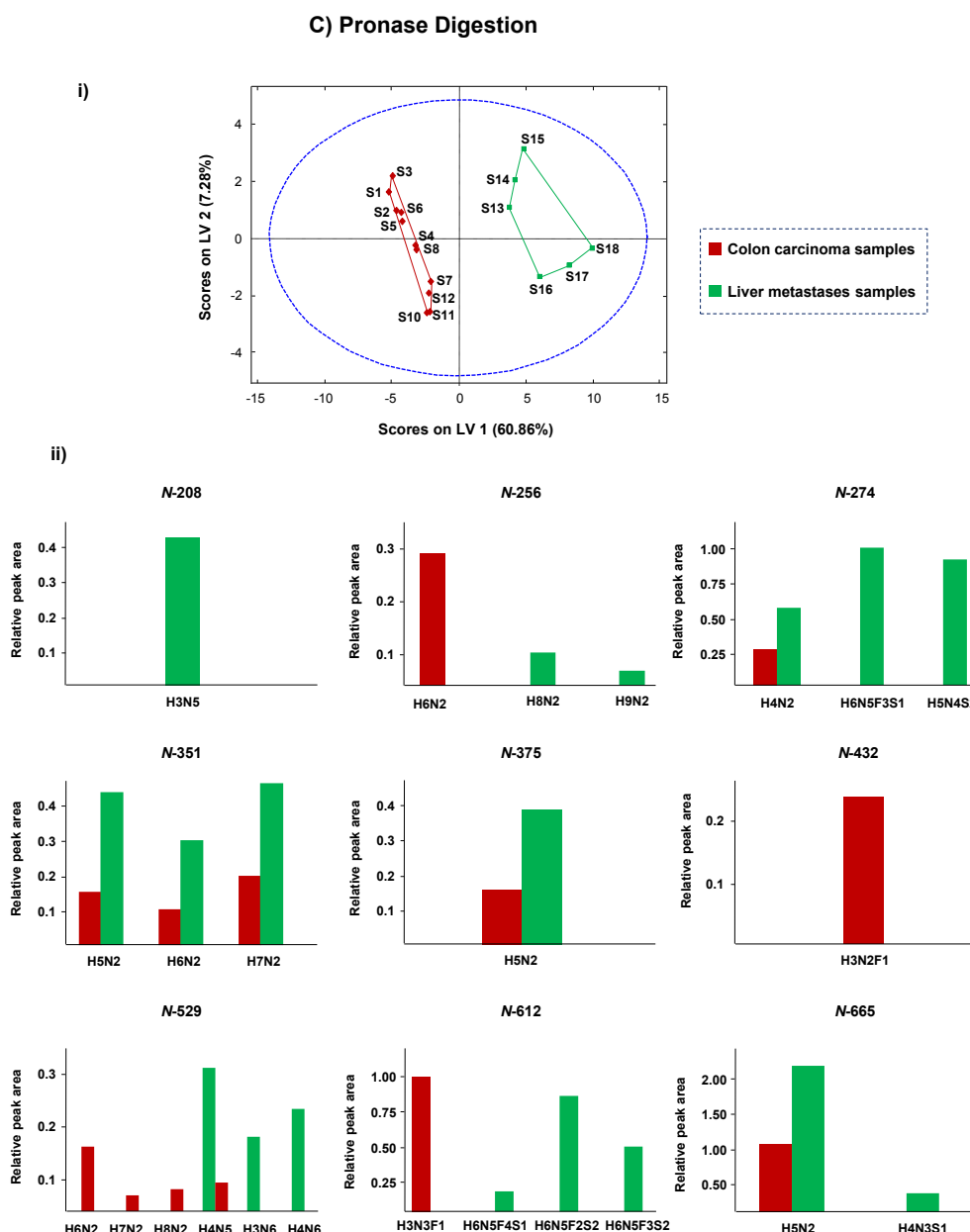


Figure 4 (continue). (i) PLS-DA scores plots for the model of the CEA samples and (ii) relative peak areas for the identified *N*-glycopeptides that are significant (VIP>1) to distinguish between colon carcinoma CEA1 and CEA2 samples and the liver metastases CEA3 sample after digestion with (C) pronase.

processes such as cellular recognition, cell adhesion and cell signaling^{40,41}. Moreover, an increase in global sialylation and fucosylation (due to the altered expression of glycosyltransferases and fucosyltransferases)

has been closely associated with cancer^{40,41}. Particularly, increased sialylation has been often associated with cancer invasiveness, and elevated serum total sialic acid (TSA) levels have been reported in several

malignancies based on their degree of metastatic involvement, with the highest values corresponding to samples purified from liver or other metastases^{42,43}. Similarly, global fucosylation (including terminal and core fucosylation) has also been found to increase in several cancers with highly malignant or metastatic tumors^{44,45}. This explanation agreed and reinforced to a large extent the results obtained in the present work, as, in general, the highest levels of fucosylation (between 3 and 7 fucoses) and, especially, sialylation (between 1 and 2 sialic acids), were observed for the liver metastases CEA sample, which presented a higher degree of metastatic involvement. Furthermore, a multivariate strategy based on advanced chemometric tools allowed the discrimination between specific *N*-glycosylation sites, which could be useful in order to explain the metastasis process in patients with colon carcinoma. Particularly, for the *N*-glycosylation sites *N*-204/*N*-560 (identified with trypsin) and *N*-466 and *N*-580 (identified with Glu-C), relative peak areas of the *N*-glycopeptides presenting fucoses (in general, ranging from 1 to 3), were found to be upregulated in the colon carcinoma CEA1 and CEA2 samples, opposite trend to the observed for the *N*-glycopeptides with the highest content of fucoses (between 3 and 7). However, this is a preliminary study and

further analyses of a larger set of samples would be necessary to determine whether the proposed method is useful to reliably discriminate between CEA samples purified from human colon carcinoma and human liver metastases.

4. Concluding remarks

In this work, CEA *N*-glycosylation sites were characterized using *N*-glycopeptide profiles obtained after enzymatic digestion with trypsin, Glu-C and pronase and analysis by sheathless CE-MS/MS. Complementary information was obtained through the use of the three above mentioned enzymes, hence allowing an improved *N*-glycosylation site coverage and the identification of most of the potential *N*-glycosylation sites (20 out of 28), as well as their site-specific dominant *N*-glycan subtype (219 different *N*-glycopeptide glycoforms were identified). We have also demonstrated that CEA *N*-glycopeptide profiling combined with multivariate data analysis is a rapid, simple and excellent approach to identify differences on glycosylation between CEA samples purified from human colon carcinoma and human liver metastases. A hundred and twenty one *N*-glycopeptide glycoforms (out of the total of 219) were found to be the most significant to discriminate between the two CEA types, being

CEA samples purified from human liver metastases more fucosylated and sialylated. The present multienzyme multivariate sheathless CE-MS/MS bottom-up strategy shows a great potential to provide important biologic information on how *N*-glycosylation may influence CEA processing in cancer biogenesis. Furthermore, it may be applied to the characterization of other complex glycoproteins.

Acknowledgements

Laura Pont acknowledges the Spanish Ministry of Economy and Competitiveness for a FPI fellowship (BES-2012-061127) and a short stay fellowship (EEBB-I-16-11506). This study was supported by a grant from the Spanish Ministry of Education and Science (CTQ2014-56777-R), the European Union (Seventh Framework Programme HighGlycan and IBD-BIOM, Grant Agreements 278535 and 305479), a Horizon Programme Zenith project of The Netherlands Genomic Initiative (Project No. 93511033) and SCIEX.

References

(1) Adamczyk, B.; Tharmalingam, T.; Rudd, P. *M. Biochim. Biophys. Acta.* **2012**, *1820*, 1347–1353.

(2) Lau, K. S.; Dennis, J. W. *Glycobiology.* **2008**, *18*, 750–760.

(3) Zhang, Y.; Jiao, J.; Yang, P.; Lu, H. *Clin. Proteomics.* **2014**, *11*, 1–14.

(4) Freeze, H. H. *J. Biol. Chem.* **2013**, *288*, 6936–6944.

(5) Defaus, S.; Gupta, P.; Andreu, D.; Gutiérrez-Gallego, R. *Analyst.* **2014**, *139*, 2944–2967.

(6) Corfield, A. P.; Berry, M. *Trends Biochem. Sci.* **2015**, *40*, 351–359.

(7) Hammarström, S. *Semin. Cancer Biol.* **1999**, *9*, 67–81.

(8) Lopez, N. E.; Peterson, C. Y. *Clin. Colon Rectal Surg.* **2016**, *29*, 196–204.

(9) Okamura, K.; Takayama, K.; Izumi, M.; Harada, T.; Furuyama, K.; Nakanishi, Y. *Lung Cancer.* **2013**, *80*, 45–49.

(10) Becerra, A. Z.; Probst, C. P.; Tejani, M. A.; Aquina, C. T.; Hensley, B. J.; Noyes, K.; Monson, J. R.; Fleming, F. J. *Ann. Surg. Oncol.* **2016**, *23*, 1554–1561.

(11) Kannicht, C.; Lucka, L.; Nuck, R.; Reutter, W.; Gohlke, M. *Glycobiology.* **1999**, *9*, 897–906.

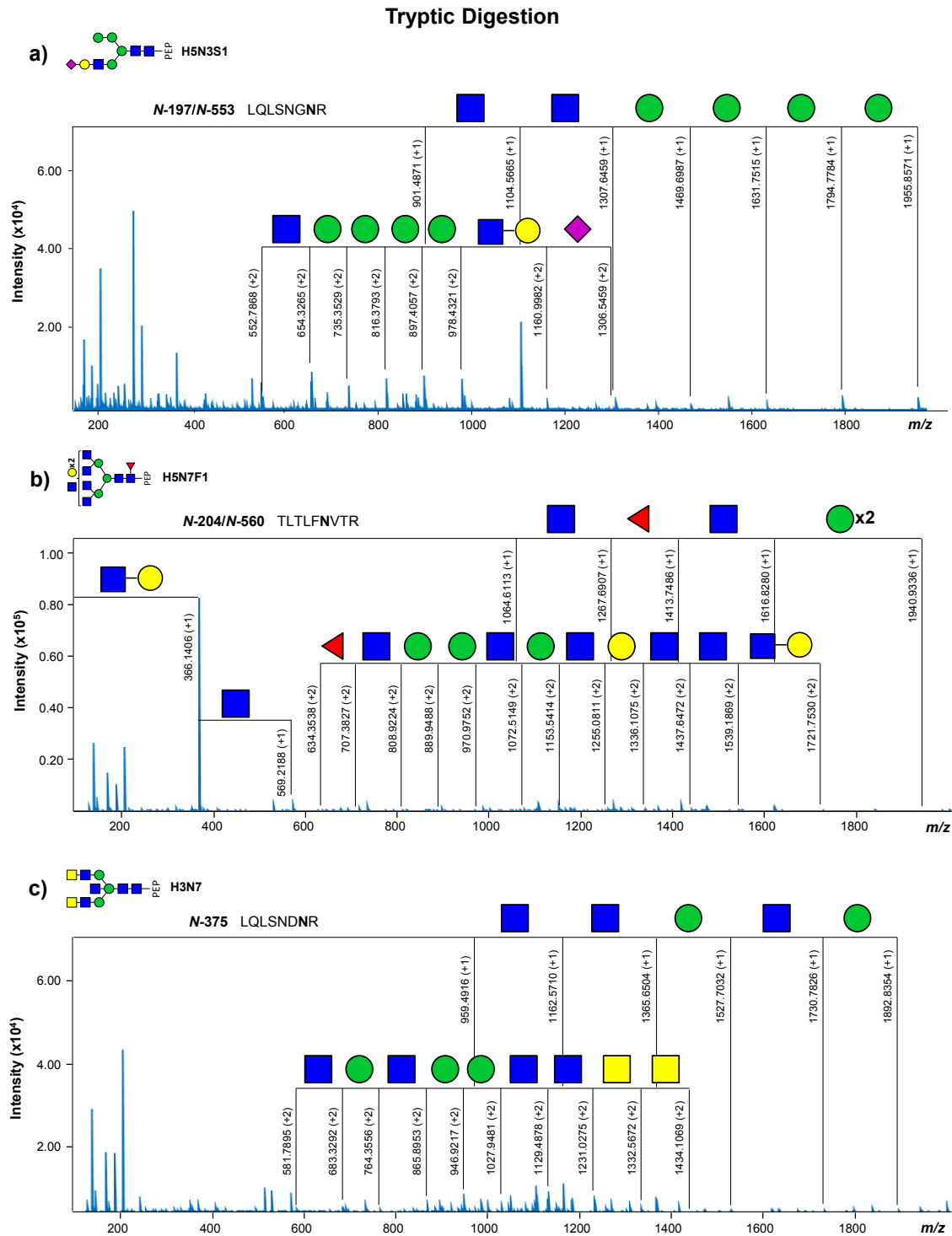
(12) Huang, C.; Zhan, T.; Liu, Y.; Li, Q.; Wu, H.; Ji, D.; Li, Y. *Clin. Proteomics.* **2015**, *12*, 1–7.

(13) Jeong, H. J.; Kim, Y. G.; Yang, Y. H.; Kim, B. G. *Anal. Chem.* **2012**, *84*, 3453–3460.

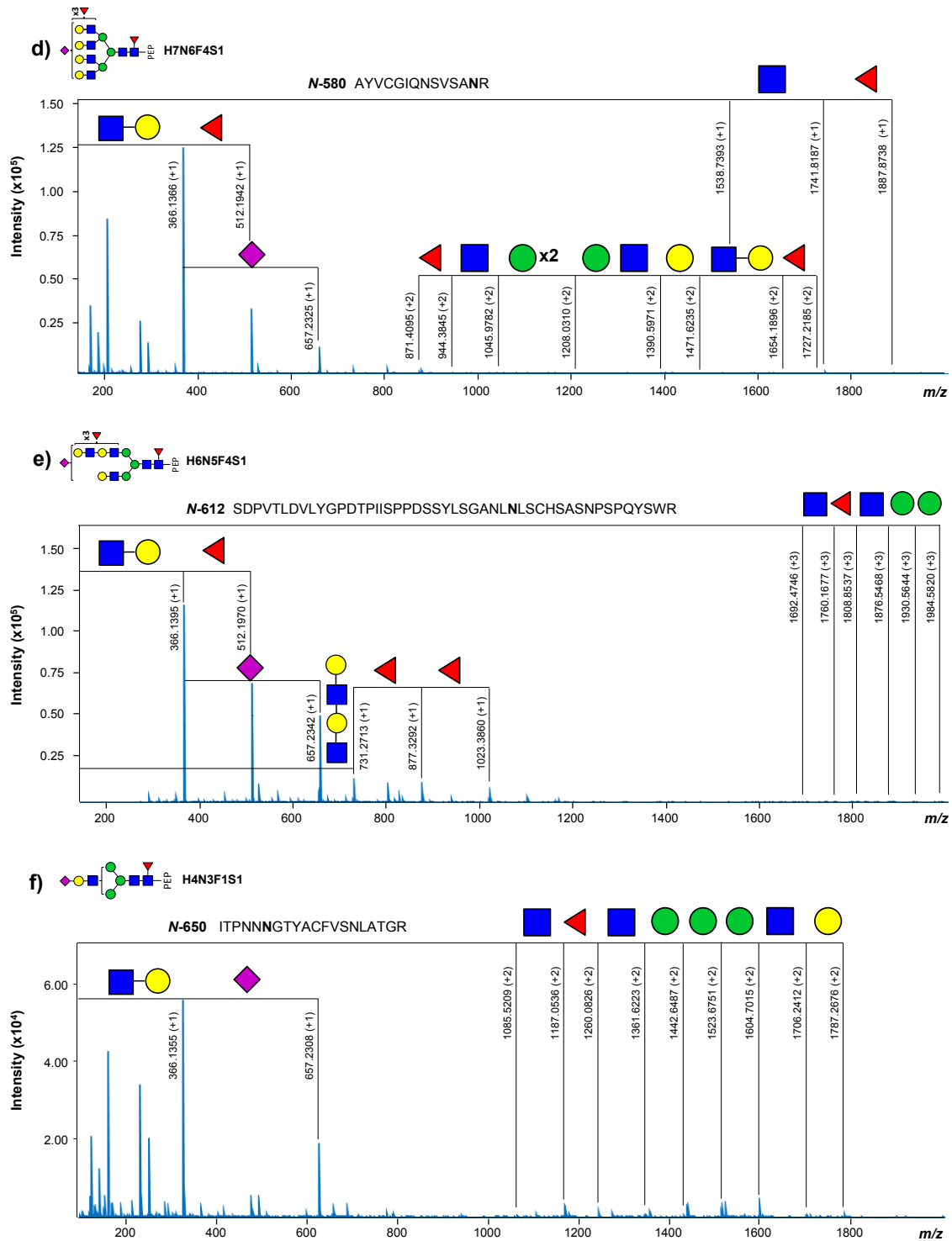
- (14) Wuhrer, M.; Catalina, M. I.; Deelder, A. M.; Hokke, C. H. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* **2007**, *849*, 115–128.
- (15) Wuhrer, M. *Glycoconj. J.* **2013**, *30*, 11–22.
- (16) Giménez, E.; Sanz-Nebot, V.; Rizzi, A. *Anal. Bioanal. Chem.* **2013**, *405*, 7307–7319.
- (17) Barroso, A.; Giménez, E.; Benavente, F.; Barbosa, J.; Sanz-Nebot, V. *Talanta*. **2016**, *160*, 614–623.
- (18) Štěpánová, S.; Kašička, V. *J. Sep. Sci.* **2016**, *39*, 198–211.
- (19) Heemskerck, A. A. M.; Deelder, A. M.; Mayboroda, O. A. *Mass Spectrom. Rev.* **2016**, *35*, 259–271.
- (20) Sun, L.; Zhu, G.; Yan, X.; Zhang, Z.; Wojcik, R.; Champion, M. M.; Dovichi, N. J. *Proteomics*. **2016**, *16*, 188–196.
- (21) Wenz, C.; Barbas, C.; Lopez-Gonzalvez, A.; Garcia, A.; Benavente, F.; Sanz-Nebot, V.; Blanc, T.; Freckleton, G.; Britz-McKibbin, P.; Shanmuganathan, M.; De L’Escaille, F.; Far, J.; Haselberg, R.; Huang, S.; Huhn, C.; Pattky, M.; Michels, D.; Mou, S.; Yang, F.; Neusüss, C.; Tromsdorf, N.; Baidoo, E. E. K.; Keasling, J. D.; Park, S. S. *J. Sep. Sci.* **2015**, *38*, 3262–3270.
- (22) Ramautar, R.; Busnel, J. M.; Deelder, A. M.; Mayboroda, O. A. *Anal. Chem.* **2012**, *84*, 885–892.
- (23) Faserl, K.; Sarg, B.; Kremser, L.; Lindner, H. *Anal. Chem.* **2011**, *83*, 7297–7305.
- (24) Moini, M. *Anal. Chem.* **2007**, *179*, 4241–4246.
- (25) Busnel, J. M.; Schoenmaker, B.; Ramautar, R.; Carrasco-Pancorbo, A.; Ratnayake, C.; Feitelson, J. S.; Chapman, J. D.; Deelder, A. M.; Mayboroda, O. A. *Anal. Chem.* **2010**, *82*, 9476–9483.
- (26) Kammeijer, G. S. M.; Kohler, I.; Jansen, B. C.; Hensbergen, P. J.; Mayboroda, O. A.; Falck, D.; Wuhrer, M. *Anal. Chem.* **2016**, *88*, 5849–5856.
- (27) Joliffe, I. T.; Morgan, B. J. *Stat. Methods Med. Res.* **1992**, *1*, 69–95.
- (28) Barker, M.; Rayens, W. *J. Chemom.* **2003**, *17*, 166–173.
- (29) Wold, S.; Sjöström, M.; Eriksson, L. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109–130.
- (30) Palermo, G.; Piraino, P.; Zucht, H. D. *Adv. Appl. Bioinforma. Chem.* **2009**, *2*, 57–70.
- (31) Cawley, G. C.; Talbot, N. L. C. *Pattern Recognit.* **2003**, *36*, 2585–2592.
- (32) Tsiatsiani, L.; Heck, A. J. R. *FEBS J.* **2015**, *282*, 2612–2626.
- (33) Giansanti, P.; Tsiatsiani, L.; Low, T. Y.; Heck, A. J. R. *Nat. Protoc.* **2016**, *11*, 993–1006.
- (34) Dalpathado, D. S.; Desaire, H. *Analyst.* **2008**, *133*, 731–738.
- (35) Zauner, G.; Koeleman, C. A. M.; Deelder, A. M.; Wuhrer, M. *J. Sep. Sci.* **2010**, *33*, 903–910.

- (36) Temporini, C.; Perani, E.; Calleri, E.; Dolcini, L.; Lubda, D.; Caccialanza, G.; Massolini, G. *Anal. Chem.* **2007**, *79*, 355–363.
- (37) Gattiker, A.; Bienvenut, W. V.; Bairoch, A.; Gasteiger, E. *Proteomics*. **2002**, *2*, 1435–1444.
- (38) Giménez, E.; Ramos-Hernan, R.; Benavente, F.; Barbosa, J.; Sanz-Nebot, V. *Rapid Commun. Mass Spectrom.* **2011**, *25*, 2307–2316.
- (39) Barroso, A.; Giménez, E.; Benavente, F.; Barbosa, J.; Sanz-Nebot, V. *Anal. Chim. Acta.* **2015**, *854*, 169–177.
- (40) Li, M.; Song, L.; Qin, X. *J. Biosci.* **2010**, *35*, 665–673.
- (41) Pinho, S. S.; Reis, C. A. *Nat Rev Cancer.* **2015**, *15*, 540–555.
- (42) Plucinsky, M. C.; Riley, W. M.; Prorok, J. J.; Alhadeff, J. A. *Cancer.* **1986**, *58*, 2680–2685.
- (43) Vajaria, B. N.; Patel, K. R.; Begum, R.; Patel, P. S. *Pathol. Oncol. Res.* **2016**, *22*, 443–447.
- (44) Miyoshi, E.; Moriwaki, K.; Naka, J. *Biochem.* **2008**, *143*, 725–729.
- (45) Miyoshi, E.; Moriwaki, K.; Terao, N.; Tan, C. C.; Terao, M.; Nakagawa, T.; Matsumoto, H.; Shinzaki, S.; Kamada, Y. *Biomolecules.* **2012**, *30*, 34–45.

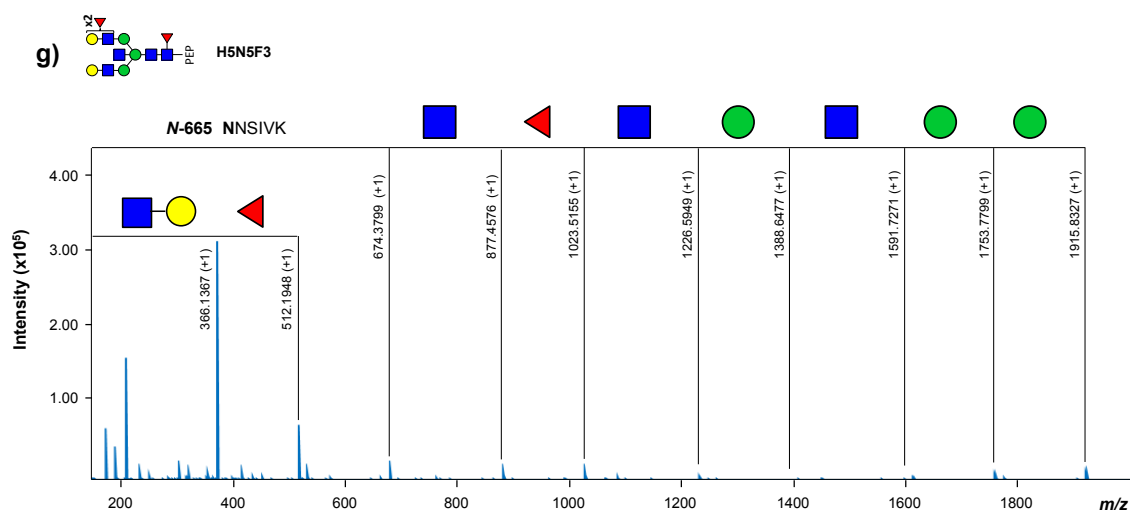
Supplementary Figures and Tables



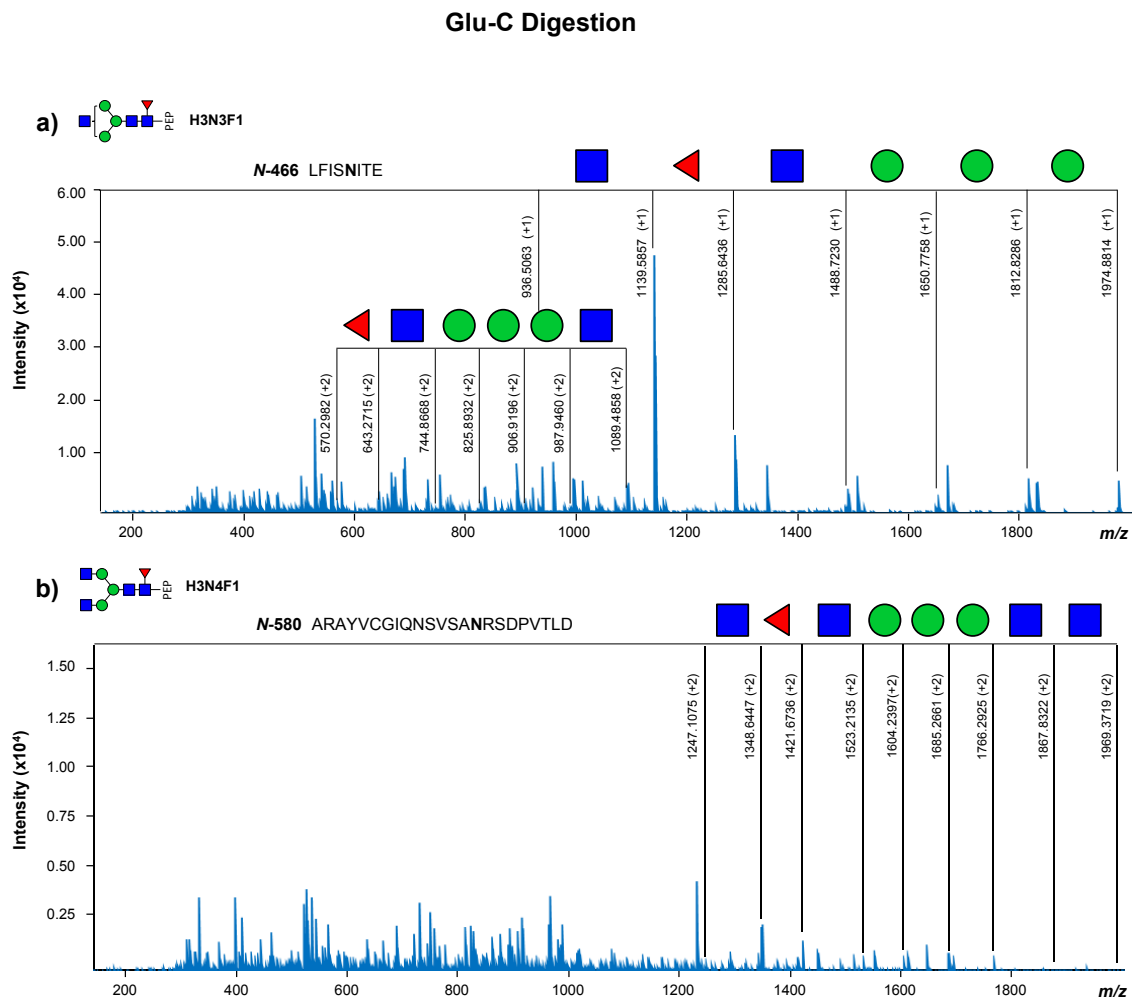
Supplementary Figure S-1-A. MS/MS fragmentation spectra (with experimental m/z) of some of the CEA *N*-glycopeptides that show differentiation between CEA samples from colon carcinoma and liver metastases after digestion with trypsin (a: H5N3S1 (CEA3), b: H5N7F1 (CEA1), c: H3N7 (CEA1)).



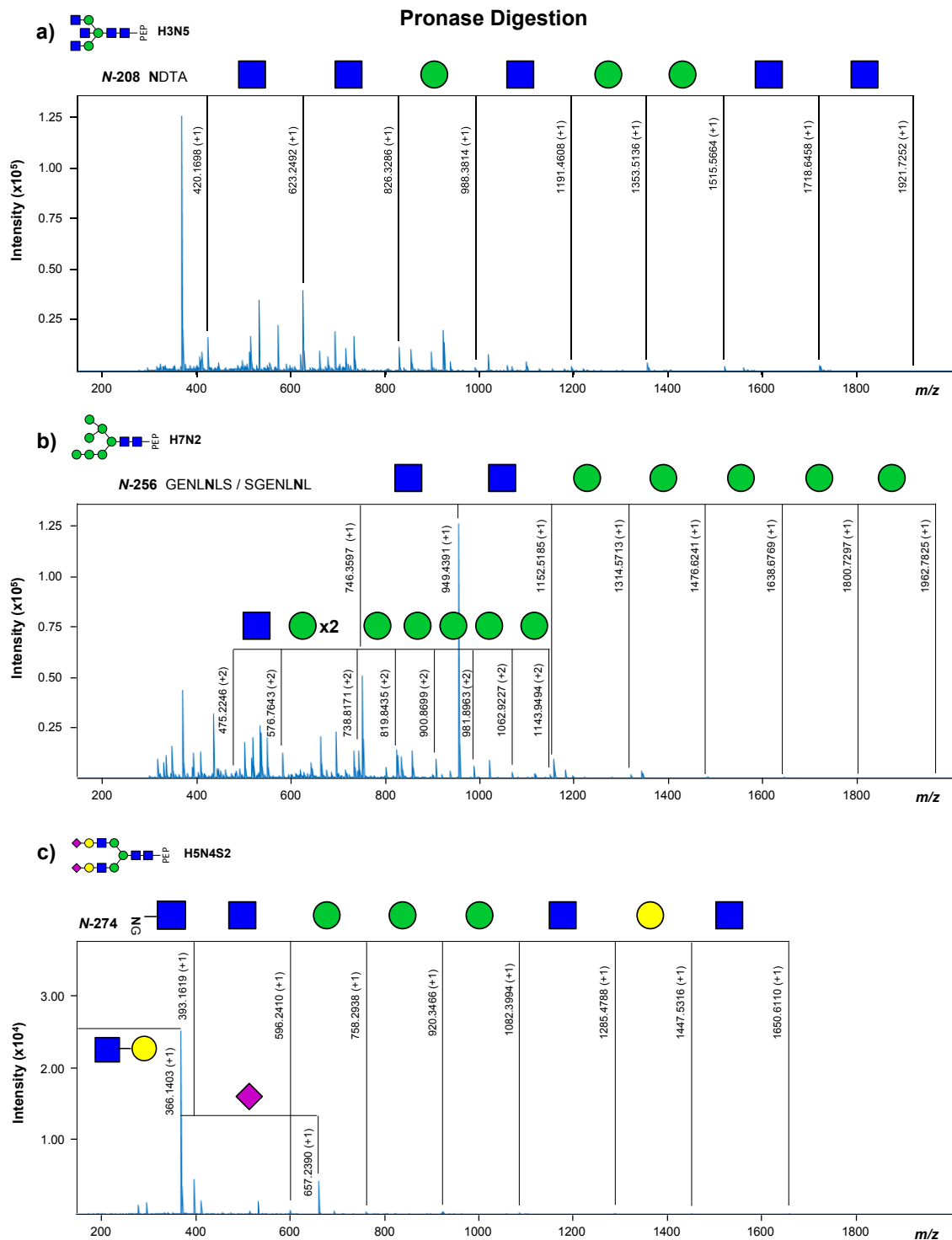
Supplementary Figure S-1-A (continue). MS/MS fragmentation spectra (with experimental m/z) of some of the CEA N-glycopeptides that show differentiation between CEA samples from colon carcinoma and liver metastases after digestion with trypsin (d: H7N6F4S1 (CEA3), e: H6N5F4S1 (CEA3), f: H4N3F1S1 (CEA3).



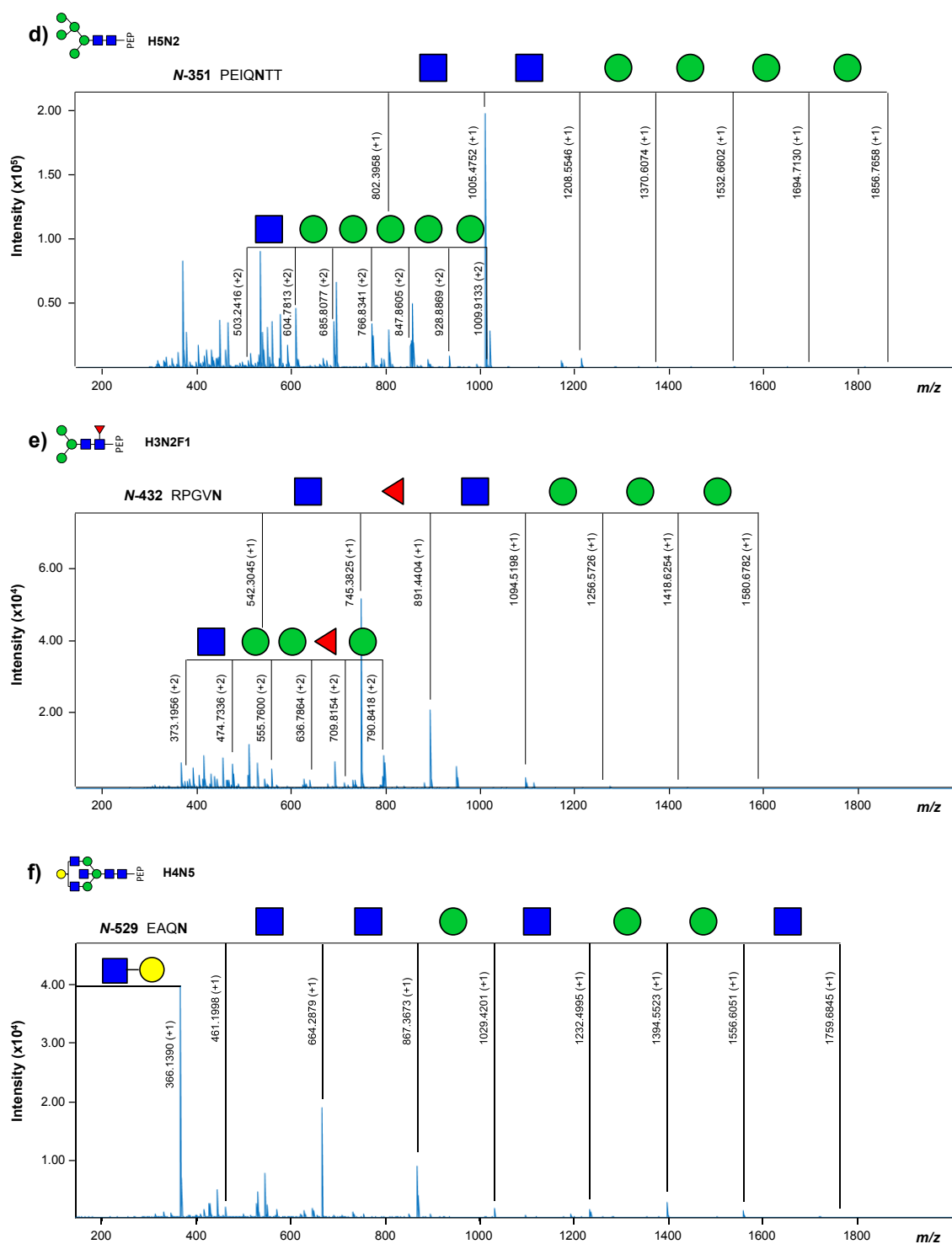
Supplementary Figure S-1-A (continue). MS/MS fragmentation spectra (with experimental m/z) of some of the CEA *N*-glycopeptides that show differentiation between CEA samples from colon carcinoma and liver metastases after digestion with trypsin (g: H5N5F3 (CEA3)).



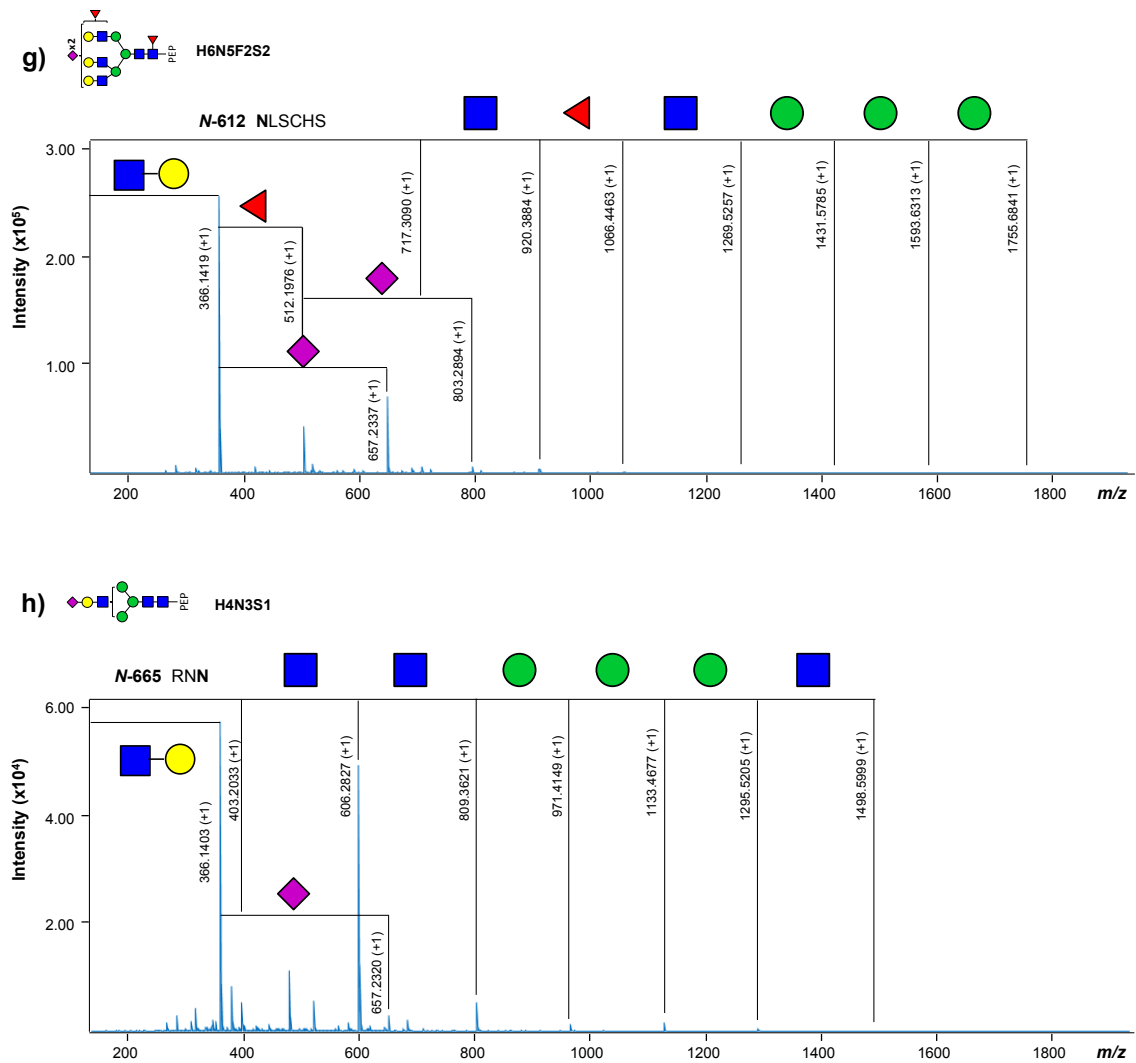
Supplementary Figure S-1-B. MS/MS fragmentation spectra (with experimental m/z) of some of the CEA *N*-glycopeptides that show differentiation between CEA samples from colon carcinoma and liver metastases after digestion with Glu-C (a: H3N3F1 (CEA1), b: H3N4F1 (CEA1)).



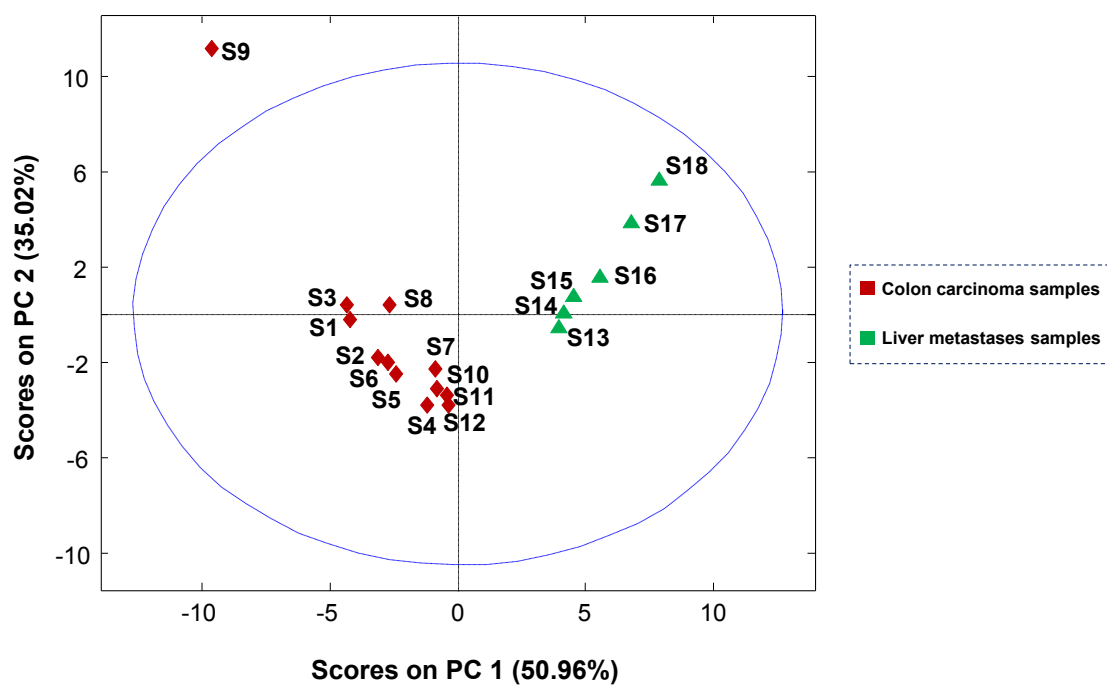
Supplementary Figure S-1-C. MS/MS fragmentation spectra (with experimental m/z) of some of the CEA *N*-glycopeptides that show differentiation between CEA samples from colon carcinoma and liver metastases after digestion with pronase (a: H3N5 (CEA3), b: H7N2 (CEA1), c: H5N4S2 (CEA3)).



Supplementary Figure S-1-C (continue). MS/MS fragmentation spectra (with experimental m/z) of some of the CEA N -glycopeptides that show differentiation between CEA samples from colon carcinoma and liver metastases after digestion with pronase (d: H5N2 (CEA3), e: H3N2F1 (CEA2), f: H4N5 (CEA2)).

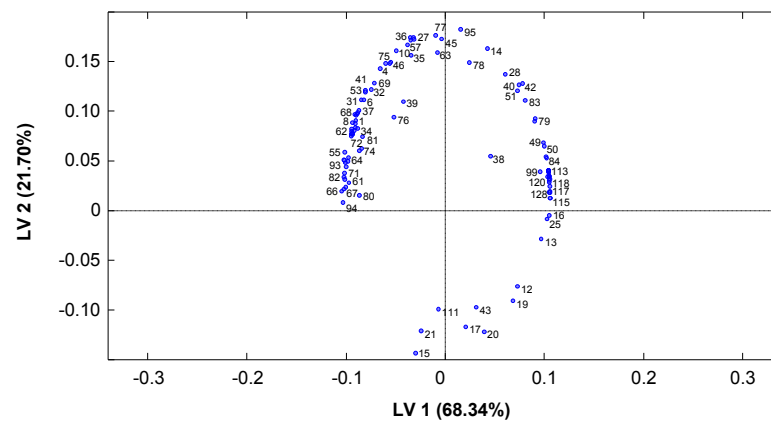


Supplementary Figure S-1-C (continue). MS/MS fragmentation spectra (with experimental m/z) of some of the CEA *N*-glycopeptides that show differentiation between CEA samples from colon carcinoma and liver metastases after digestion with pronase (g: H6N5F2S2 (CEA3), h: H4N3S1 (CEA3)).

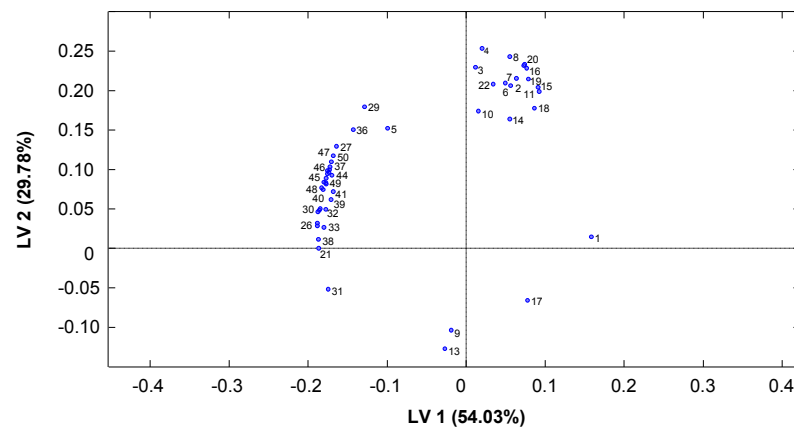


Supplementary Figure S-2. PCA scores plot for the model of the CEA samples after digestion with pronase. Sample 9 (S9) was detected as an outlier and, therefore, it was discarded for further analysis.

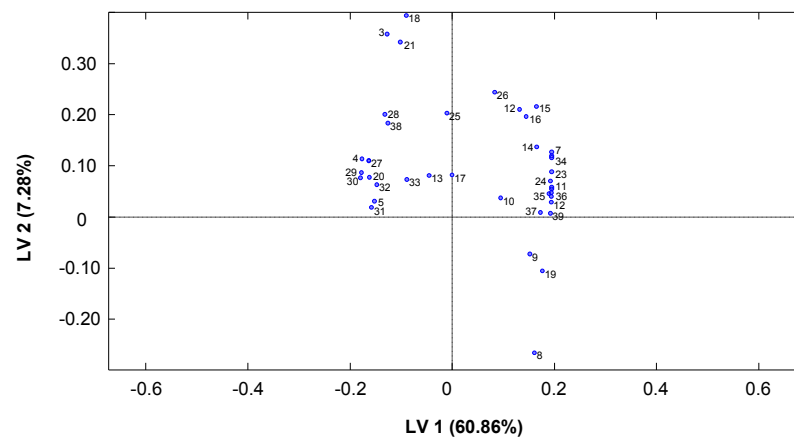
A) Tryptic digestion



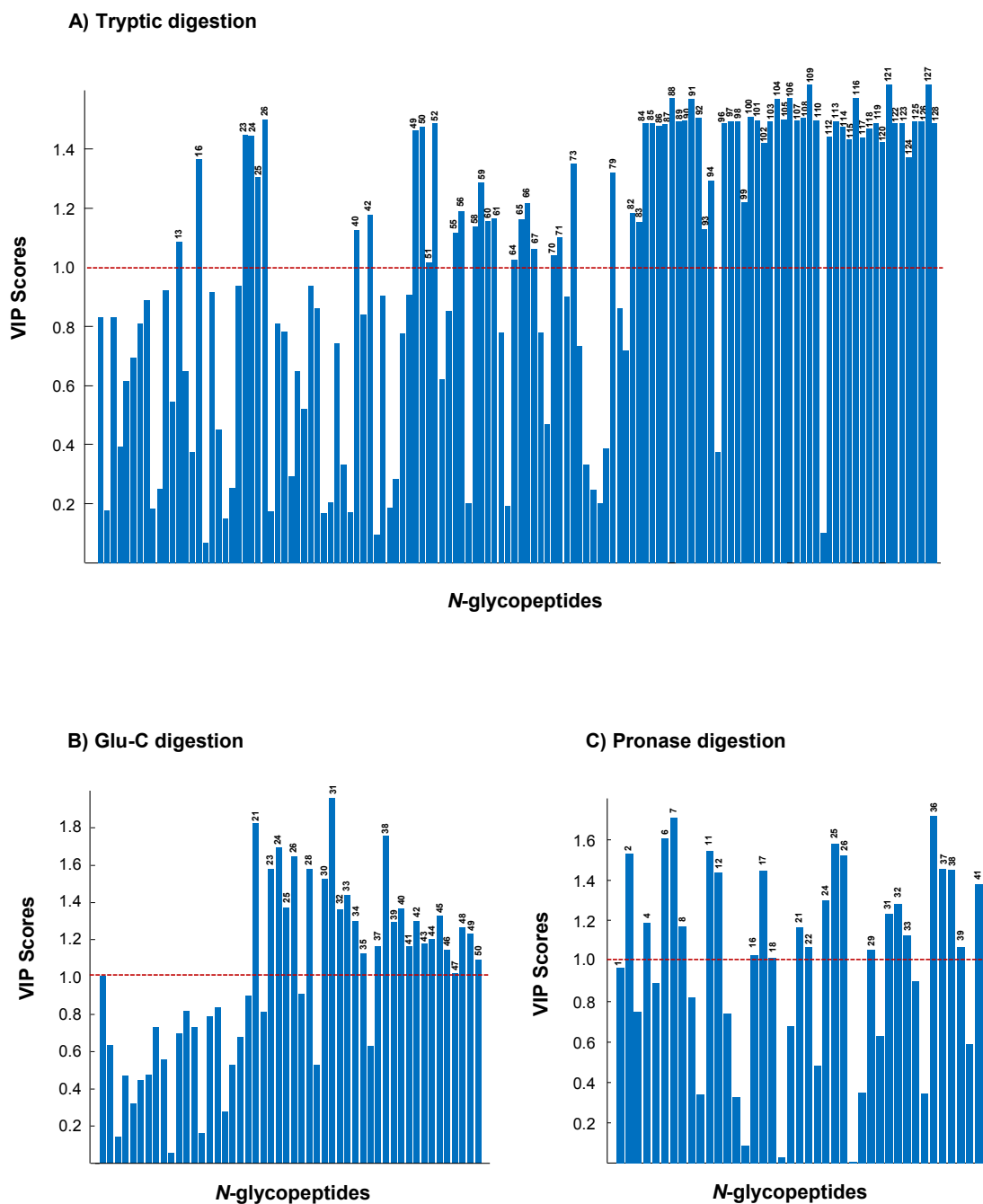
B) Glu-C digestion



C) Pronase digestion






















Supplementary Figure S-3. PLS-DA loadings plots for the classification of the CEA samples after digestion with (A) trypsin, (B) Glu-C and (C) pronase. Variable numbers are referred to the list of identified *N*-glycopeptides (Num) reported in Supplementary Tables S-1 (trypsin), S-2 (Glu-C) and S-3-A (pronase).



Supplementary Figure S-4. PLS-DA model VIP scores values for the *N*-glycopeptides identified after digestion with (A) trypsin, (B) Glu-C and (C) pronase. VIP scores values higher than 1 were used as a feature selection tool to choose only the most relevant candidate *N*-glycopeptides for each PLS-DA model (71 out of 128 for trypsin digestion, 26 out of 50 for Glu-C digestion and 24 out of 41 for pronase digestion). Bar numbers are referred to the list of identified *N*-glycopeptides (Num) reported in Supplementary Tables S-1 (trypsin), S-2 (Glu-C) and S-3-A (pronase).

Supplementary Table S-1. Observed average m/z ($n=6$), relative error and %relative abundances (normalizing the integrated peak areas by the total integrated peak areas for the N -glycosylation site) for the CEA N -glycopeptides identified by sheathless CE-MS/MS after digestion with trypsin.





















N-197/N-553 Peptide: LQLSNGNR																	
		CEA1 Human Colon Carcinoma						CEA2 Human Colon Carcinoma					CEA3 Human Liver Metastases				
Num ^a	Glycan Species ^b		Molecular ion	Observed m/z	Error (ppm) ^c	%Relative Abundance	MS/MS ^d	Molecular ion	Observed m/z	Error (ppm) ^c	%Relative Abundance	MS/MS ^d	Molecular ion	Observed m/z	Error (ppm) ^c	%Relative Abundance	MS/MS ^d
1	H3N2		[M+2H] ²⁺	897.4060	1.0	0.9	Yes	[M+2H] ²⁺	897.3996	6.1	1.3	No					
2	H4N2		[M+2H] ²⁺	978.4297	1.8	1.5	Yes	[M+2H] ²⁺	978.4227	9.0	2.1	Yes	[M+2H] ²⁺	978.4280	3.6	2.5	Yes
3	H3N3		[M+2H] ²⁺	998.9424	2.4	2.2	Yes	[M+2H] ²⁺	998.9362	8.6	3.1	Yes					
4	H5N2		[M+2H] ²⁺	1059.4553	2.5	41.4	Yes	[M+2H] ²⁺	1059.4491	8.3	59.8	Yes	[M+2H] ²⁺	1059.4537	4.0	42.1	Yes
5	H4N3		[M+2H] ²⁺	1079.9694	1.6	1.8	Yes	[M+2H] ²⁺	1079.9612	9.2	2.3	No	[M+2H] ²⁺	1079.9748	3.4	1.1	No
6	H6N2		[M+2H] ²⁺	1140.4822	1.9	4.3	Yes	[M+2H] ²⁺	1140.4730	9.9	4.2	No	[M+2H] ²⁺	1140.4783	5.3	2.3	Yes
7	H3N5		[M+3H] ³⁺	801.6901	5.9	3.2	Yes	[M+3H] ³⁺	801.6810	5.5	3.0	No	[M+3H] ³⁺	801.6835	2.4	1.1	No
8	H5N3		[M+2H] ²⁺	1160.9953	2.0	3.0	Yes	[M+2H] ²⁺	1160.9872	8.9	3.3	No					
9	H4N4		[M+3H] ³⁺	788.0081	2.3	3.0	No	[M+3H] ³⁺	788.0060	4.9	2.5	No	[M+3H] ³⁺	788.0070	3.7	3.6	Yes
10	H5N4		[M+3H] ³⁺	842.0263	1.4	6.1	Yes	[M+3H] ³⁺	842.0227	5.7	5.1	No	[M+3H] ³⁺	842.0243	3.8	6.6	Yes
11	H4N5		[M+3H] ³⁺	855.7021	1.1	4.3	Yes	[M+3H] ³⁺	855.6976	6.3	4.1	Yes	[M+3H] ³⁺	855.7028	0.2	1.2	No
12	H7N2		[M+2H] ²⁺	1221.5117	0.8	1.0	No	[M+2H] ²⁺	1221.5011	7.9	0.2	No	[M+2H] ²⁺	1221.5046	5.0	1.1	No
13	H5N4F1		[M+3H] ³⁺	890.7136	1.3	1.5	Yes	[M+3H] ³⁺	890.7062	8.2	0.1	No	[M+3H] ³⁺	890.7104	3.4	2.9	No
14	H6N4		[M+3H] ³⁺	896.0438	1.5	3.4	Yes	[M+3H] ³⁺	896.0381	7.8	2.6	No	[M+3H] ³⁺	896.0417	3.8	5.8	Yes
15	H4N6		[M+3H] ³⁺	923.3949	0.2	3.8	Yes	[M+3H] ³⁺	923.3879	8.9	0.6	No	[M+3H] ³⁺	923.3893	7.4	1.3	No
16	H6N3F1		[M+3H] ³⁺	877.0378	0.2	1.2	No	[M+3H] ³⁺	877.0377	0.3	0.1	No	[M+3H] ³⁺	877.0350	3.4	3.9	Yes
17	H5N5		[M+3H] ³⁺	909.7192	1.6	2.0	No	[M+3H] ³⁺	909.7124	9.0	0.4	No	[M+3H] ³⁺	909.7176	3.3	1.5	No
18	H5N6		[M+3H] ³⁺	977.4088	5.1	2.6	Yes	[M+3H] ³⁺	977.4048	9.2	2.0	No	[M+3H] ³⁺	977.4104	3.4	0.8	No
19	H5N5F2		[M+3H] ³⁺	1007.0906	1.9	2.7	No	[M+3H] ³⁺	1007.0841	8.4	0.3	No	[M+3H] ³⁺	1007.0873	5.2	2.9	No














20	H5N6F2		$[M+3H]^{3+}$	1074.7827	2.8	2.4	No	$[M+3H]^{3+}$	1074.7807	4.6	0.2	No	$[M+3H]^{3+}$	1074.7809	4.5	1.8	No
21	H4N7F2		$[M+3H]^{3+}$	1088.4570	3.9	1.3	No	$[M+3H]^{3+}$	1088.4514	9.0	0.2	No	$[M+3H]^{3+}$	1088.4567	4.1	0.6	No
22	H5N7F2		$[M+3H]^{3+}$	1142.4729	5.2	2.7	Yes	$[M+3H]^{3+}$	1142.4691	8.5	2.1	No	$[M+3H]^{3+}$	1142.4691	8.5	0.7	No
23	H4N3S1		$[M+3H]^{3+}$	817.3416	8.5	1.1	No	$[M+3H]^{3+}$	817.3498	1.5	0.1	No	$[M+3H]^{3+}$	817.3433	6.4	5.7	No
24	H5N3S1		$[M+3H]^{3+}$	871.3652	1.1	0.8	No	$[M+3H]^{3+}$	871.3608	6.5	0.1	No	$[M+3H]^{3+}$	871.3625	4.2	3.7	Yes
25	H6N3S1		$[M+3H]^{3+}$	925.3827	1.2	1.8	No	$[M+3H]^{3+}$	925.3755	8.9	0.2	No	$[M+3H]^{3+}$	925.3790	5.2	4.5	No
26	H6N3F1S1												$[M+3H]^{3+}$	974.0643	5.6	2.4	No



N-197/N-553 Deamidated Peptide: LQLSNGNR																	
		CEA1 Human Colon Carcinoma						CEA2 Human Colon Carcinoma					CEA3 Human Liver Metastases				
Num ^a	Glycan Species ^b		Molecular ion	Observed m/z	Error (ppm) ^c	%Relative Abundance	MS/MS ^d	Molecular ion	Observed m/z	Error (ppm) ^c	%Relative Abundance	MS/MS ^d	Molecular ion	Observed m/z	Error (ppm) ^c	%Relative Abundance	MS/MS ^d
27	H3N2		[M+2H] ²⁺	897.9084	0.7	1.0	Yes	[M+2H] ²⁺	897.9050	4.5	1.1	Yes	[M+2H] ²⁺	897.9019	8.0	1.3	No
28	H4N2		[M+2H] ²⁺	978.9323	3.2	1.8	Yes	[M+2H] ²⁺	978.9260	9.7	1.8	Yes	[M+2H] ²⁺	978.9300	5.6	4.4	Yes
29	H3N3		[M+2H] ²⁺	999.4455	3.2	2.2	Yes	[M+2H] ²⁺	999.4397	9.0	2.6	Yes	[M+2H] ²⁺	999.4458	2.9	0.3	No
30	H5N2		[M+2H] ²⁺	1059.9687	6.4	47.2	Yes	[M+2H] ²⁺	1059.9552	6.3	54.0	Yes	[M+2H] ²⁺	1059.9571	4.5	47.2	Yes
31	H4N3		[M+2H] ²⁺	1080.4643	10.0	1.7	Yes	[M+2H] ²⁺	1080.4654	9.0	1.8	Yes	[M+2H] ²⁺	1080.4652	9.2	0.6	No
32	H6N2		[M+2H] ²⁺	1140.9990	9.4	3.6	Yes	[M+2H] ²⁺	1140.9781	8.9	3.8	Yes	[M+2H] ²⁺	1140.9780	9.0	2.0	No
33	H3N5		[M+3H] ³⁺	802.0154	7.5	3.2	Yes	[M+3H] ³⁺	802.0136	9.7	2.6	Yes					
34	H5N3		[M+2H] ²⁺	1161.5032	1.4	2.8	Yes	[M+2H] ²⁺	1161.4902	9.8	2.5	Yes	[M+2H] ²⁺	1161.4990	2.2	0.3	No
35	H4N4		[M+3H] ³⁺	788.3395	8.1	3.1	No	[M+3H] ³⁺	788.3444	1.9	2.1	No	[M+3H] ³⁺	788.3402	7.2	2.9	Yes
36	H5N4		[M+3H] ³⁺	842.3550	10.0	5.9	Yes	[M+3H] ³⁺	842.3559	9.0	4.8	Yes	[M+3H] ³⁺	842.3608	3.2	6.4	Yes
37	H4N5		[M+3H] ³⁺	856.0320	8.2	4.1	Yes	[M+3H] ³⁺	856.0353	4.3	3.7	Yes	[M+3H] ³⁺	856.0321	8.1	1.1	No
38	H7N2		[M+2H] ²⁺	1222.0032	9.4	0.6	No	[M+2H] ²⁺	1222.0052	7.8	0.2	No	[M+2H] ²⁺	1222.0057	7.4	0.6	No
39	H5N4F1		[M+3H] ³⁺	891.0450	5.0	0.8	No	[M+3H] ³⁺	891.0469	2.9	1.0	No	[M+3H] ³⁺	891.0467	3.1	0.9	Yes



40	H6N4		[M+3H] ³⁺	896.3785	2.9	3.1	Yes	[M+3H] ³⁺	896.3760	5.7	1.9	Yes	[M+3H] ³⁺	896.3735	8.5	6.0	Yes
41	H4N6		[M+3H] ³⁺	923.7259	6.7	2.9	Yes	[M+3H] ³⁺	923.7265	6.1	3.3	Yes					
42	H6N3F1		[M+3H] ³⁺	877.3665	8.5	0.9	No	[M+3H] ³⁺	877.3703	4.1	1.0	No	[M+3H] ³⁺	877.3631	6.7	3.4	Yes
43	H5N5		[M+3H] ³⁺	910.0490	8.4	2.0	No	[M+3H] ³⁺	910.0521	4.9	0.5	Yes	[M+3H] ³⁺	910.0475	10.0	1.4	No
44	H5N6		[M+3H] ³⁺	977.7460	3.8	2.4	Yes	[M+3H] ³⁺	977.7426	7.3	1.8	Yes					
45	H5N5F2		[M+3H] ³⁺	1007.4186	9.9	2.1	Yes	[M+3H] ³⁺	1007.4216	6.9	2.1	No	[M+3H] ³⁺	1007.4190	9.5	3.2	Yes
46	H5N6F2		[M+3H] ³⁺	1075.1116	9.4	1.8	No	[M+3H] ³⁺	1075.1138	7.3	2.0	No	[M+3H] ³⁺	1075.1199	1.6	1.8	Yes
47	H4N7F2		[M+3H] ³⁺	1088.7871	9.3	1.5	Yes	[M+3H] ³⁺	1088.7883	8.2	0.8	No					
48	H5N7F2		[M+3H] ³⁺	1142.8032	10.0	2.0	Yes	[M+3H] ³⁺	1142.8050	8.6	1.7	No					
49	H4N3S1		[M+3H] ³⁺	817.6804	5.1	1.1	No	[M+3H] ³⁺	817.6807	4.7	1.0	No	[M+3H] ³⁺	817.6795	6.2	6.7	Yes
50	H5N3S1		[M+3H] ³⁺	871.6951	8.1	0.6	No	[M+3H] ³⁺	871.6978	5.0	0.5	No	[M+3H] ³⁺	871.6962	6.8	3.6	Yes
51	H6N3S1		[M+3H] ³⁺	925.7108	9.7	1.6	No	[M+3H] ³⁺	925.7133	7.0	1.3	Yes	[M+3H] ³⁺	925.7111	9.4	3.8	Yes
52	H6N3F1S1		[M+3H] ³⁺	974.4039	1.9	0.05	No	[M+3H] ³⁺	974.3998	6.1	0.1	No	[M+3H] ³⁺	974.3932	6.7	2.2	Yes


N-204/N-560 Peptide: TLTLFNVTR																	
		CEA1 Human Colon Carcinoma					CEA2 Human Colon Carcinoma					CEA3 Human Liver Metastases					
Num ^a	Glycan Species ^b	Molecular ion	Observed <i>m/z</i>	Error (ppm) ^c	%Relative Abundance	MS/MS ^d	Molecular ion	Observed <i>m/z</i>	Error (ppm) ^c	%Relative Abundance	MS/MS ^d	Molecular ion	Observed <i>m/z</i>	Error (ppm) ^c	%Relative Abundance	MS/MS ^d	
53	H3N2F1		[M+2H] ²⁺	1052.0051	8.2	1.2	Yes	[M+2H] ²⁺	1051.9891	7.0	2.2	Yes					
54	H3N3F1		[M+2H] ²⁺	1153.5334	2.4	3.5	Yes	[M+2H] ²⁺	1153.5293	5.9	6.6	Yes					
55	H3N4F1		[M+2H] ²⁺	1255.0698	4.8	5.3	Yes	[M+2H] ²⁺	1255.0671	6.9	6.2	No					
56	H3N5F1		[M+3H] ³⁺	904.7419	4.9	6.1	Yes	[M+3H] ³⁺	904.7396	7.4	5.3	Yes					
57	H4N4F1		[M+3H] ³⁺	891.0648	7.0	3.6	Yes	[M+3H] ³⁺	891.0647	6.8	3.5	No	[M+3H] ³⁺	891.0698	1.1	4.6	No
58	H4N5F1		[M+3H] ³⁺	958.7605	3.6	6.2	Yes	[M+3H] ³⁺	958.7592	4.9	7.5	Yes					
59	H3N6F1		[M+3H] ³⁺	972.4341	5.5	3.2	Yes	[M+3H] ³⁺	972.4299	9.8	3.0	No					



60	H5N5F1		$[M+3H]^{3+}$	1012.7804	1.1	4.9	Yes	$[M+3H]^{3+}$	1012.7734	8.0	5.2	No					
61	H4N6F1		$[M+3H]^{3+}$	1026.4506	6.2	3.6	Yes	$[M+3H]^{3+}$	1026.4474	9.4	3.2	No					
62	H5N5F2		$[M+3H]^{3+}$	1061.4622	5.0	4.0	Yes	$[M+3H]^{3+}$	1061.4648	2.5	3.7	No					
63	H5N5F3		$[M+3H]^{3+}$	1110.1481	4.8	3.0	Yes	$[M+3H]^{3+}$	1110.1509	2.3	2.5	No	$[M+3H]^{3+}$	1110.1488	4.2	4.1	Yes
64	H5N7F1		$[M+3H]^{3+}$	1148.1610	5.9	2.1	Yes	$[M+3H]^{3+}$	1148.1635	3.7	2.2	No					
65	H5N6F2		$[M+3H]^{3+}$	1129.1515	8.1	3.8	Yes	$[M+3H]^{3+}$	1129.1514	8.2	3.1	No					
66	H5N6F3		$[M+3H]^{3+}$	1177.8355	9.4	3.1	Yes	$[M+3H]^{3+}$	1177.8362	8.8	2.4	No					
67	H5N7F2		$[M+3H]^{3+}$	1196.8467	5.9	2.6	Yes	$[M+3H]^{3+}$	1196.8471	5.5	2.1	No					
68	H6N6F2		$[M+3H]^{3+}$	1183.1720	5.3	2.7	Yes	$[M+3H]^{3+}$	1183.1718	5.4	3.1	No					
69	H6N6F3		$[M+3H]^{3+}$	1231.8596	3.7	3.3	Yes	$[M+3H]^{3+}$	1231.8683	3.3	3.7	No					
70	H5N7F3		$[M+3H]^{3+}$	1245.5333	5.1	2.6	Yes	$[M+3H]^{3+}$	1245.5330	5.4	2.4	No					
71	H6N7F2		$[M+3H]^{3+}$	1250.8635	6.3	3.1	Yes	$[M+3H]^{3+}$	1250.8619	7.6	2.9	No					
72	H6N7F3		$[M+3H]^{3+}$	1299.5497	5.9	3.9	Yes	$[M+3H]^{3+}$	1299.5457	8.9	3.9	Yes					
73	H6N6F5		$[M+3H]^{3+}$	1329.2308	4.0	1.7	No	$[M+3H]^{3+}$	1329.2289	5.4	1.4	No	$[M+3H]^{3+}$	1329.2284	5.8	4.5	Yes
74	H6N7F4		$[M+3H]^{3+}$	1348.2359	5.5	3.6	Yes	$[M+3H]^{3+}$	1348.2339	7.0	2.6	No					
75	H7N7F3		$[M+3H]^{3+}$	1353.5711	2.8	2.9	Yes	$[M+3H]^{3+}$	1353.5672	5.7	3.2	Yes	$[M+3H]^{3+}$	1353.5699	3.7	3.2	Yes
76	H6N7F5		$[M+3H]^{3+}$	1396.9250	3.0	2.4	Yes	$[M+3H]^{3+}$	1396.9200	6.6	1.7	No					
77	H7N7F4		$[M+3H]^{3+}$	1402.2572	2.6	2.8	Yes	$[M+3H]^{3+}$	1402.2535	5.3	2.9	No	$[M+3H]^{3+}$	1402.2567	3.0	4.4	Yes
78	H7N7F5		$[M+3H]^{3+}$	1450.9426	2.9	2.5	Yes	$[M+3H]^{3+}$	1450.9411	4.0	2.4	No					
79	H7N7F6		$[M+3H]^{3+}$	1499.6263	4.4	1.6	No	$[M+3H]^{3+}$	1499.6215	7.6	1.3	No	$[M+3H]^{3+}$	1499.6270	3.9	4.5	Yes





80	H6N8F5		$[M+3H]^{3+}$	1464.6118	7.2	1.7	Yes	$[M+3H]^{3+}$	1464.6114	7.5	1.0	No					
81	H7N8F5		$[M+3H]^{3+}$	1518.6335	4.3	1.7	Yes	$[M+3H]^{3+}$	1518.6270	8.6	1.5	No					
82	H4N4F1S1		$[M+3H]^{3+}$	988.0987	3.9	2.6	Yes	$[M+3H]^{3+}$	988.1011	1.5	2.7	No					
83	H5N5F2S1		$[M+3H]^{3+}$	1158.4935	5.0	1.7	No	$[M+3H]^{3+}$	1158.4941	4.5	2.4	Yes	$[M+3H]^{3+}$	1158.4943	4.3	6.5	Yes
84	H7N6F3S1		$[M+3H]^{3+}$	1382.9111	1.8	0.9	No	$[M+3H]^{3+}$	1382.9025	8.0	1.0	No	$[M+3H]^{3+}$	1382.9077	4.2	8.1	Yes
85	H7N6F4S1		$[M+3H]^{3+}$	1431.5932	4.5	0.9	No	$[M+3H]^{3+}$	1431.5903	6.5	0.6	No	$[M+3H]^{3+}$	1431.5914	5.7	10.5	Yes
													$[M+4H]^{4+}$	1073.9449	6.3		Yes
86	H7N6F5S1		$[M+3H]^{3+}$	1480.2820	2.4	0.7	No	$[M+3H]^{3+}$	1480.2724	8.9	0.5	No	$[M+3H]^{3+}$	1480.2779	5.2	8.8	Yes
87	H8N7F4S1		$[M+3H]^{3+}$	1553.3057	3.0	0.4	No						$[M+3H]^{3+}$	1553.3062	2.6	8.3	Yes
													$[M+4H]^{4+}$	1165.2292	4.7		Yes
88	H8N8F4S1												$[M+4H]^{4+}$	1215.9976	5.7	5.0	Yes
89	H8N7F5S1												$[M+4H]^{4+}$	1201.7426	5.5	8.8	Yes
90	H8N7F6S1												$[M+4H]^{4+}$	1238.2597	3.2	6.0	Yes
91	H9N8F6S1												$[M+4H]^{4+}$	1329.5385	6.2	8.6	Yes
92	H9N8F7S1												$[M+4H]^{4+}$	1366.0530	6.0	4.1	Yes

N-375 Peptide: LQLSNDNR																
		CEA1 Human Colon Carcinoma					CEA2 Human Colon Carcinoma					CEA3 Human Liver Metastases				
Num ^a	Glycan Species ^b	Molecular ion	Observed <i>m/z</i>	Error (ppm) ^c	%Relative Abundance	MS/MS ^d	Molecular ion	Observed <i>m/z</i>	Error (ppm) ^c	%Relative Abundance	MS/MS ^d	Molecular ion	Observed <i>m/z</i>	Error (ppm) ^c	%Relative Abundance	MS/MS ^d
93	H3N3		[M+2H] ²⁺	1027.9444	3.0	20.4	Yes	[M+2H] ²⁺	1027.9387	8.6	19.4	No				
94	H3N7		[M+3H] ³⁺	956.4025	4.5	41.7	Yes	[M+3H] ³⁺	956.3975	9.7	30.5	No				

95	H7N6F1		$[M+3H]^{3+}$	1153.4688	1.1	24.4	No	$[M+3H]^{3+}$	1153.4589	9.7	39.5	No	$[M+3H]^{3+}$	1153.4790	7.7	26.9	Yes
96	H8N7F4S1		$[M+4H]^{4+}$	1138.9643	8.3	13.5	No	$[M+4H]^{4+}$	1138.9626	6.8	10.6	No	$[M+4H]^{4+}$	1138.9501	4.2	73.1	Yes










N-580 Peptide: AYVCGIQNSVSA N R																
		CEA1 Human Colon Carcinoma					CEA2 Human Colon Carcinoma					CEA3 Human Liver Metastases				
Num ^a	Glycan Species ^b	Molecular ion	Observed m/z	Error (ppm) ^c	%Relative Abundance	MS/MS ^d	Molecular ion	Observed m/z	Error (ppm) ^c	%Relative Abundance	MS/MS ^d	Molecular ion	Observed m/z	Error (ppm) ^c	%Relative Abundance	MS/MS ^d
97	H7N6F4S1 											$[M+4H]^{4+}$	1192.4792	3.8	100	Yes

N-612 Peptide: SDPVTLDVLYGPDTP I ISPDPSSYLSGAN L NLSCHSASNPSPQYSWR																
		CEA1 Human Colon Carcinoma					CEA2 Human Colon Carcinoma					CEA3 Human Liver Metastases				
Num ^a	Glycan Species ^b	Molecular ion	Observed m/z	Error (ppm) ^c	%Relative Abundance	MS/MS ^d	Molecular ion	Observed m/z	Error (ppm) ^c	%Relative Abundance	MS/MS ^d	Molecular ion	Observed m/z	Error (ppm) ^c	%Relative Abundance	MS/MS ^d
98	H6N5F3S1 											$[M+4H]^{4+}$	1948.8437	4.1	68.5	Yes
												$[M+5H]^{5+}$	1559.2852	1.4		Yes
99	H6N5F4S1 											$[M+5H]^{5+}$	1588.4939	0.4	31.5	Yes

N-650 Peptide: ITPNN G TYACFVSNLATGR																
		CEA1 Human Colon Carcinoma					CEA2 Human Colon Carcinoma					CEA3 Human Liver Metastases				
Num ^a	Glycan Species ^b	Molecular ion	Observed m/z	Error (ppm) ^c	%Relative Abundance	MS/MS ^d	Molecular ion	Observed m/z	Error (ppm) ^c	%Relative Abundance	MS/MS ^d	Molecular ion	Observed m/z	Error (ppm) ^c	%Relative Abundance	MS/MS ^d
100	H4N3F2 											$[M+3H]^{3+}$	1240.5323	5.0	4.7	Yes
101	H6N5F4 											$[M+3H]^{3+}$	1581.3278	2.6	5.3	Yes
102	H4N3F1S1 											$[M+3H]^{3+}$	1288.8777	5.2	8.2	Yes
103	H5N3F1S1 											$[M+3H]^{3+}$	1342.8935	6.3	10.0	Yes

104	HSN4F2S1												[M+4H] ⁴⁺	1094.7081	4.3	10.9	Yes
105	H6N5F2S1												[M+4H] ⁴⁺	1185.9855	8.7	5.3	Yes
106	H6N5F3S1												[M+4H] ⁴⁺	1222.5031	5.9	16.6	Yes
107	H6N5F4S1												[M+4H] ⁴⁺	1259.0174	5.9	11.4	Yes
108	HSN4F1S2												[M+4H] ⁴⁺	1130.9635	7.7	4.6	Yes
109	H6N5F2S2												[M+4H] ⁴⁺	1258.7646	4.0	14.6	Yes
110	H6N5F3S2												[M+4H] ⁴⁺	1295.2774	5.2	8.3	Yes

N-665 Peptide: NNSIVK																
		CEA1 Human Colon Carcinoma					CEA2 Human Colon Carcinoma					CEA3 Human Liver Metastases				
Num ^a	Glycan Species ^b	Molecular ion	Observed <i>m/z</i>	Error (ppm) ^c	%Relative Abundance	MS/MS ^d	Molecular ion	Observed <i>m/z</i>	Error (ppm) ^c	%Relative Abundance	MS/MS ^d	Molecular ion	Observed <i>m/z</i>	Error (ppm) ^c	%Relative Abundance	MS/MS ^d
111	HSN2	[M+2H] ²⁺	945.9042	2.9	23.9	Yes	[M+2H] ²⁺	945.8977	9.8	37.5	No	[M+2H] ²⁺	945.9024	4.8	1.3	No
112	H5N5F2	[M+3H] ³⁺	931.3930	1.2	15.8	No	[M+3H] ³⁺	931.3850	7.4	26.7	No	[M+3H] ³⁺	931.3889	3.2	5.1	Yes
113	H5N5F3	[M+3H] ³⁺	980.0778	0.1	12.0	No	[M+3H] ³⁺	980.0719	6.1	19.9	No	[M+3H] ³⁺	980.0742	3.8	10.7	Yes
114	H5N5F4	[M+3H] ³⁺	1028.7642	0.3	3.9	No						[M+3H] ³⁺	1028.7598	4.0	3.5	Yes
115	H6N6F2	[M+3H] ³⁺	1053.1054	2.6	7.5	No						[M+3H] ³⁺	1053.0983	4.1	3.0	Yes
116	H6N6F3	[M+3H] ³⁺	1101.7902	1.4	7.9	No						[M+3H] ³⁺	1101.7839	4.3	4.1	Yes
117	H6N6F4	[M+3H] ³⁺	1150.4799	4.6	7.7	No						[M+3H] ³⁺	1150.4724	1.9	3.8	Yes
118	H7N7F2	[M+3H] ³⁺	1174.8160	2.2	4.4	No						[M+3H] ³⁺	1174.8096	3.2	3.3	Yes
119	H7N7F4	[M+3H] ³⁺	1272.1855	0.1	4.0	No						[M+3H] ³⁺	1272.1801	4.1	5.0	Yes

120	H7N7F5		[M+3H] ³⁺	1320.8733	1.5	4.6	No						[M+3H] ³⁺	1320.8644	5.2	5.1	Yes
121	H5N5F2S1		[M+3H] ³⁺	1028.4223	1.4	8.2	No	[M+3H] ³⁺	1028.4290	5.1	16.0	No	[M+3H] ³⁺	1028.4197	3.9	8.2	Yes
122	H6N5F2S1												[M+3H] ³⁺	1082.4354	5.5	4.5	Yes
123	H6N5F3S1												[M+3H] ³⁺	1131.1235	3.4	5.2	Yes
124	H7N6F2S1												[M+3H] ³⁺	1204.1469	4.3	2.7	Yes
125	H7N6F3S1												[M+3H] ³⁺	1252.8318	5.0	7.5	Yes
													[M+4H] ⁴⁺	939.8775	3.2		Yes
126	H7N6F4S1												[M+3H] ³⁺	1301.5175	5.0	12.0	Yes
													[M+4H] ⁴⁺	976.3905	4.6		Yes
127	H7N6F5S1												[M+3H] ³⁺	1350.2041	4.4	9.2	Yes
													[M+4H] ⁴⁺	1012.9059	3.5		Yes
128	H7N6F6S1												[M+3H] ³⁺	1398.8870	6.4	5.6	Yes
													[M+4H] ⁴⁺	1049.4187	5.0		Yes

^aNum is referred to the *N*-glycopeptide number.

^bGlycan compositions are given in terms of number of hexoses (H), N-acetylglucosamines and N-acetylgalactosamines (N), fucoses (F) and sialic acids (S).

^cRelative error was calculated in ppm as: $|m/z_{\text{exp}} - m/z_{\text{theo}}|/m/z_{\text{theo}} \times 10^6$ (exp=experimental and theo=theoretical).

^d*N*-glycopeptide structures confirmed by MS/MS (Yes) (At least 3 observed fragments).

Supplementary Table S-2. Observed average m/z ($n=6$), relative error and %relative abundances (normalizing the integrated peak areas by the total integrated peak areas for the N -glycosylation site) for the CEA N -glycopeptides identified by sheathless CE-MS/MS after digestion with Glu-C.

N-152/N-508 Peptide: LPKPSISSNNSKPVE (MC 0)/LPKPSISSNNSKPVED (MC 1)/LPKPSISSNNSKPVEDKD (MC 2)/LPKPSISSNNSKPVEDKDAVFTCEPE (MC 3)																				
		CEA1 Human Colon Carcinoma						CEA2 Human Colon Carcinoma						CEA3 Human Liver Metastases						
Num ^a	Glycan Species ^b	MC ^c	Molecular ion	Observed m/z	Error (ppm) ^d	%Relative Abundance	MS/MS ^e	MC ^c	Molecular ion	Observed m/z	Error (ppm) ^d	%Relative Abundance	MS/MS ^e	MC ^c	Molecular ion	Observed m/z	Error (ppm) ^d	%Relative Abundance	MS/MS ^e	
1-4	H5N2		0	[M+3H] ³⁺	938.4234	9.8	0.2	No						0	[M+3H] ³⁺	938.4260	7.0	0.6	No	
			1	[M+3H] ³⁺	976.7675	7.6	1.0	No	1	[M+3H] ³⁺	976.7666	8.5	1.5	No						
			2	[M+4H] ⁴⁺	793.6081	7.0	7.4	Yes	2	[M+4H] ⁴⁺	793.6070	8.4	7.4	Yes	2	[M+4H] ⁴⁺	793.6094	5.4	5.5	Yes
			3	[M+4H] ⁴⁺	1044.7106	9.5	6.4	Yes	3	[M+4H] ⁴⁺	1044.7165	3.8	4.0	Yes	3	[M+4H] ⁴⁺	1044.7116	8.5	5.0	Yes
5-8	H6N2		0	[M+3H] ³⁺	992.4462	4.1	0.5	No	0	[M+3H] ³⁺	992.4414	8.9	0.7	No	0	[M+3H] ³⁺	992.4470	3.2	0.3	No
			1	[M+3H] ³⁺	1030.7845	7.8	0.6	No	1	[M+3H] ³⁺	1030.7876	4.8	0.8	No						
			2	[M+4H] ⁴⁺	834.1202	8.0	4.7	Yes	2	[M+4H] ⁴⁺	834.1186	9.9	5.2	Yes	2	[M+4H] ⁴⁺	834.1211	6.9	4.4	Yes
			3	[M+4H] ⁴⁺	1085.2258	7.3	3.8	Yes	3	[M+4H] ⁴⁺	1085.2287	4.6	2.6	Yes	3	[M+4H] ⁴⁺	1085.2322	9.7	3.5	Yes
9-12	H7N2		0	[M+3H] ³⁺	1046.4602	7.3	0.4	No	0	[M+3H] ³⁺	1046.4594	8.1	1.5	Yes	0	[M+3H] ³⁺	1046.4618	5.8	0.4	No
			1	[M+3H] ³⁺	1084.8002	9.2	1.1	No	1	[M+3H] ³⁺	1084.8077	2.3	1.9	No	1	[M+3H] ³⁺	1084.8026	7.0	1.0	No
			2	[M+4H] ⁴⁺	874.6319	9.4	11.6	Yes	2	[M+4H] ⁴⁺	874.6314	9.9	14.3	Yes	2	[M+4H] ⁴⁺	874.6330	8.1	12.7	Yes
			3	[M+4H] ⁴⁺	1125.7357	10.0	8.8	Yes	3	[M+4H] ⁴⁺	1125.7390	7.0	6.0	Yes	3	[M+4H] ⁴⁺	1125.7395	6.6	9.1	Yes
13-16	H8N2		0	[M+3H] ³⁺	1100.4775	7.2	0.5	No	0	[M+3H] ³⁺	1100.4790	5.8	2.3	No	0	[M+3H] ³⁺	1100.4785	6.3	0.5	No
			1	[M+3H] ³⁺	1138.8173	9.2	1.7	Yes	1	[M+3H] ³⁺	1138.8191	7.6	3.1	Yes	1	[M+3H] ³⁺	1138.8194	7.4	1.7	Yes
			2	[M+4H] ⁴⁺	915.1443	9.8	16.7	Yes	2	[M+4H] ⁴⁺	915.1467	7.2	19.8	Yes	2	[M+4H] ⁴⁺	915.1460	8.0	19.0	Yes
			3	[M+4H] ⁴⁺	1166.2498	8.8	13.9	Yes	3	[M+4H] ⁴⁺	1166.2556	3.9	8.6	Yes	3	[M+4H] ⁴⁺	1166.2558	3.7	14.6	Yes
17-20	H9N2		0	[M+3H] ³⁺	1154.5074	3.8	0.4	No	0	[M+3H] ³⁺	1154.4953	6.7	1.6	No	0	[M+3H] ³⁺	1154.4930	8.7	0.6	No
			1	[M+3H] ³⁺	1192.8364	7.5	1.1	No	1	[M+3H] ³⁺	1192.8394	5.0	1.7	No	1	[M+3H] ³⁺	1192.8379	6.3	1.1	No
			2	[M+4H] ⁴⁺	955.6573	9.6	10.5	Yes	2	[M+4H] ⁴⁺	955.6597	7.1	11.8	Yes	2	[M+4H] ⁴⁺	955.6594	7.4	10.9	Yes
			3	[M+4H] ⁴⁺	1206.7617	9.6	8.5	Yes	3	[M+4H] ⁴⁺	1206.7681	4.3	5.1	Yes	3	[M+4H] ⁴⁺	1206.7719	1.2	9.0	Yes

N-466 Peptide: LFISNITE																				
		CEA1 Human Colon Carcinoma							CEA2 Human Colon Carcinoma					CEA3 Human Liver Metastases						
Num ^a	Glycan Species ^b	MC ^c	Molecular ion	Observed m/z	Error (ppm) ^d	%Relative Abundance	MS/MS ^e	MC ^c	Molecular ion	Observed m/z	Error (ppm) ^d	%Relative Abundance	MS/MS ^e	MC ^c	Molecular ion	Observed m/z	Error (ppm) ^d	%Relative Abundance	MS/MS ^e	
21	H3N3F1		0	[M+2H] ²⁺	1089.4786	4.1	5.8	Yes	0	[M+2H] ²⁺	1089.4735	8.7	7.1	Yes	0	[M+2H] ²⁺	1089.4725	9.7	7.4	No
22	H4N3F1		0	[M+2H] ²⁺	1170.5024	6.0	1.2	No	0	[M+2H] ²⁺	1170.4988	9.1	1.2	No	0	[M+2H] ²⁺	1170.4988	9.1	12.7	No
23	H3N4F1		0	[M+2H] ²⁺	1191.0168	5.0	7.3	No	0	[M+2H] ²⁺	1191.0134	7.8	6.8	No						
24	H3N5F1		0	[M+3H] ³⁺	862.0427	1.8	11.0	Yes	0	[M+3H] ³⁺	862.0383	6.9	11.1	No						
25	H4N4F1		0	[M+2H] ²⁺	1272.0440	4.0	3.3	No	0	[M+3H] ³⁺	848.3665	2.6	3.0	No						
26	H4N5F1		0	[M+3H] ³⁺	916.0592	2.9	7.6	No	0	[M+3H] ³⁺	916.0542	8.3	7.6	No						

27	HSN4F1		0	[M+2H] ²⁺	1353.0734	1.6	2.3	No	0	[M+3H] ³⁺	902.3795	7.6	1.8	No						
28	H3N6F1		0	[M+3H] ³⁺	929.7337	3.9	7.5	Yes	0	[M+3H] ³⁺	929.7296	8.4	7.2	No						
29	H4N4F2		0	[M+3H] ³⁺	897.0516	3.4	2.2	No	0	[M+3H] ³⁺	897.0460	9.7	1.6	No						
30	H4NSF2		0	[M+3H] ³⁺	964.7441	3.9	4.9	No	0	[M+3H] ³⁺	964.7382	10.0	5.1	No	0	[M+3H] ³⁺	964.7402	7.9	10.4	No
31	H4N6F1		0	[M+3H] ³⁺	983.7511	3.9	4.9	Yes	0	[M+3H] ³⁺	983.7465	8.6	7.2	Yes						
32	HSN5F1		0	[M+3H] ³⁺	970.0751	4.5	4.3	No	0	[M+3H] ³⁺	970.0828	3.4	4.7	No	0	[M+3H] ³⁺	970.0741	5.5	9.9	No
33	HSN5F2		0	[M+3H] ³⁺	1018.7612	4.1	3.5	No	0	[M+3H] ³⁺	1018.7560	9.3	4.2	No	0	[M+3H] ³⁺	1018.7573	8.0	10.7	No
34	HSN6F1		0	[M+3H] ³⁺	1037.7686	3.8	4.5	No	0	[M+3H] ³⁺	1037.7629	9.3	4.4	No						
35	H4N6F3		0	[M+3H] ³⁺	1081.1161	10.0	2.1	No	0	[M+3H] ³⁺	1081.1195	6.9	1.8	No	0	[M+3H] ³⁺	1081.1173	8.9	5.4	No
36	HSN5F3		0	[M+3H] ³⁺	1067.4406	10.0	4.1	No	0	[M+3H] ³⁺	1067.4446	6.4	3.1	No	0	[M+3H] ³⁺	1067.4420	8.8	18.8	No
37	HSN6F2		0	[M+3H] ³⁺	1086.4540	4.2	4.6	No	0	[M+3H] ³⁺	1086.4478	9.9	4.0	No	0	[M+3H] ³⁺	1086.4486	9.2	12.6	No
38	HSN7F1		0	[M+3H] ³⁺	1105.4613	4.0	4.8	Yes	0	[M+3H] ³⁺	1105.4561	8.7	5.6	Yes						
39	HSN6F3		0	[M+3H] ³⁺	1135.1376	6.1	3.8	No	0	[M+3H] ³⁺	1135.1334	9.8	4.0	No	0	[M+3H] ³⁺	1135.1375	6.2	12.0	No
40	HSN7F2		0	[M+3H] ³⁺	1154.1483	2.9	5.7	Yes	0	[M+3H] ³⁺	1154.1418	8.6	4.6	No						
41	HSN7F3		0	[M+3H] ³⁺	1202.8315	5.1	4.7	Yes	0	[M+3H] ³⁺	1202.8263	9.4	3.9	No						

N-580 Peptide: ARAYVCGIQNSVSANRSDPVTLD																			
		CEA1 Human Colon Carcinoma						CEA2 Human Colon Carcinoma						CEA3 Human Liver Metastases					
Num ^a	Glycan Species ^b	MC ^c	Molecular ion	Observed m/z	Error (ppm) ^d	%Relative Abundance	MS/MS ^e	MC ^c	Molecular ion	Observed m/z	Error (ppm) ^d	%Relative Abundance	MS/MS ^e	MC ^c	Molecular ion	Observed m/z	Error (ppm) ^d	%Relative Abundance	MS/MS ^e
42	H3N3F1		0	[M+4H] ⁴⁺	934.4188	4.8	11.8	No	0	[M+4H] ⁴⁺	934.4190	4.6	14.1	No					
43	H3N4F1		0	[M+4H] ⁴⁺	985.1851	8.2	17.3	Yes	0	[M+4H] ⁴⁺	985.1858	7.5	16.1	No					

44	H4N3F1		0	[M+4H] ⁴⁺	974.9353	1.3	4.3	No	0	[M+4H] ⁴⁺	974.9273	9.5	4.1	No						
45	H3N5F1		0	[M+4H] ⁴⁺	1035.9526	10.0	13.6	No	0	[M+4H] ⁴⁺	1035.9584	4.5	15.5	No						
46	H4N4F1		0	[M+4H] ⁴⁺	1025.6987	7.5	10.3	No	0	[M+4H] ⁴⁺	1025.6991	7.1	9.4	No						
47	H3N6F1		0	[M+4H] ⁴⁺	1086.7230	9.1	6.1	No	0	[M+4H] ⁴⁺	1086.7232	8.9	4.4	No						
48	H4N5F1		0	[M+4H] ⁴⁺	1076.4666	8.9	16.7	No	0	[M+4H] ⁴⁺	1076.4664	9.1	18.7	No						
49	H4N6F1		0	[M+4H] ⁴⁺	1127.2381	7.1	6.9	No	0	[M+4H] ⁴⁺	1127.2357	9.2	7.1	No						
50	H5N5F1		0	[M+4H] ⁴⁺	1116.9785	9.8	13.0	No	0	[M+4H] ⁴⁺	1116.9790	9.3	10.7	No						

^aNum is referred to the *N*-glycopeptide number.


^bGlycan compositions are given in terms of number of hexoses (H), N-acetylglucosamines and N-acetylgalactosamines (N), fucoses (F) and sialic acids (S).


^cMC=missed cleavage.






^dRelative error was calculated in ppm as: $|m/z_{\text{exp}} - m/z_{\text{theo}}| / m/z_{\text{theo}} \times 10^6$ (exp=experimental and theo=theoretical).








^e*N*-glycopeptide structures confirmed by MS/MS (Yes) (At least 3 observed fragments).


Supplementary Table S-3-A. Observed average m/z ($n=6$), relative error and %relative abundances (normalizing the integrated peak areas by the total integrated peak areas for the N -glycosylation site) for the CEA N -glycopeptides identified by sheathless CE-MS/MS after digestion with pronase.






N-104 Peptide: IYPN																		
			CEA1 Human Colon Carcinoma					CEA2 Human Colon Carcinoma					CEA3 Human Liver Metastases					
Num ^a	Glycan Species ^b		Peptide Sequence(s):	Molecular ion	Observed m/z	Error (ppm) ^c	%Relative Abundance	MS/MS ^d	Molecular ion	Observed m/z	Error (ppm) ^c	%Relative Abundance	MS/MS ^d	Molecular ion	Observed m/z	Error (ppm) ^c	%Relative Abundance	MS/MS ^d
1	H3N4		IYPN	[M+2H] ²⁺	902.8665	6.9	100	Yes	[M+2H] ²⁺	902.8667	6.7	100	Yes					


N-208 Peptide: NDTA																		
			CEA1 Human Colon Carcinoma					CEA2 Human Colon Carcinoma					CEA3 Human Liver Metastases					
Num ^a	Glycan Species ^b		Peptide Sequence(s):	Molecular ion	Observed m/z	Error (ppm) ^c	%Relative Abundance	MS/MS ^d	Molecular ion	Observed m/z	Error (ppm) ^c	%Relative Abundance	MS/MS ^d	Molecular ion	Observed m/z	Error (ppm) ^c	%Relative Abundance	MS/MS ^d
2	H3N5		NDTA	[M+2H] ²⁺										[M+2H] ²⁺	961.3677	0.5	100	Yes
														[M+3H] ²⁺	641.2541	9.3		No


N-256 Peptide(s): GENLNLS / SGENLNIL																		
			CEA1 Human Colon Carcinoma						CEA2 Human Colon Carcinoma					CEA3 Human Liver Metastases				
Num ^a	Glycan Species ^b		Peptide Sequence(s):	Molecular ion	Observed m/z	Error (ppm) ^c	%Relative Abundance	MS/MS ^d	Molecular ion	Observed m/z	Error (ppm) ^c	%Relative Abundance	MS/MS ^d	Molecular ion	Observed m/z	Error (ppm) ^c	%Relative Abundance	MS/MS ^d
3	H5N2		GENLNLS / SGENLNIL	[M+2H] ²⁺	981.8979	1.8	31.7	No	[M+2H] ²⁺	981.8967	3.0	29.8	No	[M+2H] ²⁺	981.9034	3.8	50.2	Yes
4	H6N2		GENLNLS / SGENLNIL	[M+2H] ²⁺	1062.9223	3.6	44.7	No	[M+2H] ²⁺	1062.9208	5.0	45.0	No					
5	H7N2		GENLNLS / SGENLNIL	[M+2H] ²⁺	1143.9418	9.3	23.6	Yes	[M+2H] ²⁺	1143.9429	8.4	25.2	No	[M+2H] ²⁺	1143.9501	2.1	6.8	Yes
6	H8N2		GENLNLS / SGENLNIL											[M+2H] ²⁺	1224.9787	0.2	29.7	Yes
7	H9N2		GENLNLS / SGENLNIL											[M+2H] ²⁺	1306.0044	0.7	13.3	No









N-274 Peptide(s): NG / NGTF																		
			CEA1 Human Colon Carcinoma						CEA2 Human Colon Carcinoma					CEA3 Human Liver Metastases				
Num ^a	Glycan Species ^b		Peptide Sequence(s):	Molecular ion	Observed m/z	Error (ppm) ^c	%Relative Abundance	MS/MS ^d	Molecular ion	Observed m/z	Error (ppm) ^c	%Relative Abundance	MS/MS ^d	Molecular ion	Observed m/z	Error (ppm) ^c	%Relative Abundance	MS/MS ^d
8	H4N2		NG	[M+2H] ²⁺	622.7287	2.7	13.1	No	[M+2H] ²⁺	622.7293	1.8	21.3	Yes	[M+2H] ²⁺	622.7311	1.1	10.6	Yes
9	H5N2		NG	[M+2H] ²⁺	703.7540	4.0	52.1	Yes	[M+2H] ²⁺	703.7545	3.3	52.2	Yes	[M+2H] ²⁺	703.7578	1.4	35.8	Yes
10	H6N2		NG	[M+2H] ²⁺	784.7823	1.2	2.6	No	[M+2H] ²⁺	784.7813	2.5	3.6	No	[M+2H] ²⁺	784.7864	4.0	1.8	No
11	H6N5F3S1		NG											[M+3H] ³⁺	969.6922	1.7	20.0	Yes
12	H5N4S2		NG											[M+3H] ³⁺	798.9556	1.8	17.8	Yes
13	H5N2		NGTF	[M+2H] ²⁺	827.8117	3.8	24.9	No	[M+2H] ²⁺	827.8117	3.8	17.7	No	[M+2H] ²⁺	827.8158	1.1	11.2	Yes
14	H6N2		NGTF	[M+2H] ²⁺	908.8350	6.9	7.2	No	[M+2H] ²⁺	908.8361	5.7	5.3	No	[M+2H] ²⁺	908.8421	0.9	2.8	No

N-330 Peptide: ITSNN																		
			CEA1 Human Colon Carcinoma					CEA2 Human Colon Carcinoma					CEA2 Human Liver Metastases					
Num ^a	Glycan Species ^b		Peptide Sequence(s):	Molecular ion	Observed <i>m/z</i>	Error (ppm) ^c	%Relative Abundance	MS/MS ^d	Molecular ion	Observed <i>m/z</i>	Error (ppm) ^c	%Relative Abundance	MS/MS ^d	Molecular ion	Observed <i>m/z</i>	Error (ppm) ^c	%Relative Abundance	MS/MS ^d
15	H5N2		ITSNN	[M+2H] ²⁺	882.8350	16.4	100	Yes	[M+2H] ²⁺	882.8349	16.5	100	Yes	[M+2H] ²⁺	882.8410	9.6	100	Yes

N-351 Peptide: PEIQNTT																		
			CEA1 Human Colon Carcinoma						CEA2 Human Colon Carcinoma					CEA3 Human Liver Metastases				
Num ^a	Glycan Species ^b		Peptide Sequence(s):	Molecular ion	Observed m/z	Error (ppm) ^c	%Relative Abundance	MS/MS ^d	Molecular ion	Observed m/z	Error (ppm) ^c	%Relative Abundance	MS/MS ^d	Molecular ion	Observed m/z	Error (ppm) ^c	%Relative Abundance	MS/MS ^d
16	H5N2		PEIQNTT	[M+2H] ²⁺	1009.9130	0.2	8.7	No	[M+2H] ²⁺	1009.9138	1.0	10.5	No	[M+2H] ²⁺	1009.9189	6.1	20.5	Yes
17	H6N2		PEIQNTT	[M+2H] ²⁺	1090.9377	1.4	5.7	No	[M+2H] ²⁺	1090.9381	1.0	7.9	No	[M+2H] ²⁺	1090.9455	5.8	14.3	No
18	H7N2		PEIQNTT	[M+2H] ²⁺	1171.9660	0.3	10.6	No	[M+2H] ²⁺	1171.9624	2.7	15.0	No	[M+2H] ²⁺	1171.9738	7.0	21.9	Yes
19	H8N2		PEIQNTT	[M+2H] ²⁺	1252.9913	0.6	24.2	No	[M+2H] ²⁺	1252.9899	1.7	24.9	No	[M+2H] ²⁺	1253.0006	6.9	18.3	Yes
				[M+3H] ³⁺	835.6667	3.2		No	[M+3H] ³⁺	835.6657	2.1		No	[M+3H] ³⁺	835.6716	9.1		No
20	H9N2		PEIQNTT	[M+2H] ²⁺	1334.0118	5.0	50.8	No	[M+2H] ²⁺	1334.0112	5.4	41.9	No	[M+2H] ²⁺	1334.0277	7.0	24.9	No
				[M+3H] ³⁺	889.6806	1.1		Yes	[M+3H] ³⁺	889.6802	1.6		Yes	[M+3H] ³⁺	889.6872	6.3		No

N-375 Peptide: NDNR																		
			CEA1 Human Colon Carcinoma					CEA2 Human Colon Carcinoma					CEA3 Human Liver Metastases					
Num ^a	Glycan Species ^b		Peptide Sequence(s):	Molecular ion	Observed m/z	Error (ppm) ^c	%Relative Abundance	MS/MS ^d	Molecular ion	Observed m/z	Error (ppm) ^c	%Relative Abundance	MS/MS ^d	Molecular ion	Observed m/z	Error (ppm) ^c	%Relative Abundance	MS/MS ^d
21	H5N2		NDNR	[M+2H] ²⁺	867.8250	7.6	100	No	[M+2H] ²⁺	867.8262	6.2	100	Yes	[M+2H] ²⁺	867.8316	0.02	100	Yes

N-432 Peptide: RPGVN																		
			CEA1 Human Colon Carcinoma					CEA2 Human Colon Carcinoma					CEA3 Human Liver Metastases					
Num ^a	Glycan Species ^b		Peptide Sequence(s):	Molecular ion	Observed <i>m/z</i>	Error (ppm) ^c	%Relative Abundance	MS/MS ^d	Molecular ion	Observed <i>m/z</i>	Error (ppm) ^c	%Relative Abundance	MS/MS ^d	Molecular ion	Observed <i>m/z</i>	Error (ppm) ^c	%Relative Abundance	MS/MS ^d
22	H3N2F1		RPGVN	[M+2H] ²⁺	790.8429	1.5	100	No	[M+2H] ²⁺	790.8432	1.1	100	Yes					

N-529 Peptide: EAQN																		
			CEA1 Human Colon Carcinoma						CEA2 Human Colon Carcinoma					CEA3 Human Liver Metastases				
Num ^a	Glycan Species ^b		Peptide Sequence(s):	Molecular ion	Observed <i>m/z</i>	Error (ppm) ^c	%Relative Abundance	MS/MS ^d	Molecular ion	Observed <i>m/z</i>	Error (ppm) ^c	%Relative Abundance	MS/MS ^d	Molecular ion	Observed <i>m/z</i>	Error (ppm) ^c	%Relative Abundance	MS/MS ^d
23	H5N2		EAQN	[M+2H] ²⁺	839.3210	6.8	70.9	Yes	[M+2H] ²⁺	839.3201	5.8	56.8	Yes	[M+2H] ²⁺	839.3228	9.0	65.7	No
24	H6N2		EAQN											[M+2H] ²⁺	920.3479	6.8	4.1	No
25	H7N2		EAQN											[M+2H] ²⁺	1001.3743	6.2	1.0	No
26	H8N2		EAQN											[M+2H] ²⁺	1082.4022	7.1	1.5	No
27	H5N4		EAQN	[M+2H] ²⁺	1042.3937	0.9	7.9	No	[M+2H] ²⁺	1042.3945	0.1	10.7	No	[M+2H] ²⁺	1042.4003	5.4	12.1	No
				[M+3H] ³⁺	695.2666	1.2		Yes	[M+3H] ³⁺	695.2656	0.2		Yes	[M+3H] ³⁺	695.2677	2.8		No
28	H4N5		EAQN	[M+2H] ²⁺	1062.9056	2.2	5.4	No	[M+2H] ²⁺	1062.9029	4.7	8.7	No	[M+2H] ²⁺	1062.9130	4.8	1.8	No
				[M+3H] ³⁺	708.9416	0.5		Yes	[M+3H] ³⁺	708.9406	0.9		Yes	[M+3H] ³⁺	708.9443	4.3		No
29	H3N6		EAQN	[M+2H] ²⁺	1083.4117	8.8	2.8	No										
				[M+3H] ³⁺	722.6165	0.4		Yes	[M+3H] ³⁺	722.6158	1.3	4.7	Yes					
30	H6N4		EAQN	[M+2H] ²⁺	1123.4164	4.1	4.2	No	[M+2H] ²⁺	1123.4167	3.9	5.9	No	[M+2H] ²⁺	1123.4244	3.0	10.2	No
				[M+3H] ³⁺	749.2795	5.1		Yes	[M+3H] ³⁺	749.2833	0.1		No	[M+3H] ³⁺	749.2848	1.9		Yes





31	H4N3F1		EAQN	[M+2H] ²⁺	1135.9293	6.7	5.0	No	[M+2H] ²⁺	1135.9307	5.4	6.5	No	[M+2H] ²⁺	1135.9363	0.5	3.6	No
				[M+3H] ³⁺	757.6253	2.5		Yes	[M+3H] ³⁺	757.6265	1.0		No	[M+3H] ³⁺	757.6305	4.3		No
32	H4N6		EAQN	[M+2H] ²⁺	1164.4407	5.9	3.8	No	[M+2H] ²⁺	1164.4427	4.2	6.6	No					
				[M+3H] ³⁺	776.6328	2.0		Yes	[M+3H] ³⁺	776.6329	1.9		No					


N-612 Peptide: NLSCHS																		
			CEA1 Human Colon Carcinoma					CEA2 Human Colon Carcinoma					CEA3 Human Liver Metastases					
Num ^a	Glycan Species ^b		Peptide Sequence(s):	Molecular ion	Observed <i>m/z</i>	Error (ppm) ^c	%Relative Abundance	MS/MS ^d	Molecular ion	Observed <i>m/z</i>	Error (ppm) ^c	%Relative Abundance	MS/MS ^d	Molecular ion	Observed <i>m/z</i>	Error (ppm) ^c	%Relative Abundance	MS/MS
33	H3N3F1		NLSCHS	[M+2H] ²⁺	979.8783	2.5	45.7	No	[M+2H] ²⁺	979.8791	1.7	48.6	Yes					
34	H4N3F1		NLSCHS	[M+2H] ²⁺	1060.9061	1.0	29.2	No	[M+2H] ²⁺	1060.9059	1.2	22.4	Yes					
35	H3N4F1		NLSCHS	[M+2H] ²⁺	1081.4175	2.8	25.1	No	[M+2H] ²⁺	1081.4173	2.9	29.0	Yes	[M+2H] ²⁺	1081.4134	6.5	10.8	No
36	H6N5F4S1		NLSCHS											[M+3H] ³⁺	1194.1247	5.1	9.5	No
37	H6N5F2S2		NLSCHS											[M+3H] ³⁺	1193.7834	4.1	51.0	Yes
38	H6N5F3S2		NLSCHS											[M+3H] ³⁺	1242.4688	3.5	28.7	Yes



N-665 Peptide: RNN																		
			CEA1 Human Colon Carcinoma						CEA2 Human Colon Carcinoma					CEA3 Human Liver Metastases				
Num ^a	Glycan Species ^b		Peptide Sequence(s):	Molecular ion	Observed <i>m/z</i>	Error (ppm) ^c	%Relative Abundance	MS/MS ^d	Molecular ion	Observed <i>m/z</i>	Error (ppm) ^c	%Relative Abundance	MS/MS ^d	Molecular ion	Observed <i>m/z</i>	Error (ppm) ^c	%Relative Abundance	MS/MS ^d
39	H5N2																	


Supplementary Table S-3-B. Observed average m/z ($n=6$) and relative error for the CEA N -glycopeptides (grouped by N -glycosylation sites) identified by sheathless CE-MS/MS after digestion with pronase. These N -glycopeptides were discarded for the Results and Discussion Section, as multiple overlaps between N -glycosylation sites were observed.


NAS (N-104) / NGT (N-274/N-650) / SAN (N-580)															
Num ^a	Glycan Species ^b		Peptide Sequence(s):	CEA1 Human Colon Carcinoma				CEA2 Human Colon Carcinoma				CEA3 Human Liver Metastases			
				Molecular ion	Observed m/z	Error (ppm) ^c	MS/MS ^d	Molecular ion	Observed m/z	Error (ppm) ^c	MS/MS ^d	Molecular ion	Observed m/z	Error (ppm) ^c	MS/MS ^d
1	H5N2		NAS / NGT / SAN	[M+2H] ²⁺	754.2787	2.6	Yes	[M+2H] ²⁺	754.2782	3.3	Yes	[M+2H] ²⁺	754.2813	0.8	Yes
2	H6N2		NAS / NGT / SAN	[M+2H] ²⁺	835.3044	3.2	No	[M+2H] ²⁺	835.3043	3.3	Yes	[M+2H] ²⁺	835.3082	1.3	Yes
3	H7N2		NAS / NGT / SAN	[M+2H] ²⁺	916.3289	5.0	No	[M+2H] ²⁺	916.3283	5.7	No	[M+2H] ²⁺	916.3345	1.1	Yes
4	H4N3F1		NAS / NGT / SAN	[M+2H] ²⁺	847.8192	4.4	No	[M+2H] ²⁺	847.8199	3.5	No	[M+2H] ²⁺	847.8234	0.6	Yes
5	H5N3		NAS / NGT / SAN	[M+2H] ²⁺	855.8179	2.9	No	[M+2H] ²⁺	855.8172	3.7	Yes	[M+2H] ²⁺	855.8222	2.2	No
6	H4N3F2		NAS / NGT / SAN	[M+2H] ²⁺	920.8476	4.6	No	[M+2H] ²⁺	920.8467	5.6	No	[M+2H] ²⁺	920.8517	0.2	Yes
7	H5N3F1		NAS / NGT / SAN	[M+2H] ²⁺	928.8445	5.2	No	[M+2H] ²⁺	928.8441	5.6	No	[M+2H] ²⁺	928.8497	0.4	Yes
8	H6N3		NAS / NGT / SAN	[M+2H] ²⁺	936.8428	4.2	No	[M+2H] ²⁺	936.8421	5.0	No	[M+2H] ²⁺	936.8472	0.5	Yes
9	H4N5		NAS / NGT / SAN	[M+2H] ²⁺	977.8686	4.8	No	[M+2H] ²⁺	977.8696	3.8	No	[M+2H] ²⁺	977.8738	0.5	Yes
10	H6N3F1		NAS / NGT / SAN	[M+2H] ²⁺	1009.8710	4.7	No	[M+2H] ²⁺	1009.8712	4.5	No	[M+2H] ²⁺	1009.8765	0.8	Yes
11	H5N4F1S2		NAS / NGT / SAN									[M+3H] ³⁺	881.3239	1.9	Yes


















N-152/N-508 Peptide(s): NNSKPV / SNNSKPV / NNSKPVED / SNNSKPVED / SNNSKPVEDK / SNNSKPVEDKDA															
			CEA1 Human Colon Carcinoma					CEA2 Human Colon Carcinoma				CEA3 Human Liver Metastases			
Num ^a	Glycan Species ^b		Peptide Sequence(s):	Molecular ion	Observed <i>m/z</i>	Error (ppm) ^c	MS/MS ^d	Molecular ion	Observed <i>m/z</i>	Error (ppm) ^c	MS/MS ^d	Molecular ion	Observed <i>m/z</i>	Error (ppm) ^c	MS/MS ^d
1-6	H5N2		NNSKPV	[M+2H] ²⁺	937.8875	4.4	No	[M+2H] ²⁺	937.8889	3.0	Yes	[M+2H] ²⁺	937.8886	1.4	Yes
			SNNSKPV	[M+2H] ²⁺	981.4011	6.7	Yes	[M+2H] ²⁺	981.4015	6.3	Yes	[M+2H] ²⁺	981.4100	2.4	Yes
			NNSKPVED					[M+2H] ²⁺	1059.9214	4.7	No	[M+2H] ²⁺	1059.9300	3.4	Yes
			SNNSKPVED	[M+2H] ²⁺	1103.4371	4.4	No	[M+2H] ²⁺	1103.4352	6.2	No	[M+2H] ²⁺	1103.4443	2.1	Yes
			SNNSKPVEDK	[M+3H] ³⁺	778.6559	8.6	Yes	[M+3H] ³⁺	778.6576	6.4	Yes	[M+3H] ³⁺	778.6609	0.4	Yes
			SNNSKPVEDKDA	[M+3H] ³⁺	840.6794	5.4	Yes	[M+3H] ³⁺	840.6828	1.3	No	[M+3H] ³⁺	840.6859	2.3	No
7-12	H6N2		NNSKPV	[M+2H] ²⁺	1018.9122	5.8	No	[M+2H] ²⁺	1018.9132	4.8	No	[M+2H] ²⁺	1018.9185	0.4	No
			SNNSKPV	[M+2H] ²⁺	1062.4269	6.8	Yes	[M+2H] ²⁺	1062.4269	6.8	Yes	[M+2H] ²⁺	1062.4351	1.0	Yes
			NNSKPVED					[M+2H] ²⁺	1140.9482	4.1	No				
			SNNSKPVED	[M+2H] ²⁺	1184.4588	8.6	No	[M+2H] ²⁺	1184.4583	9.0	No	[M+2H] ²⁺	1184.4749	5.0	No
			SNNSKPVEDK	[M+3H] ³⁺	832.6756	5.5	Yes	[M+3H] ³⁺	832.6778	2.9	No	[M+3H] ³⁺	832.6780	0.7	Yes
			SNNSKPVEDKDA	[M+3H] ³⁺	894.6974	4.6	No	[M+3H] ³⁺	894.7000	1.7	No	[M+3H] ³⁺	894.7052	4.1	No
13-18	H7N2		NNSKPV	[M+2H] ²⁺	1099.9366	7.2	No	[M+2H] ²⁺	1099.9375	6.4	Yes	[M+2H] ²⁺	1099.9398	0.9	Yes
			SNNSKPV	[M+2H] ²⁺	1143.4541	5.6	Yes	[M+2H] ²⁺	1143.4530	6.6	Yes	[M+2H] ²⁺	1143.4610	0.4	Yes
			NNSKPVED					[M+2H] ²⁺	1221.9687	8.6	No				
			SNNSKPVED	[M+2H] ²⁺	1265.4812	10.9	Yes	[M+2H] ²⁺	1265.4830	9.5	Yes	[M+2H] ²⁺	1265.4853	2.3	Yes
			SNNSKPVEDK	[M+3H] ³⁺	886.6918	6.8	Yes	[M+3H] ³⁺	886.6964	1.6	No	[M+3H] ³⁺	886.6983	0.6	No
			SNNSKPVEDKDA	[M+3H] ³⁺	948.7169	2.4	No	[M+3H] ³⁺	948.7181	1.1	No	[M+3H] ³⁺	948.7225	3.5	No
19-24	H8N2		NNSKPV	[M+2H] ²⁺	1180.9646	5.3	No	[M+2H] ²⁺	1180.9624	7.2	Yes	[M+2H] ²⁺	1180.9671	2.5	Yes
			SNNSKPV	[M+2H] ²⁺	1224.4780	7.3	Yes	[M+2H] ²⁺	1224.4773	7.8	Yes	[M+2H] ²⁺	1224.4862	0.6	Yes
				[M+3H] ³⁺	816.6585	2.6	Yes	[M+3H] ³⁺	816.6576	3.7	Yes	[M+3H] ³⁺	816.6621	1.9	Yes
			NNSKPVED	[M+2H] ²⁺	1302.9945	8.6	No	[M+2H] ²⁺	1302.9930	9.7	No	[M+2H] ²⁺	1303.0039	1.3	Yes
				[M+3H] ³⁺	869.0036	3.2	No	[M+3H] ³⁺	869.0034	3.5	No	[M+3H] ³⁺	869.0070	0.7	No
			SNNSKPVED	[M+2H] ²⁺	1346.5117	7.6	Yes	[M+2H] ²⁺	1346.5089	9.7	Yes	[M+2H] ²⁺	1346.5122	1.0	Yes
				[M+3H] ³⁺	898.0130	4.5	Yes	[M+3H] ³⁺	898.0112	6.5	No	[M+3H] ³⁺	898.0172	0.2	Yes
			SNNSKPVEDK	[M+3H] ³⁺	940.7098	6.0	Yes	[M+3H] ³⁺	940.7071	8.8	Yes	[M+3H] ³⁺	940.7201	5.0	No
			SNNSKPVEDKDA	[M+3H] ³⁺	1002.7296	7.1	No	[M+3H] ³⁺	1002.7266	10.1	No	[M+3H] ³⁺	1002.7356	1.1	No

25-30	H9N2		NNSKPV	$[M+2H]^{2+}$	1261.9912	4.8	No	$[M+2H]^{2+}$	1261.9898	6.0	No	$[M+2H]^{2+}$	1261.9999	2.0	No
			SNNSKPV	$[M+2H]^{2+}$	1305.5023	8.4	Yes	$[M+2H]^{2+}$	1305.5013	9.2	Yes	$[M+2H]^{2+}$	1305.5165	2.4	Yes
				$[M+3H]^{3+}$	870.6734	5.5	Yes	$[M+3H]^{3+}$	870.6729	6.1	Yes	$[M+3H]^{3+}$	870.6809	3.1	No
			NNSKPVED	$[M+2H]^{2+}$	1384.0245	5.5	No	$[M+2H]^{2+}$	1384.0203	8.5	No	$[M+2H]^{2+}$	1384.0319	0.1	Yes
			SNNSKPVED	$[M+2H]^{2+}$	1427.5384	6.7	No	$[M+2H]^{2+}$	1427.5349	9.2	Yes	$[M+2H]^{2+}$	1427.5379	1.3	Yes
				$[M+3H]^{3+}$	952.0302	5.0	No	$[M+3H]^{3+}$	952.0278	7.6	No	$[M+3H]^{3+}$	952.0340	1.1	Yes
			SNNSKPVEDK	$[M+3H]^{3+}$	994.7303	2.7	Yes	$[M+3H]^{3+}$	994.7301	2.9	No	$[M+3H]^{3+}$	994.7330	0.03	No
			SNNSKPVEDKA	$[M+3H]^{3+}$	1056.7460	7.9	No	$[M+3H]^{3+}$	1056.7500	4.1	No	$[M+3H]^{3+}$	1056.7544	0.04	No


N-152/N-330/N-466/N-508 Peptide(s): SN															
			CEA1 Human Colon Carcinoma				CEA2 Human Colon Carcinoma				CEA3 Human Liver Metastases				
Num ^a	Glycan Species ^b		Peptide Sequence(s):	Molecular ion	Observed <i>m/z</i>	Error (ppm) ^c	MS/MS ^d	Molecular ion	Observed <i>m/z</i>	Error (ppm) ^c	MS/MS ^d	Molecular ion	Observed <i>m/z</i>	Error (ppm) ^c	MS/MS ^d
1	H5N2		SN									[M+2H] ²⁺	718.7636	2.1	Yes
2	H6N2		SN									[M+2H] ²⁺	799.7899	1.7	No



NR (N-197/N-309/N-375/N-553/N-580) / RN (N-208/N-665)															
			CEA1 Human Colon Carcinoma				CEA2 Human Colon Carcinoma				CEA3 Human Liver Metastases				
Num ^a	Glycan Species ^b		Peptide Sequence(s):	Molecular ion	Observed <i>m/z</i>	Error (ppm) ^c	MS/MS ^d	Molecular ion	Observed <i>m/z</i>	Error (ppm) ^c	MS/MS ^d	Molecular ion	Observed <i>m/z</i>	Error (ppm) ^c	MS/MS ^d
1	H7N6F4S1		NR / RN									[M+3H] ³⁺	1173.1175	0.3	Yes


GNR / NGNR (N-197/N-553) / GRN / GRNN (N-665)															
			CEA1 Human Colon Carcinoma				CEA2 Human Colon Carcinoma				CEA3 Human Liver Metastases				
Num ^a	Glycan Species ^b		Peptide Sequence(s):	Molecular ion	Observed <i>m/z</i>	Error (ppm) ^c	MS/MS ^d	Molecular ion	Observed <i>m/z</i>	Error (ppm) ^c	MS/MS ^d	Molecular ion	Observed <i>m/z</i>	Error (ppm) ^c	MS/MS ^d
1	H3N2		GNR / GRN	[M+2H] ²⁺	619.7533	2.1	Yes	[M+2H] ²⁺	619.7582	5.8	No	[M+2H] ²⁺	619.7560	2.3	No





2	H4N2		GNR / GRN	[M+2H] ²⁺	700.7822	1.7	Yes	[M+2H] ²⁺	700.7858	6.8	Yes	[M+2H] ²⁺	700.7857	6.7	No
3-4	H5N2		GNR / GRN	[M+2H] ²⁺	781.8040	4.4	Yes	[M+2H] ²⁺	781.8056	2.3	Yes	[M+2H] ²⁺	781.8068	0.8	Yes
			NGNR / GRNN	[M+2H] ²⁺	838.8250	4.6	Yes	[M+2H] ²⁺	838.8238	6.0	No	[M+2H] ²⁺	838.8316	3.3	Yes
5	H6N2		GNR / GRN	[M+2H] ²⁺	862.8310	3.3	Yes	[M+2H] ²⁺	862.8335	0.4	Yes	[M+2H] ²⁺	862.8355	1.9	Yes
6	H7N2		GNR / GRN	[M+2H] ²⁺	943.8557	4.8	No	[M+2H] ²⁺	943.8584	2.0	No	[M+2H] ²⁺	943.8609	0.7	No
7	H8N2		GNR / GRN	[M+2H] ²⁺	1024.8841	2.5	No					[M+2H] ²⁺	1024.8872	0.5	No
8	H3N3		GNR / GRN	[M+2H] ²⁺	721.2919	3.3	Yes	[M+2H] ²⁺	721.2915	3.9	Yes				
9	H4N3		GNR / GRN	[M+2H] ²⁺	802.3205	0.2	Yes	[M+2H] ²⁺	802.3225	2.2	Yes	[M+2H] ²⁺	802.3217	1.3	No
10	H5N3		GNR / GRN	[M+2H] ²⁺	883.3445	3.0	Yes	[M+2H] ²⁺	883.3455	1.8	Yes				
11	H4N4		GNR / GRN	[M+2H] ²⁺	903.8619	1.7	No	[M+2H] ²⁺	903.8612	0.9	Yes	[M+2H] ²⁺	903.8627	2.6	Yes
				[M+3H] ³⁺	602.9090	0.9	No								
12	H5N4		GNR / GRN	[M+2H] ²⁺	984.8813	5.6	No	[M+2H] ²⁺	984.8829	4.0	Yes	[M+2H] ²⁺	984.8887	1.9	Yes
				[M+3H] ³⁺	656.9245	4.1	Yes								
13	H4N5		GNR / GRN	[M+3H] ³⁺	670.6013	2.1	Yes	[M+3H] ³⁺	670.6019	1.2	Yes	[M+3H] ³⁺	670.6025	0.3	No
14	H6N3F1		GNR / GRN	[M+2H] ²⁺	1037.3960	6.2	No	[M+2H] ²⁺	1037.3980	4.3	No	[M+2H] ²⁺	1037.4046	2.0	Yes
				[M+3H] ³⁺	691.9332	6.4	No					[M+3H] ³⁺	691.9379	0.4	No
15	H6N4		GNR / GRN	[M+2H] ²⁺	1065.9076	5.3	No	[M+2H] ²⁺	1065.9090	3.9	No	[M+2H] ²⁺	1065.9170	3.6	No
				[M+3H] ³⁺	710.9416	4.5	No	[M+3H] ³⁺	710.9422	3.6	Yes	[M+3H] ³⁺	710.9459	1.6	No
16	H4N6		GNR / GRN	[M+2H] ²⁺	1106.9364	3.0	Yes	[M+2H] ²⁺	1106.9358	3.6	No	[M+2H] ²⁺	1106.9448	4.6	No
				[M+3H] ³⁺	738.2940	2.5	Yes	[M+3H] ³⁺	738.2934	3.3	Yes	[M+3H] ³⁺	738.2974	2.1	No
17	H4N7F2		GNR / GRN	[M+2H] ²⁺	1354.5248	9.3	No	[M+2H] ²⁺	1354.5280	6.9	No				
				[M+3H] ³⁺	903.3520	9.8	No	[M+3H] ³⁺	903.3535	8.2	No				
18	H5N7F2		GNR / GRN	[M+2H] ²⁺	1435.5548	6.2	Yes	[M+2H] ²⁺	1435.5540	6.8	No				
				[M+3H] ³⁺	957.3724	6.4	Yes	[M+3H] ³⁺	957.3714	7.4	Yes				
19	H7N7F2		GNR / GRN	[M+2H] ²⁺	1597.6169	0.2	No					[M+2H] ²⁺	1597.6185	1.2	No
				[M+3H] ³⁺	1065.4111	2.4	Yes	[M+3H] ³⁺	1065.4094	4.0	No	[M+3H] ³⁺	1065.4124	1.2	No










20	H4N3S1		GNR / GRN	[M+2H] ²⁺	947.8610	7.8	No	[M+2H] ²⁺	947.8598	9.1	No	[M+2H] ²⁺	947.8665	2.0	Yes
				[M+3H] ³⁺	632.2459	3.7	No	[M+3H] ³⁺	632.2470	2.0	No	[M+3H] ³⁺	632.2476	1.0	Yes
21	H5N3S1		GNR / GRN	[M+2H] ²⁺	1028.8872	7.4	No	[M+2H] ²⁺	1028.8878	6.8	No	[M+2H] ²⁺	1028.8940	0.8	Yes
22	H6N3S1		GNR / GRN	[M+3H] ³⁺	740.2783	7.0	Yes	[M+3H] ³⁺	740.2809	3.5	Yes	[M+3H] ³⁺	740.2835	0.1	No

GNRT (N-197/N-553) / SANR / ANRS (N-579) / TGRN (N-665)															
			CEA1 Human Colon Carcinoma				CEA2 Human Colon Carcinoma				CEA3 Human Liver Metastases				
Num ^a	Glycan Species ^b		Peptide Sequence(s):	Molecular ion	Observed <i>m/z</i>	Error (ppm) ^c	MS/MS ^d	Molecular ion	Observed <i>m/z</i>	Error (ppm) ^c	MS/MS ^d	Molecular ion	Observed <i>m/z</i>	Error (ppm) ^c	MS/MS ^d
1	H5N2		GNRT / SANR / ANRS / TGRN	[M+2H] ²⁺	832.3288	2.9	Yes	[M+2H] ²⁺	832.3300	1.5	Yes	[M+2H] ²⁺	832.3331	2.3	No

RND (N-208) / DNR (N-375)															
			CEA1 Human Colon Carcinoma				CEA2 Human Colon Carcinoma				CEA3 Human Liver Metastases				
Num ^a	Glycan Species ^b		Peptide Sequence(s):	Molecular ion	Observed m/z	Error (ppm) ^c	MS/MS ^d	Molecular ion	Observed m/z	Error (ppm) ^c	MS/MS ^d	Molecular ion	Observed m/z	Error (ppm) ^c	MS/MS ^d
1	H3N5		RND / DNR	[M+2H] ²⁺	953.3688	7.9	No	[M+2H] ²⁺	953.3705	6.1	Yes				
2	H7N6F3S1		RND / DNR									[M+3H] ³⁺	1162.7711	2.0	Yes

N-246/N-351/N-529 Peptide(s): NT															
			CEA1 Human Colon Carcinoma				CEA2 Human Colon Carcinoma				CEA3 Human Liver Metastases				
Num ^a	Glycan Species ^b		Peptide Sequence(s):	Molecular ion	Observed <i>m/z</i>	Error (ppm) ^c	MS/MS ^d	Molecular ion	Observed <i>m/z</i>	Error (ppm) ^c	MS/MS ^d	Molecular ion	Observed <i>m/z</i>	Error (ppm) ^c	MS/MS ^d
1	H5N2		NT	[M+2H] ²⁺	725.7675	3.4	No	[M+2H] ²⁺	725.7678	3.0	Yes	[M+2H] ²⁺	725.7714	2.0	Yes

NLN (N-256/N-612) / ANR (N-580)															
			CEA1 Human Colon Carcinoma				CEA2 Human Colon Carcinoma				CEA3 Human Liver Metastases				
Num ^a	Glycan Species ^b		Peptide Sequence(s):	Molecular ion	Observed m/z	Error (ppm) ^c	MS/MS ^d	Molecular ion	Observed m/z	Error (ppm) ^c	MS/MS ^d	Molecular ion	Observed m/z	Error (ppm) ^c	MS/MS ^d
1	H3N2F1		NLN / ANR	[M+2H] ²⁺	699.7898	5.8	Yes	[M+2H] ²⁺	699.7893	5.1	Yes				
2	H3N3F1		NLN / ANR	[M+2H] ²⁺	801.3280	3.2	Yes	[M+2H] ²⁺	801.3298	5.4	Yes				
3	H3N4F1		NLN / ANR	[M+2H] ²⁺	902.8666	1.6	Yes	[M+2H] ²⁺	902.8667	1.7	Yes				
4	H5N6F3		NLN / ANR	[M+2H] ²⁺	1414.0489	4.5	No	[M+2H] ²⁺	1414.0492	4.3	No				
				[M+3H] ³⁺	943.0353	4.4	Yes	[M+3H] ³⁺	943.0353	4.4	No				

N-351/N-529 Peptide(s): NTT / QNT / QNTT															
			CEA1 Human Colon Carcinoma				CEA2 Human Colon Carcinoma				CEA3 Human Liver Metastases				
Num ^a	Glycan Species ^b		Peptide Sequence(s):	Molecular ion	Observed <i>m/z</i>	Error (ppm) ^c	MS/MS ^d	Molecular ion	Observed <i>m/z</i>	Error (ppm) ^c	MS/MS ^d	Molecular ion	Observed <i>m/z</i>	Error (ppm) ^c	MS/MS ^d
1	H4N2		NTT	[M+2H] ²⁺	695.2637	5.3	Yes	[M+2H] ²⁺	695.2646	4.0	No	[M+2H] ²⁺	695.2680	0.9	Yes
2-4	H5N2		NTT	[M+2H] ²⁺	776.2901	4.7	Yes	[M+2H] ²⁺	776.2899	5.0	Yes	[M+2H] ²⁺	776.2930	1.0	Yes
			QNT	[M+2H] ²⁺	789.7987	0.7	Yes	[M+2H] ²⁺	789.7982	1.3	Yes	[M+2H] ²⁺	789.8005	1.6	Yes
			QNTT	[M+2H] ²⁺	840.3197	4.0	Yes	[M+2H] ²⁺	840.3196	4.1	Yes	[M+2H] ²⁺	840.3244	1.6	Yes
5-7	H6N2		NTT	[M+2H] ²⁺	857.3166	4.2	No	[M+2H] ²⁺	857.3166	4.2	No	[M+2H] ²⁺	857.3209	0.8	No
			QNT									[M+2H] ²⁺	870.8288	3.6	Yes
			QNTT	[M+2H] ²⁺	921.3454	4.4	No	[M+2H] ²⁺	921.3461	3.7	No				
8-10	H7N2		NTT	[M+2H] ²⁺	938.3421	4.8	No	[M+2H] ²⁺	938.3415	5.4	No	[M+2H] ²⁺	938.3472	0.6	No
			QNT									[M+2H] ²⁺	951.8570	5.2	Yes
			QNTT	[M+2H] ²⁺	1002.3816	5.7	No	[M+2H] ²⁺	1002.3814	5.5	No				
11	H8N2		NTT	[M+2H] ²⁺	1019.3686	4.3	No	[M+2H] ²⁺	1019.3690	3.9	No				
12	H3N2F1		QNTT	[M+2H] ²⁺	751.2950	5.6	Yes	[M+2H] ²⁺	751.2961	4.1	Yes	[M+2H] ²⁺	751.3002	1.3	No
13	H3N3F1		QNTT	[M+2H] ²⁺	852.8350	4.6	Yes	[M+2H] ²⁺	852.8352	4.3	Yes				
14	H4N3F1		QNTT	[M+2H] ²⁺	933.8588	7.0	Yes	[M+2H] ²⁺	933.8589	6.9	No	[M+2H] ²⁺	933.8640	1.4	No
15	H5N3		NTT					[M+2H] ²⁺	877.8292	4.8	Yes				















16	H4N3F2		NTT									[M+2H] ²⁺	942.8644	0.6	Yes
17	H4N3F3		NTT									[M+2H] ²⁺	1015.8939	0.01	No
18	H6N3F1		NTT					[M+2H] ²⁺	1031.8835	5.2	No	[M+2H] ²⁺	1031.8899	1.0	Yes
19-20	H5N4F1		NTT	[M+2H] ²⁺	1052.3972	4.7	Yes	[M+2H] ²⁺	1052.3954	6.4	No	[M+2H] ²⁺	1052.4027	0.6	No
			QNTT	[M+2H] ²⁺	1116.4241	6.5	No	[M+2H] ²⁺	1116.4231	7.4	No	[M+2H] ²⁺	1116.4317	0.3	Yes
21	H5N6F3		QNT	[M+2H] ²⁺	1415.0444	0.3	No	[M+2H] ²⁺	1415.0425	1.6	No				
				[M+3H] ³⁺	943.6994	0.2	No	[M+3H] ³⁺	943.6994	0.2	Yes				
22	H4N3F1S1		NTT									[M+2H] ²⁺	1015.3847	1.0	Yes
23	H5N4S2		QNT									[M+3H] ³⁺	856.3181	0.6	Yes

















-190-







ANNS / NNSA (N-480) / NNGT (N-650)															
			CEA1 Human Colon Carcinoma				CEA2 Human Colon Carcinoma				CEA3 Human Liver Metastases				
Num ^a	Glycan Species ^b		Peptide Sequence(s):	Molecular ion	Observed <i>m/z</i>	Error (ppm) ^c	MS/MS ^d	Molecular ion	Observed <i>m/z</i>	Error (ppm) ^c	MS/MS ^d	Molecular ion	Observed <i>m/z</i>	Error (ppm) ^c	MS/MS ^d
1	H5N2		ANNS / NNSA / NNGT	[M+2H] ²⁺	811.3002	2.4	Yes	[M+2H] ²⁺	811.3009	1.6	Yes	[M+2H] ²⁺	811.3044	2.7	Yes
2	H3N2F1		ANNS / NNSA / NNGT	[M+2H] ²⁺	722.2749	4.7	Yes	[M+2H] ²⁺	722.2759	3.3	No	[M+2H] ²⁺	722.2828	6.2	No
3	H3N3F1		ANNS / NNSA / NNGT	[M+2H] ²⁺	823.8151	3.5	Yes	[M+2H] ²⁺	823.8154	3.1	Yes	[M+2H] ²⁺	823.8188	1.0	Yes
4	H4N3F1		ANNS / NNSA / NNGT	[M+2H] ²⁺	904.8394	5.5	Yes	[M+2H] ²⁺	904.8396	5.3	No	[M+2H] ²⁺	904.8453	1.0	Yes
5	H3N4F1		ANNS / NNSA / NNGT	[M+2H] ²⁺	925.3521	6.0	Yes	[M+2H] ²⁺	925.3521	6.0	Yes	[M+2H] ²⁺	925.3561	1.7	Yes
6	H5N3F1		ANNS / NNSA / NNGT	[M+2H] ²⁺	985.8659	5.0	Yes	[M+2H] ²⁺	985.8665	4.4	No				
7	H6N6F2		ANNS / NNSA / NNGT	[M+2H] ²⁺	1444.5463	0.7	No					[M+2H] ²⁺	1444.5496	3.0	No
				[M+3H] ³⁺	963.3651	1.1	No	[M+3H] ³⁺	963.3681	2.0	No	[M+3H] ³⁺	963.3641	2.1	Yes
8	H6N6F3		ANNS / NNSA / NNGT	[M+2H] ²⁺	1517.5676	4.3	No					[M+2H] ²⁺	1517.5782	2.6	No
				[M+3H] ³⁺	1012.0470	5.0	No	[M+3H] ³⁺	1012.0494	2.7	No	[M+3H] ³⁺	1012.0507	1.4	Yes

9	H6N5F2S2		ANNS / NNSA / NNGT									[M+3H] ³⁺	1089.7356	0.9	Yes
10	H6N5F3S2		ANNS / NNSA / NNGT									[M+3H] ³⁺	1138.4222	0.3	Yes
















SNN (N-152/N-330/N-466/N-508) / NNS (N-292/N-480/N-665)															
			CEA1 Human Colon Carcinoma					CEA2 Human Colon Carcinoma				CEA3 Human Liver Metastases			
Num ^a	Glycan Species ^b		Peptide Sequence(s):	Molecular ion	Observed m/z	Error (ppm) ^c	MS/MS ^d	Molecular ion	Observed m/z	Error (ppm) ^c	MS/MS ^d	Molecular ion	Observed m/z	Error (ppm) ^c	MS/MS ^d
1	H3N2F1		SNN / NNS	[M+2H] ²⁺	686.7586	1.6	Yes	[M+2H] ²⁺	686.7585	1.8	Yes				
2	H3N3		SNN / NNS	[M+2H] ²⁺	715.2730	3.6	No	[M+2H] ²⁺	715.2707	0.4	No				
3	H3N4F1		SNN / NNS	[M+2H] ²⁺	889.8442	5.8	Yes	[M+2H] ²⁺	889.8430	4.4	Yes				
4	H4N4F1		SNN / NNS	[M+2H] ²⁺	970.8681	2.7	Yes	[M+2H] ²⁺	970.8664	0.9	Yes				
				[M+3H] ³⁺	647.5812	2.4	No								
5	H3N5F1		SNN / NNS	[M+2H] ²⁺	991.3832	4.5	Yes	[M+2H] ²⁺	991.3821	3.4	Yes				
				[M+3H] ³⁺	661.2579	4.2	No								
6	H4N5F1		SNN / NNS	[M+2H] ²⁺	1072.4077	2.4	Yes	[M+2H] ²⁺	1072.4059	0.7	Yes	[M+2H] ²⁺	1072.4071	1.8	No
				[M+3H] ³⁺	715.2753	3.5	Yes	[M+3H] ³⁺	715.2836	15.2	Yes				
7	H4N5F2		SNN / NNS	[M+2H] ²⁺	1145.4339	0.2	No	[M+2H] ²⁺	1145.4346	0.4	No	[M+2H] ²⁺	1145.4411	6.1	No
				[M+3H] ³⁺	763.9598	1.4	Yes	[M+3H] ³⁺	763.9593	0.7	Yes	[M+3H] ³⁺	763.9605	2.3	Yes
8	H5N5F1		SNN / NNS	[M+2H] ²⁺	1153.4334	1.6	No	[M+2H] ²⁺	1153.4316	0.01	No	[M+2H] ²⁺	1153.4380	5.6	No
				[M+3H] ³⁺	769.2922	2.4	Yes	[M+3H] ³⁺	769.2907	0.4	Yes	[M+3H] ³⁺	769.2915	1.5	No
9	H4N6F1		SNN / NNS	[M+2H] ²⁺	1173.9478	2.5	No	[M+2H] ²⁺	1173.9468	1.7	No				
				[M+3H] ³⁺	782.9699	5.1	Yes	[M+3H] ³⁺	782.9668	1.2	Yes				
10	H5N5F2		SNN / NNS	[M+2H] ²⁺	1226.4600	0.4	No	[M+2H] ²⁺	1226.4596	0.8	No	[M+2H] ²⁺	1226.4632	2.2	No
				[M+3H] ³⁺	817.9769	0.7	Yes	[M+3H] ³⁺	817.9760	0.4	Yes	[M+3H] ³⁺	817.9775	1.4	Yes
11	H5N6F1		SNN / NNS	[M+2H] ²⁺	1254.9740	2.2	No	[M+2H] ²⁺	1254.9733	1.6	No				
				[M+3H] ³⁺	836.9870	4.2	Yes	[M+3H] ³⁺	836.9855	2.4	Yes	[M+3H] ³⁺	836.9875	4.8	No
12	H5N5F3		SNN / NNS	[M+2H] ²⁺	1299.4872	1.8	No	[M+2H] ²⁺	1299.4872	1.8	No	[M+2H] ²⁺	1299.4912	1.3	No
				[M+3H] ³⁺	866.6613	1.2	Yes	[M+3H] ³⁺	866.6609	1.6	Yes	[M+3H] ³⁺	866.6630	0.8	Yes
















13	H5N6F2		SNN / NNS	[M+2H] ²⁺	1328.0036	2.5	No	[M+2H] ²⁺	1328.0017	1.1	No				
				[M+3H] ³⁺	885.6716	2.4	Yes	[M+3H] ³⁺	885.6704	1.1	Yes	[M+3H] ³⁺	885.6704	1.1	No
14	H6N6F1		SNN / NNS	[M+2H] ²⁺	1336.0045	5.1	No	[M+2H] ²⁺	1336.0015	2.9	No				
				[M+3H] ³⁺	891.0048	4.2	No	[M+3H] ³⁺	891.0033	2.5	No	[M+3H] ³⁺	891.0034	2.6	Yes
15	H5N5F4		SNN / NNS	[M+2H] ²⁺	1372.5191	0.5	No	[M+2H] ²⁺	1372.5182	0.2	No	[M+2H] ²⁺	1372.5218	2.4	No
				[M+3H] ³⁺	915.3475	0.9	No	[M+3H] ³⁺	915.3469	1.5	No	[M+3H] ³⁺	915.3477	0.6	Yes
16	H5N6F3		SNN / NNS	[M+3H] ³⁺	934.3579	2.6	Yes	[M+3H] ³⁺	934.3560	0.6	Yes	[M+3H] ³⁺	934.3574	2.1	No
17	H6N6F2		SNN / NNS	[M+3H] ³⁺	939.6905	3.6	Yes	[M+3H] ³⁺	939.6887	1.7	Yes	[M+3H] ³⁺	939.6880	1.0	Yes
18	H6N6F3		SNN / NNS	[M+3H] ³⁺	988.3745	1.5	Yes	[M+3H] ³⁺	988.3727	0.3	Yes	[M+3H] ³⁺	988.3741	1.1	Yes
19	H5N7F3		SNN / NNS	[M+3H] ³⁺	1002.0514	2.8	No	[M+3H] ³⁺	1002.0484	0.2	Yes	[M+3H] ³⁺	1002.0507	2.1	No
20	H6N6F4		SNN / NNS	[M+3H] ³⁺	1037.0595	0.5	Yes	[M+3H] ³⁺	1037.0606	1.5	Yes	[M+3H] ³⁺	1037.0636	4.4	Yes
21	H6N7F3		SNN / NNS	[M+3H] ³⁺	1056.0685	2.2	Yes	[M+3H] ³⁺	1056.0694	3.1	Yes	[M+3H] ³⁺	1056.0698	3.4	No
22	H7N7F2		SNN / NNS	[M+3H] ³⁺	1061.4004	2.4	No	[M+3H] ³⁺	1061.4020	4.0	No	[M+3H] ³⁺	1061.3998	1.9	Yes
23	H6N6F5		SNN / NNS	[M+3H] ³⁺	1085.7442	0.7	No	[M+3H] ³⁺	1085.7460	0.9	No	[M+3H] ³⁺	1085.7494	4.1	Yes
24	H7N6F4		SNN / NNS	[M+3H] ³⁺	1091.0783	1.5	No	[M+3H] ³⁺	1091.0814	4.4	No	[M+3H] ³⁺	1091.0798	2.9	Yes
25	H6N7F4		SNN / NNS	[M+3H] ³⁺	1104.7525	0.3	Yes	[M+3H] ³⁺	1104.7538	1.5	Yes	[M+3H] ³⁺	1104.7521	0.03	No
26	H7N7F3		SNN / NNS	[M+3H] ³⁺	1110.0863	2.3	Yes	[M+3H] ³⁺	1110.0881	3.9	Yes	[M+3H] ³⁺	1110.0863	2.3	Yes

27	H7N6F5		SNN / NNS	$[M+3H]^{3+}$	1139.7646	1.8	No	$[M+3H]^{3+}$	1139.7671	4.0	No	$[M+3H]^{3+}$	1139.7646	1.8	Yes
28	H6N7F5		SNN / NNS	$[M+3H]^{3+}$	1153.4371	0.9	Yes	$[M+3H]^{3+}$	1153.4374	0.6	Yes	$[M+3H]^{3+}$	1153.4421	3.5	No
29	H7N7F4		SNN / NNS	$[M+3H]^{3+}$	1158.7725	2.4	Yes	$[M+3H]^{3+}$	1158.7731	2.9	Yes	$[M+3H]^{3+}$	1158.7719	1.9	Yes
30	H7N6F6		SNN / NNS	$[M+3H]^{3+}$	1188.4507	1.8	No	$[M+3H]^{3+}$	1188.4548	5.3	No	$[M+3H]^{3+}$	1188.4490	0.4	Yes
31	H7N7F5		SNN / NNS	$[M+3H]^{3+}$	1207.4591	2.8	Yes	$[M+3H]^{3+}$	1207.4593	3.0	No	$[M+3H]^{3+}$	1207.4584	2.2	Yes
32	H7N7F6		SNN / NNS	$[M+3H]^{3+}$	1256.1449	2.6	No	$[M+3H]^{3+}$	1256.1401	1.3	No	$[M+3H]^{3+}$	1256.1444	2.2	Yes
33	H8N7F5		SNN / NNS	$[M+3H]^{3+}$	1261.4778	3.6	No	$[M+3H]^{3+}$	1261.4745	0.9	No	$[M+3H]^{3+}$	1261.4778	3.6	Yes
34	H7N8F5		SNN / NNS	$[M+3H]^{3+}$	1275.1523	2.7	Yes	$[M+3H]^{3+}$	1275.1482	0.5	Yes	$[M+3H]^{3+}$	1275.1568	6.2	No
35	H8N7F6		SNN / NNS									$[M+3H]^{3+}$	1310.1654	4.7	Yes
36	H7N8F6		SNN / NNS	$[M+3H]^{3+}$	1323.8372	1.8	Yes	$[M+3H]^{3+}$	1323.8323	1.9	Yes	$[M+3H]^{3+}$	1323.8413	4.9	No
37	H8N8F5		SNN / NNS									$[M+3H]^{3+}$	1329.1738	5.5	Yes
38	H8N8F6		SNN / NNS									$[M+3H]^{3+}$	1377.8589	4.7	Yes
39	H9N8F6		SNN / NNS									$[M+3H]^{3+}$	1431.8784	5.9	Yes
40	H4N4F1S1		SNN / NNS	$[M+3H]^{3+}$	744.6126	1.6	No	$[M+3H]^{3+}$	744.6119	0.6	No				
41	H5N5F1S1		SNN / NNS	$[M+3H]^{3+}$	866.3230	1.0	Yes	$[M+3H]^{3+}$	866.3211	1.2	No				
42	H5N5F2S1		SNN / NNS	$[M+3H]^{3+}$	915.0086	0.5	Yes	$[M+3H]^{3+}$	915.0070	1.2	No	$[M+3H]^{3+}$	915.0112	3.3	Yes

43	H7N6F3S1		SNN / NNS					[M+3H] ³⁺	1139.4170	4.8	No	[M+3H] ³⁺	1139.4259	3.0	Yes
44	H7N6F4S1		SNN / NNS	[M+3H] ³⁺	1188.1088	0.3	No					[M+3H] ³⁺	1188.1121	3.1	No
45	H7N6F5S1		SNN / NNS									[M+3H] ³⁺	1236.7970	2.1	Yes
46	H5N4F1S2		SNN / NNS									[M+3H] ³⁺	895.6599	1.1	Yes
47	H5N4F2S2		SNN / NNS									[M+3H] ³⁺	944.3448	2.1	Yes
48	H7N6F4S2		SNN / NNS									[M+3H] ³⁺	1285.1449	3.6	Yes


















NVT (N-204/N-560) / NLS (N-432/N-612) / TVN (N-292) / ISN (N-466)															
			CEA1 Human Colon Carcinoma					CEA2 Human Colon Carcinoma				CEA3 Human Liver Metastases			
Num ^a	Glycan Species ^b		Peptide Sequence(s):	Molecular ion	Observed <i>m/z</i>	Error (ppm) ^c	MS/MS ^d	Molecular ion	Observed <i>m/z</i>	Error (ppm) ^c	MS/MS ^d	Molecular ion	Observed <i>m/z</i>	Error (ppm) ^c	MS/MS ^d
1	H3N2F1		NVT / NLS / TVN / ISN	[M+2H] ²⁺	686.2697	15.4	Yes	[M+2H] ²⁺	686.2743	8.7	Yes				
2	H3N3		NVT / NLS / TVN / ISN	[M+2H] ²⁺	714.7862	6.8	No	[M+2H] ²⁺	714.7847	8.9	No				
3	H3N4F1		NVT / NLS / TVN / ISN	[M+2H] ²⁺	889.3534	7.1	Yes	[M+2H] ²⁺	889.3535	6.9	Yes				
4	H4N4F1		NVT / NLS / TVN / ISN	[M+2H] ²⁺	970.3775	8.8	Yes	[M+2H] ²⁺	970.3766	9.8	Yes				
			NVT / NLS / TVN / ISN	[M+3H] ³⁺	647.2551	7.6	No								
5	H3N5F1		NVT / NLS / TVN / ISN	[M+2H] ²⁺	990.8918	7.6	Yes	[M+2H] ²⁺	990.8914	8.0	Yes				
			NVT / NLS / TVN / ISN	[M+3H] ³⁺	660.9312	6.6	No								
6	H4N5F1		NVT / NLS / TVN / ISN	[M+2H] ²⁺	1071.9178	7.4	Yes	[M+2H] ²⁺	1071.9173	7.9	Yes				
			NVT / NLS / TVN / ISN	[M+3H] ³⁺	714.9491	5.7	Yes								
7	H4N5F2		NVT / NLS / TVN / ISN	[M+2H] ²⁺	1144.9450	8.5	No	[M+2H] ²⁺	1144.9444	9.0	No	[M+2H] ²⁺	1144.9428	10.4	No
			NVT / NLS / TVN / ISN	[M+3H] ³⁺	763.6371	2.7	Yes				[M+3H] ³⁺	763.6320	9.3	Yes	
8	H5N5F1		NVT / NLS / TVN / ISN	[M+2H] ²⁺	1152.9416	9.2	No	[M+2H] ²⁺	1152.9412	9.5	No	[M+2H] ²⁺	1152.9474	4.1	No
			NVT / NLS / TVN / ISN	[M+3H] ³⁺	768.9667	5.3	Yes	[M+3H] ³⁺	768.9626	10.6	Yes	[M+3H] ³⁺	768.9634	9.6	No
















9	H4N6F1		NVT / NLS / TVN / ISN	[M+2H] ²⁺	1173.4536	10.1	No	[M+2H] ²⁺	1173.4541	9.7	No				
			NVT / NLS / TVN / ISN	[M+3H] ³⁺	782.6416	6.0	Yes	[M+3H] ³⁺	782.6356	13.7	Yes				
10	H5N5F2		NVT / NLS / TVN / ISN	[M+2H] ²⁺	1225.9711	8.2	No	[M+2H] ²⁺	1225.9714	7.9	No	[M+2H] ²⁺	1225.9768	3.5	No
			NVT / NLS / TVN / ISN	[M+3H] ³⁺	817.6527	4.9	Yes	[M+3H] ³⁺	817.6523	5.4	Yes	[M+3H] ³⁺	817.6551	2.0	Yes
11	H5N6F1		NVT / NLS / TVN / ISN	[M+2H] ²⁺	1254.4781	11.0	No	[M+2H] ²⁺	1254.4800	9.5	No				
			NVT / NLS / TVN / ISN	[M+3H] ³⁺	836.6596	5.1	Yes	[M+3H] ³⁺	836.6585	6.4	Yes				
12	H5N5F3		NVT / NLS / TVN / ISN	[M+2H] ²⁺	1298.9984	9.0	No	[M+2H] ²⁺	1298.9977	9.5	No	[M+2H] ²⁺	1299.0065	2.8	Yes
			NVT / NLS / TVN / ISN	[M+3H] ³⁺	866.3363	7.4	Yes	[M+3H] ³⁺	866.3365	7.2	Yes	[M+3H] ³⁺	866.3403	2.8	Yes
13	H5N6F2		NVT / NLS / TVN / ISN	[M+2H] ²⁺	1327.5097	8.4	No	[M+2H] ²⁺	1327.5090	8.9	No				
			NVT / NLS / TVN / ISN	[M+3H] ³⁺	885.3434	7.3	Yes	[M+3H] ³⁺	885.3438	6.8	Yes				
14	H6N6F1		NVT / NLS / TVN / ISN	[M+2H] ²⁺	1335.5055	9.6	No	[M+2H] ²⁺	1335.5048	10.1	No	[M+2H] ²⁺	1335.5135	3.6	No
			NVT / NLS / TVN / ISN	[M+3H] ³⁺	890.6720	10.7	No	[M+3H] ³⁺	890.6739	8.5	No	[M+3H] ³⁺	890.6747	7.6	Yes
15	H5N5F4		NVT / NLS / TVN / ISN	[M+2H] ²⁺	1372.0243	10.7	No					[M+2H] ²⁺	1372.0298	6.7	No
			NVT / NLS / TVN / ISN	[M+3H] ³⁺	915.0192	10.4	No	[M+3H] ³⁺	915.0191	10.5	No	[M+3H] ³⁺	915.0257	3.3	Yes
16	H5N6F3		NVT / NLS / TVN / ISN	[M+3H] ³⁺	934.0285	7.9	Yes	[M+3H] ³⁺	934.0274	9.0	Yes	[M+3H] ³⁺	934.0297	6.6	No
17	H6N6F2		NVT / NLS / TVN / ISN	[M+3H] ³⁺	939.3624	5.4	Yes	[M+3H] ³⁺	939.3604	7.5	Yes	[M+3H] ³⁺	939.3619	5.9	Yes
18	H6N6F3		NVT / NLS / TVN / ISN	[M+3H] ³⁺	988.0478	5.7	Yes	[M+3H] ³⁺	988.0457	7.8	Yes	[M+3H] ³⁺	988.0508	2.7	Yes
19	H5N7F3		NVT / NLS / TVN / ISN	[M+3H] ³⁺	1001.7198	9.1	No	[M+3H] ³⁺	1001.7178	11.1	Yes				
20	H6N6F4		NVT / NLS / TVN / ISN	[M+3H] ³⁺	1036.7285	10.5	Yes	[M+3H] ³⁺	1036.7298	9.3	Yes	[M+3H] ³⁺	1036.7331	6.1	Yes
21	H6N7F3		NVT / NLS / TVN / ISN	[M+3H] ³⁺	1055.7372	8.9	Yes	[M+3H] ³⁺	1055.7372	8.9	Yes	[M+3H] ³⁺	1055.7351	10.9	No
22	H7N7F2		NVT / NLS / TVN / ISN	[M+3H] ³⁺	1061.0680	9.6	No	[M+3H] ³⁺	1061.0711	6.7	No	[M+3H] ³⁺	1061.0725	5.4	Yes
23	H6N6F5		NVT / NLS / TVN / ISN	[M+3H] ³⁺	1085.4117	12.6	No	[M+3H] ³⁺	1085.4141	10.4	No	[M+3H] ³⁺	1085.4207	4.3	Yes

24	H7N6F4		NVT / NLS / TVN / ISN	[M+3H] ³⁺	1090.7486	7.7	No	[M+3H] ³⁺	1090.7521	4.5	No	[M+3H] ³⁺	1090.7495	6.9	Yes
25	H6N7F4		NVT / NLS / TVN / ISN	[M+3H] ³⁺	1104.4219	9.6	Yes	[M+3H] ³⁺	1104.4225	9.1	Yes	[M+3H] ³⁺	1104.4251	6.7	No
26	H7N7F3		NVT / NLS / TVN / ISN	[M+3H] ³⁺	1109.7536	9.5	Yes	[M+3H] ³⁺	1109.7544	8.8	Yes	[M+3H] ³⁺	1109.7572	6.3	Yes
27	H7N6F5		NVT / NLS / TVN / ISN	[M+3H] ³⁺	1139.4345	7.4	No	[M+3H] ³⁺	1139.4345	7.4	No	[M+3H] ³⁺	1139.4378	4.6	Yes
28	H6N7F5		NVT / NLS / TVN / ISN	[M+3H] ³⁺	1153.1062	10.7	Yes	[M+3H] ³⁺	1153.1065	10.4	Yes	[M+3H] ³⁺	1153.1129	4.9	No
29	H7N7F4		NVT / NLS / TVN / ISN	[M+3H] ³⁺	1158.4410	7.9	Yes	[M+3H] ³⁺	1158.4420	7.0	Yes	[M+3H] ³⁺	1158.4459	3.7	Yes
30	H7N6F6		NVT / NLS / TVN / ISN	[M+3H] ³⁺	1188.1205	7.1	No					[M+3H] ³⁺	1188.1258	2.7	Yes
31	H7N7F5		NVT / NLS / TVN / ISN	[M+3H] ³⁺	1207.1279	6.8	Yes	[M+3H] ³⁺	1207.1276	7.1	No	[M+3H] ³⁺	1207.1335	2.2	Yes
32	H7N7F6		NVT / NLS / TVN / ISN	[M+3H] ³⁺	1255.8120	8.0	No	[M+3H] ³⁺	1255.8067	12.2	No	[M+3H] ³⁺	1255.8126	7.5	Yes
33	H8N7F5		NVT / NLS / TVN / ISN					[M+3H] ³⁺	1261.1465	5.7	No	[M+3H] ³⁺	1261.1500	2.9	Yes
34	H7N8F5		NVT / NLS / TVN / ISN	[M+3H] ³⁺	1274.8200	7.2	Yes	[M+3H] ³⁺	1274.8149	11.2	Yes				
35	H8N7F6		NVT / NLS / TVN / ISN									[M+3H] ³⁺	1309.8336	4.6	Yes
36	H7N8F6		NVT / NLS / TVN / ISN	[M+3H] ³⁺	1323.5054	7.4	Yes	[M+3H] ³⁺	1323.5008	10.9	Yes	[M+3H] ³⁺	1323.5061	6.9	No
37	H8N8F5		NVT / NLS / TVN / ISN									[M+3H] ³⁺	1328.8404	4.8	Yes
38	H8N8F6		NVT / NLS / TVN / ISN									[M+3H] ³⁺	1377.5295	2.4	Yes

39	H9N8F6		NVT / NLS / TVN / ISN									[M+3H] ³⁺	1431.5474	2.1	Yes
40	H4N4F1S1		NVT / NLS / TVN / ISN	[M+3H] ³⁺	744.2860	7.8	No	[M+3H] ³⁺	744.2870	6.5	No				
41	H5N5F1S1		NVT / NLS / TVN / ISN	[M+3H] ³⁺	865.9979	5.4	Yes	[M+3H] ³⁺	865.9970	6.4	No				
42	H5N5F2S1		NVT / NLS / TVN / ISN	[M+3H] ³⁺	914.6816	7.6	Yes	[M+3H] ³⁺	914.6808	8.5	No	[M+3H] ³⁺	914.6873	1.4	Yes
43	H7N6F3S1		NVT / NLS / TVN / ISN					[M+3H] ³⁺	1139.0956	6.4	No	[M+3H] ³⁺	1139.0999	2.6	Yes
44	H7N6F4S1		NVT / NLS / TVN / ISN	[M+3H] ³⁺	1187.7711	14.9	No					[M+3H] ³⁺	1187.7844	3.7	Yes
45	H7N6F5S1		NVT / NLS / TVN / ISN									[M+3H] ³⁺	1236.4713	2.8	Yes
46	H5N4F1S2		NVT / NLS / TVN / ISN									[M+3H] ³⁺	895.3288	13.9	Yes
47	H5N4F2S2		NVT / NLS / TVN / ISN									[M+3H] ³⁺	944.01800	9.8	Yes
48	H7N6F4S2		NVT / NLS / TVN / ISN									[M+3H] ³⁺	1284.8105	7.9	Yes

All N-Glycosylation sites Peptide(s): N														
			CEA1 Human Colon Carcinoma				CEA2 Human Colon Carcinoma				CEA3 Human Liver Metastases			
Num ^a	Glycan Species ^b	Peptide Sequence(s):	Molecular ion	Observed <i>m/z</i>	Error (ppm) ^c	MS/MS ^d	Molecular ion	Observed <i>m/z</i>	Error (ppm) ^c	MS/MS ^d	Molecular ion	Observed <i>m/z</i>	Error (ppm) ^c	MS/MS ^d
1	H4N2		N								[M+2H] ²⁺	594.2210	2.2	Yes
2	H5N2		N								[M+2H] ²⁺	675.2472	1.6	Yes
3	H6N2		N								[M+2H] ²⁺	756.2740	1.9	No
4	H7N2		N								[M+2H] ²⁺	837.3009	2.3	Yes
5	H8N2		N								[M+2H] ²⁺	918.326	0.7	Yes

6	H9N2		N									[M+2H] ²⁺	999.3524	0.6	Yes
7	H3N3F1		N	[M+2H] ²⁺	687.7606	1.9	Yes	[M+2H] ²⁺	687.7605	2.1	Yes	[M+2H] ²⁺	687.7628	1.3	No
8	H4N3F1		N	[M+2H] ²⁺	768.7862	2.8	Yes	[M+2H] ²⁺	768.7853	4.0	Yes	[M+2H] ²⁺	768.7897	1.8	Yes
9	H3N4F1		N	[M+2H] ²⁺	789.2997	2.4	Yes	[M+2H] ²⁺	789.2989	3.5	Yes	[M+2H] ²⁺	789.3039	2.9	No
				[M+3H] ³⁺	526.5338	6.2	No								
10	H5N3F1		N	[M+2H] ²⁺	849.8115	3.8	Yes	[M+2H] ²⁺	849.8110	4.4	Yes	[M+2H] ²⁺	849.8148	0.05	No
11	H4N4F1		N	[M+2H] ²⁺	870.3241	4.5	Yes	[M+2H] ²⁺	870.3232	5.6	Yes	[M+2H] ²⁺	870.3295	1.7	No
				[M+3H] ³⁺	580.5527	3.4	No								
12	H3N5F1		N	[M+2H] ²⁺	890.8374	4.4	Yes	[M+2H] ²⁺	890.8359	6.1	Yes	[M+2H] ²⁺	890.8427	1.6	No
				[M+3H] ³⁺	594.2257	7.6	No								
13	H4N4F2		N	[M+2H] ²⁺	943.3513	6.0	Yes	[M+2H] ²⁺	943.3504	7.0	Yes	[M+2H] ²⁺	943.3579	1.0	No
				[M+3H] ³⁺	629.2392	2.3	No	[M+3H] ³⁺	629.2379	4.4	No				
14	H4N5F1		N	[M+2H] ²⁺	971.8616	6.3	Yes	[M+2H] ²⁺	971.8606	7.3	Yes	[M+2H] ²⁺	971.8686	0.9	No
				[M+3H] ³⁺	648.2452	4.0	No	[M+3H] ³⁺	648.2450	4.3	Yes				
15	H4N5F2		N	[M+2H] ²⁺	1044.8896	6.8	Yes	[M+2H] ²⁺	1044.8884	7.9	Yes	[M+2H] ²⁺	1044.8972	0.5	No
				[M+3H] ³⁺	696.9307	4.4	No	[M+3H] ³⁺	696.9302	5.1	No				
16	H5N5F2		N	[M+2H] ²⁺	1125.9147	7.5	Yes	[M+2H] ²⁺	1125.9133	8.7	Yes	[M+2H] ²⁺	1125.9220	1.0	No
				[M+3H] ³⁺	750.9482	4.2	No	[M+3H] ³⁺	750.9476	5.0	No	[M+3H] ³⁺	750.9516	0.3	Yes
17	H5N5F3		N	[M+2H] ²⁺	1198.9426	7.9	No	[M+2H] ²⁺	1198.9406	9.5	No	[M+2H] ²⁺	1198.9507	1.1	Yes
				[M+3H] ³⁺	799.6341	4.1	No	[M+3H] ³⁺	799.6326	5.9	No	[M+3H] ³⁺	799.6377	0.4	Yes
18	H5N6F2		N	[M+2H] ²⁺	1227.4529	8.0	No	[M+2H] ²⁺	1227.4508	9.8	No	[M+2H] ²⁺	1227.4621	0.5	No
				[M+3H] ³⁺	818.6410	4.3	Yes	[M+3H] ³⁺	818.6397	5.9	Yes	[M+3H] ³⁺	818.6449	0.5	No
19	H6N6F1		N	[M+2H] ²⁺	1235.4524	6.3	No	[M+2H] ²⁺	1235.4492	8.9	No	[M+2H] ²⁺	1235.4605	0.2	No
				[M+3H] ³⁺	823.9734	3.3	No	[M+3H] ³⁺	823.9716	5.5	No	[M+3H] ³⁺	823.9770	1.1	Yes
20	H5N6F3		N	[M+2H] ²⁺	1300.4811	8.2	No	[M+2H] ²⁺	1300.4786	10.1	No	[M+2H] ²⁺	1300.4917	0.02	No
				[M+3H] ³⁺	867.3253	6.0	Yes	[M+3H] ³⁺	867.3238	7.7	Yes				
21	H6N6F2		N	[M+2H] ²⁺	1308.4794	7.5	No	[M+2H] ²⁺	1308.4765	9.7	No	[M+2H] ²⁺	1308.4890	0.1	No
				[M+3H] ³⁺	872.6576	5.2	Yes	[M+3H] ³⁺	872.6555	7.6	Yes	[M+3H] ³⁺	872.6620	0.1	Yes
22	H6N6F3		N	[M+2H] ²⁺	1381.5075	7.7	No	[M+2H] ²⁺	1381.5038	10.4	No	[M+2H] ²⁺	1381.5171	0.8	No
				[M+3H] ³⁺	921.3423	6.3	Yes	[M+3H] ³⁺	921.3400	8.8	Yes	[M+3H] ³⁺	921.3470	1.2	Yes

23	H6N5F5		N	[M+2H] ²⁺	1426.0252	7.8	No	[M+2H] ²⁺	1426.0223	9.9	No	[M+2H] ²⁺	1426.0344	1.4	No
				[M+3H] ³⁺	951.0237	3.4	No	[M+3H] ³⁺	951.0192	8.1	No	[M+3H] ³⁺	951.0254	1.6	No
24	H6N6F4		N	[M+2H] ²⁺	1454.5350	8.3	No	[M+2H] ²⁺	1454.5319	10.4	No	[M+2H] ²⁺	1454.5453	1.2	No
				[M+3H] ³⁺	970.0275	6.7	Yes	[M+3H] ³⁺	970.0249	9.4	Yes	[M+3H] ³⁺	970.0327	1.4	No
25	H6N7F3		N	[M+2H] ²⁺	1483.0454	8.4	No	[M+2H] ²⁺	1483.0427	10.2	No	[M+2H] ²⁺	1483.0578	0.02	No
				[M+3H] ³⁺	989.0341	7.2	No	[M+3H] ³⁺	989.0319	9.4	Yes	[M+3H] ³⁺	989.0397	1.5	No
26	H6N5F6		N									[M+3H] ³⁺	999.7102	2.7	No
27	H6N6F5		N	[M+2H] ²⁺	1527.5645	7.6	No	[M+2H] ²⁺	1527.5590	11.2	No	[M+2H] ²⁺	1527.5745	1.0	Yes
				[M+3H] ³⁺	1018.7125	7.4	Yes	[M+3H] ³⁺	1018.7102	9.6	Yes	[M+3H] ³⁺	1018.7185	1.5	Yes
28	H7N6F4		N	[M+2H] ²⁺	1535.5626	7.1	No	[M+2H] ²⁺	1535.5581	10.0	No	[M+2H] ²⁺	1535.5737	0.1	No
				[M+3H] ³⁺	1024.0450	6.5	Yes	[M+3H] ³⁺	1024.0419	9.5	Yes	[M+3H] ³⁺	1024.0510	0.6	No
29	H6N7F4		N	[M+2H] ²⁺	1556.0734	8.6	No	[M+2H] ²⁺	1556.0707	10.3	No	[M+2H] ²⁺	1556.0853	1.0	No
				[M+3H] ³⁺	1037.7188	8.1	Yes	[M+3H] ³⁺	1037.7167	10.1	Yes	[M+3H] ³⁺	1037.7248	2.3	No
30	H6N6F6		N	[M+2H] ²⁺	1600.5931	7.4	No	[M+2H] ²⁺	1600.5891	9.9	No	[M+2H] ²⁺	1600.6029	1.3	No
				[M+3H] ³⁺	1067.3975	7.9	No	[M+3H] ³⁺	1067.3947	10.6	No	[M+3H] ³⁺	1067.4039	2.0	Yes
31	H6N7F5		N	[M+2H] ²⁺	1629.1021	8.4	No	[M+2H] ²⁺	1629.0996	9.9	No	[M+2H] ²⁺	1629.1135	1.4	No
				[M+3H] ³⁺	1086.4039	8.5	No	[M+3H] ³⁺	1086.4017	10.5	No	[M+3H] ³⁺	1086.4108	2.2	No
32	H7N7F4		N	[M+2H] ²⁺	1637.0975	9.6	No	[M+2H] ²⁺	1637.0971	9.8	No	[M+2H] ²⁺	1637.1117	0.9	No
				[M+3H] ³⁺	1091.7358	8.2	No	[M+3H] ³⁺	1091.7335	10.3	Yes	[M+3H] ³⁺	1091.7422	2.4	No
33	H5N4F2S1		N	[M+2H] ²⁺	1169.9216	8.1	No					[M+2H] ²⁺	1169.9334	2.0	No
34	H6N5F3S1		N	[M+3H] ³⁺	950.6784	8.8	No					[M+3H] ³⁺	950.6878	1.1	Yes
35	H6N5F4S1		N	[M+3H] ³⁺	999.3595	13.2	No					[M+3H] ³⁺	999.3741	1.4	Yes
36	H6N6F3S1		N	[M+2H] ²⁺	1527.0541	7.7	No					[M+2H] ²⁺	1527.0695	2.4	No
				[M+3H] ³⁺	1018.3711	8.6	No					[M+3H] ³⁺	1018.3813	1.4	Yes
37	H5N4S2		N									[M+3H] ³⁺	779.9481	2.3	Yes

38	H6N5F1S2		N									[M+3H] ³⁺	950.3447	2.0	Yes
39	H6N5F2S2		N									[M+3H] ³⁺	999.0294	3.2	Yes
40	H6N5F3S2		N									[M+3H] ³⁺	1047.7153	3.1	Yes
41	H6N6F2S2		N									[M+3H] ³⁺	1066.7223	3.2	Yes
42	H6N6F3S2		N									[M+3H] ³⁺	1115.4087	2.7	Yes
43	H7N6F2S2		N									[M+3H] ³⁺	1120.7398	3.1	Yes
44	H7N6F3S2		N									[M+3H] ³⁺	1169.4254	3.3	Yes
45	H7N6F4S2		N									[M+3H] ³⁺	1218.1141	0.9	Yes

^aNum is referred to the *N*-glycopeptide number.

^bGlycan compositions are given in terms of number of hexoses (H), N-acetylglucosamines and N-acetylgalactosamines (N), fucoses (F) and sialic acids (S).

^cRelative error was calculated in ppm as: $|m/z_{\text{exp}} - m/z_{\text{theo}}|/m/z_{\text{theo}} \times 10^6$ (exp=experimental and theo=theoretical). Observed *m/z* with error>10.0 ppm are marked in red.

^d*N*-glycopeptide structures confirmed by MS/MS (Yes) (At least 3 observed fragments).

Chapter 4. Untargeted analysis for metabolomics studies

The separation, detection, identification and quantification of low molecular mass compounds in biological fluids (e.g. blood plasma) continues generating a great interest in many fields, especially in metabolomics studies for biomarker discovery. However, due to the complexity of the blood plasma matrix, especially with regard to the high concentration of salts and proteins, untreated plasma is inconvenient for direct analysis of low molecular mass compounds using MS-hyphenated techniques, such as liquid chromatography mass spectrometry (LC-MS) and capillary electrophoresis mass spectrometry (CE-MS). There are different strategies to remove salts and proteins before the analysis of the compounds in the low molecular mass fraction, being solvent precipitation and centrifugal filtration one of the best alternatives due to their good efficiency, simplicity, high throughput, and low cost.

Despite targeted strategies are widely applied for the analysis of low abundant metabolites, which are often bioactive and play important roles in biological systems, untargeted strategies can implicate previously unrecognized metabolites and, therefore, are powerful platforms to elucidate novel biomarkers and gain insight into disease pathogenesis. The present chapter is focused on the **untargeted** identification of biomarkers of Huntington's disease (HD), an inherited amyloidotic neurodegenerative disorder characterized by progressive motor and cognitive disturbances. HD is caused by an expansion of the cytosine-adenine-guanine (CAG) repeat in the exon 1 of the huntingtin gene (HTT), which encodes a stretch of glutamines in the huntingtin protein. It seems that the onset of the disease is related to the length of the CAG repeat, which defines the extent on huntingtin mutation. However, without robust and practical measures of other biochemical biomarkers of disease progression, the efficacy of therapeutic interventions in this premanifest HD cannot be readily assessed.

This chapter describes the optimization of a plasma pretreatment for optimized recovery of low molecular mass compounds prior to on-line solid-phase extraction capillary electrophoresis mass spectrometry with a C₁₈ chromatographic sorbent (C₁₈-SPE-CE-MS). C₁₈-SPE-CE-MS is presented as a novel alternative sensitive method for on-line sample clean-up and concentration

enhancement in metabolomics studies with human plasma samples. Moreover, this chapter presents the capacity of the above mentioned C₁₈-SPE-CE-MS methodology to analyze low molecular mass compounds in plasma samples from wild-type (wt) and HD mice in combination with advanced chemometric data analysis tools, such as multivariate curve resolution alternating least squares (MCR-ALS), principal component analysis (PCA) and partial least squares discriminant analysis (PLS-DA), for the identification of potential metabolite biomarker candidates involved in HD progression. These metabolite biomarkers could provide a window of opportunity for prediction of disease onset, evaluation of HD early progression and response to treatment.

This chapter includes the following publications:

- **Publication 4.1.** An update for human blood plasma pretreatment for optimized recovery of low-molecular-mass peptides prior to CE-MS and SPE-CE-MS. L. Pont, F. Benavente, J. Barbosa, V. Sanz-Nebot. *Journal of Separation Science* 36 (2013), 3896-3902.
- **Publication 4.2.** Metabolic profiling for the identification of Huntington biomarkers by on-line solid-phase extraction capillary electrophoresis mass spectrometry combined with advanced data analysis tools. L. Pont, F. Benavente, J. Jaumot, R. Tauler, J. Alberch, S. Ginés, J. Barbosa, V. Sanz-Nebot. *Electrophoresis* 37 (2016), 795-808.

Laura Pont
Fernando Benavente
José Barbosa
Victoria Sanz-Nebot

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

Received July 29, 2013
Revised September 20, 2013
Accepted October 9, 2013

Research Article

An update for human blood plasma pretreatment for optimized recovery of low-molecular-mass peptides prior to CE-MS and SPE-CE-MS

Protein precipitation and centrifugal filtration are well-established methods for concentrating and purifying peptides with a low relative molecular mass (M_r) from human blood plasma before proteomic and peptidomic studies using high-performance separation techniques, but there is little information on peptide recoveries. Here, we evaluate acetonitrile precipitation followed by a range of centrifugal filtration conditions for the analysis of low M_r peptides in human blood plasma before CE-MS and SPE coupled online to CE-MS. Three opioid peptides were used as model compounds, that is, dynorphin A 1–7, endomorphin 1, and methionine enkephalin and 3, 10, and 30 K M_r cut-off cellulose acetate filters (Amicon® Ultra-0.5) and 10 K M_r cut-off polyethersulfone filters (Vivaspin® 500) were studied. Unexpectedly, recoveries and repeatability were only optimum after passivating the 10 K M_r cut-off cellulose acetate filters with PEG to avoid peptide adsorption on the inner walls of the plastic sample reservoir.

Keywords: CZE / Biological fluids / Clean-up / Preconcentration / Pretreatment
DOI 10.1002/jssc.201300838

1 Introduction

The detection and identification of low-relative-molecular-mass (M_r) peptides in human blood plasma continues to generate a great interest in many fields of proteomic and peptidomic research, such as biomarker discovery [1–5]. However, due to the complexity of human blood plasma, especially with regard to the wide dynamic range of protein and peptide concentrations, untreated plasma is inconvenient for direct analysis of low M_r peptides using LC-MS, CE-MS, or many other separation techniques [1–16]. There are different strategies to remove salts and major high M_r proteins from plasma before the analysis of the low M_r peptide fraction [1–24]. Solvent precipitation and centrifugal filtration are widely used because of the good efficiency, simplicity, high throughput, and low cost. Furthermore, they are often combined together because the solubility and size-exclusion mechanisms that govern both methods can be regarded as complementary [21–24]. However, there is little information on low M_r peptide recoveries. Thus, for example, centrifugal filtration is often applied to desalt and concentrate high

M_r protein fractions, hence recoveries are often evaluated in terms of the retentate and no attention is given to the low M_r peptides in the filtrates [25–31]. In a previous work, we proposed a double-step pretreatment for the analysis of low M_r peptides (namely opioid peptides) in human blood plasma by CE-MS and SPE-CE-MS that consisted of precipitation with acetonitrile followed by centrifugal filtration [22–24]. Peptide recovery for the precipitation step was good when the volume ratio between acetonitrile and plasma was 6:1 v/v (49, 91, and 102% for dynorphin A (1–7) (Dyn A, Tyr-Gly-Gly-Phe-Leu-Arg-Arg, M_r 867.4715), endomorphin 1 (End 1, Tyr-Pro-Trp-Phe-NH₂, M_r 610.2904), and methionine enkephalin (Met, Tyr-Gly-Gly-Phe-Met, M_r 573.2257), respectively) [23, 24]. The first step of protein precipitation was enough before CE-MS, but not before C₁₈-SPE-CE-MS that required the subsequent centrifugal filtration step for an improved sample matrix clean-up. SPE-CE-MS is nowadays one of the best alternatives to improve between 100- and 1000-fold the LODs of CE-MS [8, 9, 22–24, 32–38]. In SPE-CE-MS a microcartridge placed near the inlet of the separation capillary contains a sorbent (e.g. C₁₈) that retains the target analyte from a large volume of sample. After rinsing, the retained analyte is eluted in a smaller volume of an appropriate solution, resulting in sample clean-up and concentration enhancement. However, the use of sorbents with limited selectivity precludes the direct analysis of plasma samples. In such cases, the double-step pretreatment proposed in this paper is required to isolate and enrich the target analytes in order to prevent microcartridge saturation [22–24]. In our previous work [24], Microcon Amicon® filters of 10 K M_r cut-off (YM-10, Millipore,

Correspondence: Dr. Victoria Sanz-Nebot, Department of Analytical Chemistry, University of Barcelona, Avda. Diagonal, 645, Barcelona 08028, Spain
E-mail: vsanz@ub.edu
Fax: +34-934021233

Abbreviations: Dyn A, dynorphin A 1–7; End 1, endomorphin 1; L_T , total length; Met, methionine enkephalin; M_r , relative molecular mass; PBS, phosphate buffered saline

Bedford, MA, USA) provided the best peptide recoveries using the complete double-step pretreatment (26, 71, and 72% for Dyn A, End 1, and Met, respectively). However, these centrifugal filters with the membrane in horizontal position (Microcon Amicon® filters) for direct flow filtration have been discontinued by the manufacturer (Millipore, Bedford, MA, USA) and substituted by others with the membrane in vertical position for tangential (or cross) flow filtration (Amicon® Ultra-0.5 filters, Millipore, USA) [28]. The new filters with the vertical design, as well as with other less evident changes, are claimed to prevent membrane fouling due to solute polarization and spinning to dryness, and are supposed to produce similar results. In the present study we revised the double-step sample pretreatment method for fractionation of low M_r peptides from human blood plasma that urgently needed an update to be reliably applied using the currently available centrifugal filters.

2 Materials and methods

2.1 Chemicals and reagents

All the chemicals used in the preparation of buffers and solutions were of analytical reagent grade. Acetonitrile, propan-2-ol, formic acid (HFor) (98–100%), acetic acid (HAc) (glacial), ammonia (25%), sodium hydrogenphosphate, sodium dihydrogenphosphate, sodium chloride, potassium dihydrogenphosphate, potassium chloride, sodium hydroxide, and trichloroacetic acid were purchased from Merck (Darmstadt, Germany). Met, End 1, and BSA were provided by Sigma (St. Louis, MO, USA). Dyn A was supplied by Bachem (Bubendorf, Switzerland). PEG (8 K) was purchased from Fluka (Buchs, Switzerland). Water with a conductivity value $<0.05 \mu\text{S}/\text{cm}$ was obtained using a Milli-Q water purification system (Millipore, Molsheim, France).

2.2 Electrolyte solutions, sheath liquid, and standard solutions

An aqueous standard solution (2500 mg/L) of each peptide was prepared and stored in a freezer at -20°C when not in use. Working standard solutions were obtained by diluting the stock solutions with water. Diluted standard solutions were used to spike human blood plasma samples. The BGE contained 50 mM of HAc and 50 mM of HFor and was adjusted to pH 3.5 with ammonia. The sheath liquid solution consisted of an hydroorganic mixture of 60:40 v/v propan-2-ol/water with a 0.05% v/v of formic acid. All solutions were passed through a $0.45 \mu\text{m}$ nylon filter (MSI, Westboro, MA, USA) before analysis and were stored at 4°C when not in use. The sheath liquid was degassed for 10 min by sonication before use. Solutions of 1% m/v of BSA in phosphate buffered saline (PBS) (0.01 M sodium hydrogenphosphate, 0.0015 M potassium dihydrogenphosphate, 0.14 M sodium

chloride, 0.0027 M potassium chloride, pH 7.2) and 5% v/v of PEG in water were used to passivate the centrifugal filters.

2.3 Human blood plasma and sample preparation

Human blood plasma samples from healthy blood donors were kindly supplied by the Banc de Sang i Teixits (Hospital de la Vall d'Hebron, Barcelona). The off-line double-step pretreatment of blank and fortified plasma samples consisted of protein precipitation followed by centrifugal filtration [22–24]. For precipitation with acetonitrile, 1200 μL of cold solvent was added to 200 μL of plasma sample. The mixture was then vortexed for 30 s and kept at -20°C for 1 h. The supernatant was collected after centrifugation of the mixture at $700 \times g$ for 8 min. The solid residue was rinsed with 100 μL of acetonitrile. After centrifugation of the mixture at $1200 \times g$ for 8 min, the supernatant was pooled with the first one. Finally, the micropipette tip used to deliver the supernatants was rinsed with 100 μL of acetonitrile and the washing volume was also collected. The pooled solution was evaporated to dryness at room temperature under a stream of air and the solid residue was reconstituted with 100 μL of water. The reconstituted sample was subjected to centrifugal filtration. A range of centrifugal filters with different membrane types and pore sizes were investigated for ultrafiltration: Amicon® Ultra-0.5 centrifugal devices of 3, 10, and 30 K M_r cut-off cellulose acetate filters (3, 10, and 30 K, respectively) (Millipore) and Vivaspin® 500 of 10 K M_r cut-off polyethersulfone membrane (10 K-PES) (GE Healthcare, Little Chalfont, UK). All the centrifugation steps were performed at a centrifugal force of $11\,000 \times g$. First, the filter was washed with water for 15 min and the filtrate was discarded. Then, the reconstituted sample was centrifuged for 10 min. The filtration residue was washed two times for 10 and 8 min with 50 μL of water. In all cases, the filtrates were pooled together and sufficient water was added until the final volume was 200 μL .

In some experiments, the 3 and 10 K devices were passivated with 1% m/v of BSA in PBS and 5% v/v of PEG in water. A volume of 500 μL of the passivation solution was pipetted into the reservoir of the centrifugal device, capped, and allowed to stand 24 h at room temperature. The reservoirs were uncapped and rinsed thoroughly with tap water. A volume of 500 μL of distilled water was added to remove any residue and the centrifuge was then spun to dead stop (30 min at $11\,000 \times g$). The remaining water was eliminated by inverting the reservoir in a vial and spinning once more at a reduced centrifugal force (2 min at $300 \times g$). In order to test the influence of temperature, the reconstituted sample resulting after precipitation with acetonitrile was incubated at 85°C for 15 min before filtration with PEG-passivated 10 K filters.

The recovery for each peptide was estimated from a comparison of the peak areas obtained by CE-MS for a plasma sample spiked at 10 $\mu\text{g}/\text{mL}$ with each peptide before and after the double-step pretreatment. For the calculations, 100%

recovery was considered for the plasma samples spiked after the sample pretreatment.

2.4 Apparatus and procedures

pH measurements were performed with a Crison 2002 potentiometer and a Crison electrode 52–03 (Crison Instruments, Barcelona, Spain). Centrifugal filtration procedures were carried out at room temperature in a 5417R centrifuge (Eppendorf Ibérica, Madrid, Spain). Sample incubation was performed with a Thermo-Shaker TS-100 (Biosan, Warren, USA).

2.4.1 CE–MS

CE–MS experiments were performed in an HP^{3D} CE system coupled with an orthogonal G1603A sheath-flow interface to an ion trap mass spectrometer (LC/MSD trap SL, Agilent Technologies, Waldbronn, Germany). The sheath liquid was delivered at a flow rate of 3.3 $\mu\text{L}/\text{min}$ by a KD Scientific 100 series infusion pump (Holliston, MA, USA). The IT mass spectrometer was operated in positive mode in the m/z range from 250 to 1250 with the parameters established in a previous work [22–24]. Instrument control, data acquisition and data processing were performed using CE/MSD Trap Software (Agilent Technologies).

A 72 cm total length (L_T) \times 75 μm id \times 360 μm od bare fused-silica capillary supplied by Polymicro Technologies (Phoenix, AZ, USA) was used for the separations. All capillary rinses were performed at high pressure (930 mbar). New capillaries were flushed for 20 min with 1 M NaOH, followed by 15 min with water and 30 min with BGE. The system was finally equilibrated by applying the separation voltage (17 kV) for 15 min. Between workdays, the capillary was conditioned by rising successively with 0.1 M NaOH (5 min), water (10 min), and BGE (15 min). Both activation and conditioning procedures were performed off-line in order to avoid the unnecessary entrance of NaOH into the MS system. Samples were hydrodynamically injected at 50 mbar for 5 s. A separation voltage of 17 kV (normal polarity, anode at the injection capillary end) was applied for the electrophoretic separations at 25°C. Between runs, the capillary was rinsed for 3 min with BGE.

2.4.2 SPE–CE–MS

The construction of the microcartridge or analyte concentrator for SPE–CE–MS was carried out as described elsewhere [22–24]. A 72 cm L_T \times 75 μm id \times 360 μm od fused-silica was prepared and activated as described above. The microcartridge (7 mm L_T \times 250 μm id \times 360 μm od, fused-silica capillary, Polymicro Technologies) was inserted at 7.5 cm from the inlet of the separation capillary using two plastic sleeves. Previously, it was filled with the sorbent found in C₁₈ Sep-pak cartridges (Waters, Milford, MA, USA) that was retained between two frits (0.1 cm).

The separation columns were first conditioned by consecutive flushes of water (1 min), methanol (1 min), water (1 min), and BGE (3 min) at 930 mbar. Spiked or blank plasma samples were then introduced at 930 mbar for 10 min (approximately 60 μL using the Hagen–Poiseuille equation [39]). A final rinse with the BGE (2 min at 930 mbar) eliminated nonretained molecules and equilibrated the capillary before the elution. Retained peptides were eluted by injecting a 60:40 v/v methanol/water solution with 50 mM HAc and 50 mM HFor at 50 mbar for 10 s. Separation was carried out at 25°C by applying a voltage of 17 kV (normal polarity). Between runs, the capillary was rinsed for 2 min with water and 2 min with acetonitrile, in order to avoid carry-over between consecutive analyses.

3 Results and discussion

3.1 CE–MS

It is widely accepted that centrifugal filtration alone does not adequately remove high abundance proteins, particularly when protein concentration is very high as is the case for whole plasma [17, 21–25]. Thus, we observed that with a single-step centrifugal filtration only Met could be detected because recoveries were extremely poor (around 10% for Met). Precipitation with acetonitrile was necessary to eliminate most of the salts and high-abundance proteins before centrifugal filtration [21–24]. Recoveries obtained for the studied peptides with a single step of precipitation with acetonitrile (49, 91, and 102% for Dyn A, End 1, and Met, respectively) [24], could not be further improved adding an extra step of precipitation with acetonitrile or trichloroacetic acid [19], especially for Dyn A. As this type of deproteinization methods are strictly based on solubility, it is difficult to finely tune it for mixtures of compounds (e.g. Dyn A was the largest and most basic peptide) or to avoid coprecipitation of the target peptides due to nonspecific binding to the plasma proteins [40]. This first step of protein precipitation was enough before CE–MS, but not before C₁₈–SPE–CE–MS that required the subsequent centrifugal filtration step for an improved sample matrix clean-up [22–24] in order to prevent microcartridge saturation.

Figure 1 shows in a bar plot graph the mean recoveries and their RSDs for Dyn A, End 1, and Met after applying the double-step sample pretreatment method to plasma samples spiked with 10 $\mu\text{g}/\text{mL}$ of each peptide using the new 3, 10, 30, and 10 K-PES filters. As can be observed, recovery values were significantly lower than those that we previously reported using YM-10 filters (Fig. 1) under optimum conditions (recovery values were 26, 71, and 72% for Dyn A, End 1, and Met, respectively) [24]. Even under certain filtration conditions, some of the peptides were not detected. For example, using 10 K and 10 K-PES filters only Met was detected and recovery values were 43 and 41%, respectively. With 3 K filters, End 1 and Met were detected (recovery values were 26 and 67%, respectively). The best recovery values were obtained

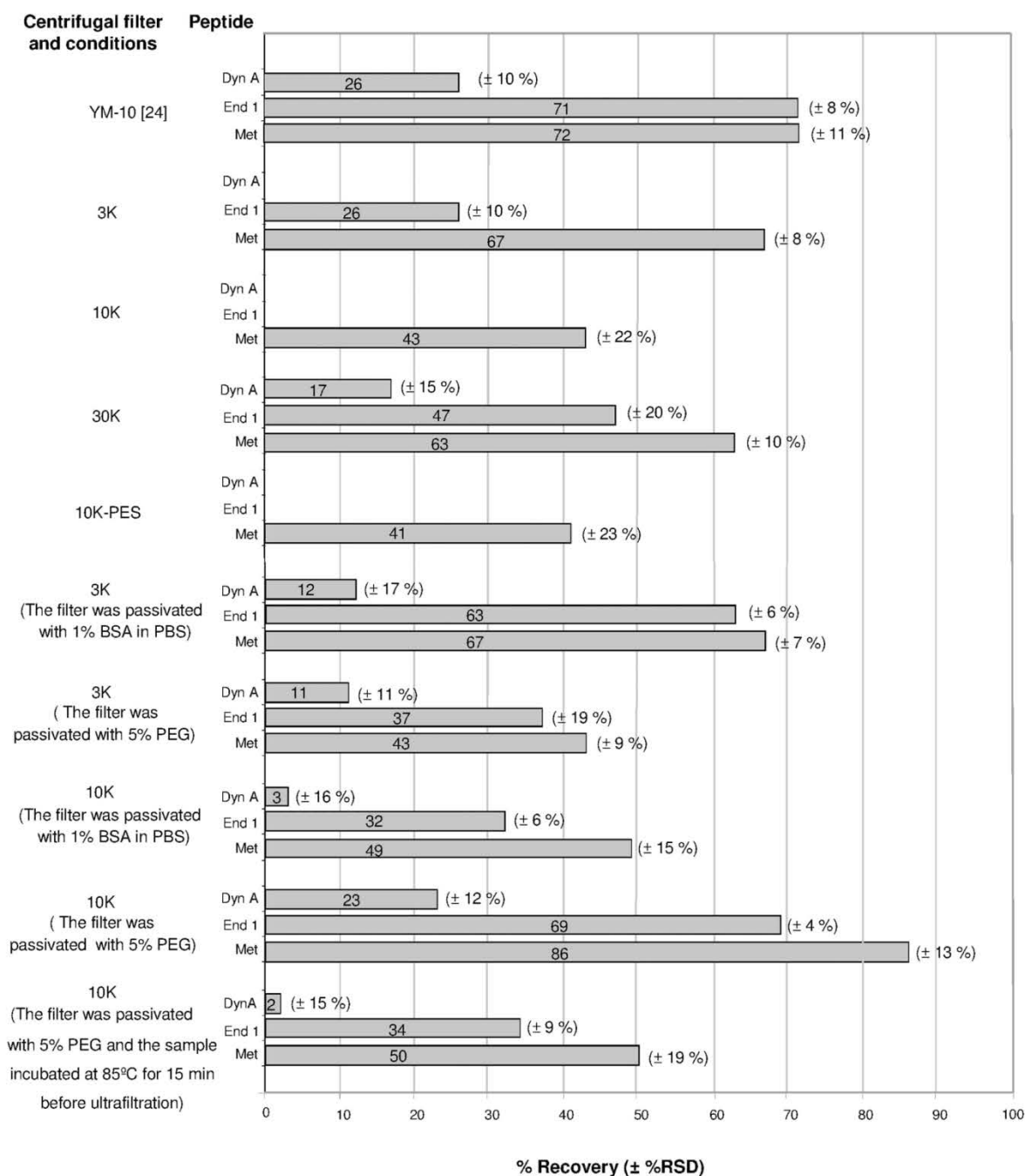


Figure 1. Evaluation of peptide recovery by CE–MS using YM-10 [24], 3, 10, 30, and 10 K-PES centrifugal filters under several conditions. Human blood plasma samples were spiked with 10 µg/mL of each peptide. Temperature was 25°C unless otherwise indicated. Recovery was estimated from a comparison of the peak areas obtained for each peptide for spiked plasma samples before and after the double-step pretreatment.

with 30 K filters (17, 47, and 63% for Dyn A, End 1, and Met, respectively), but this filter was not selected for further work because, based on our previous experience, the filtrates contained increased amounts of interferences from the plasma matrix due to the larger pore size [23, 24]. Furthermore, the repeatability of the recovery values was also low (average rela-

tive standard deviation for the recoveries of the three peptides was 15%).

Some manufacturers of centrifugal filters recommend passivating the microdevices when handling highly dilute analyte solutions to prevent analyte adsorption on the plastic walls [25, 28–31]. In our previous work, the sample

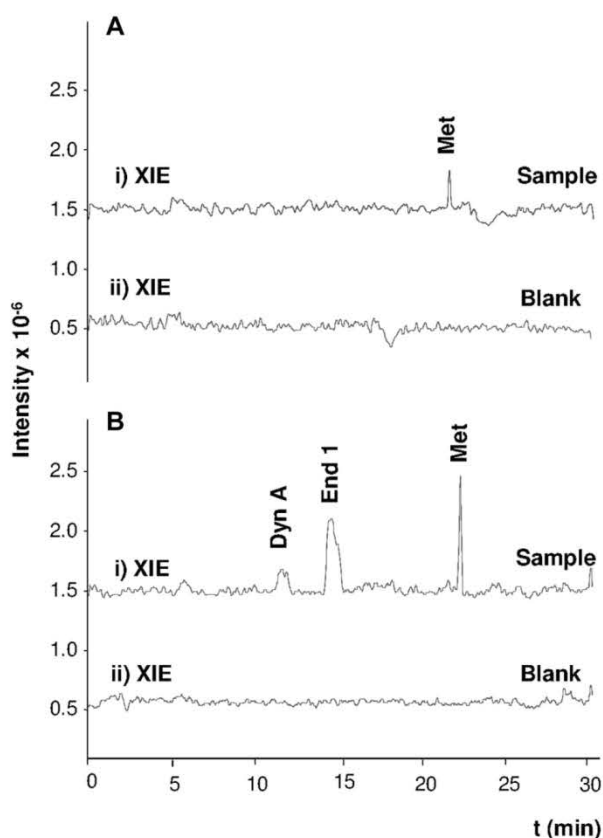


Figure 2. Extracted ion electropherograms (XIEs) by CE-MS for plasma samples using (A) nonpassivated 10 K filters and (B) PEG-passivated 10 K filters. (i) Plasma spiked with 10 µg/mL of each peptide before the double-step pretreatment and (ii) blank plasma.

reservoir of the YM series centrifugal filters was made of polycarbonate. Passivation with BSA or PEG was not necessary because it did not improve peptide recoveries [24]. Now, the sample reservoir of the K series centrifugal filters was made of a copolymer of styrene-butadiene [28] (the filtrate vial was made of polypropylene as in YM series filters). Peptide recoveries were investigated after passivating 3 and 10 K filters with 1% m/v of BSA in PBS and 5% v/v of PEG (8 K) in water. A volume of 500 µL of the passivation solution was allowed to stand in the reservoir for 24 h at 25°C. Then, the reservoirs were rinsed thoroughly with tap water, distilled water and LC-grade water [24]. As shown in Fig. 1, in all cases recoveries were clearly improved with passivation. However, any of both solutions could be proposed as a gold standard to passivate the centrifugal filters, because results for 3 K filters were significantly different for BSA and PEG, but the trend was exactly the opposite for the 10 K filters. The best recovery values for the 3 K filters were obtained with BSA, whereas for the 10 K filters were achieved with PEG (Fig. 1). In both cases, repeatability was similar (average RSDs were 10 and 9%, respectively). As the results for the PEG-passivated 10 K filters were similar to those obtained in our previous

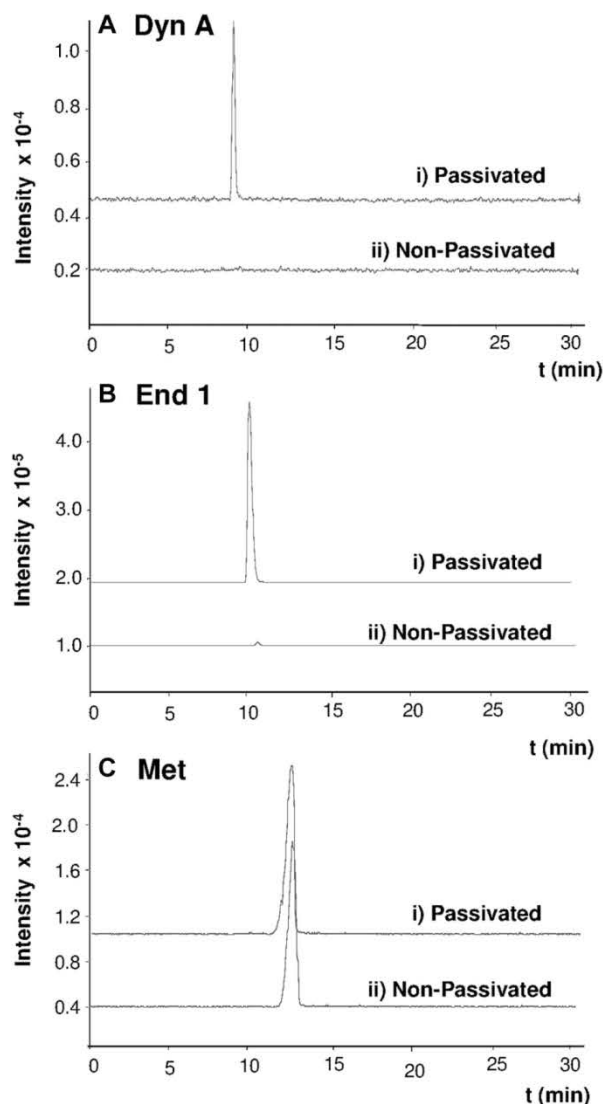


Figure 3. Extracted ion electropherograms (XIEs) by SPE-CE-MS for plasma samples spiked with 100 ng/mL of (A) Dyn A, (B) End 1, and (C) Met before the double-step pretreatment and using (i) PEG-passivated 10 K filters and (ii) nonpassivated 10 K filters.

work with YM-10 filters [24], we made further experiments to improve recoveries. Under the optimized conditions, we studied the effect of temperature before centrifugal filtration, because an increase of temperature could serve as a method for overcoming the hydrophobic, ionic and other interactions that hold peptides on the surface of abundant plasma proteins [40]. However, recoveries were significantly lower incubating the reconstituted sample after precipitation with acetonitrile at 85°C for 15 min before ultrafiltration (Fig. 1, 2, 34, and 50% for Dyn A, End 1, and Met). Figure 2A and B shows the extracted ion electropherograms by CE-MS for blank plasma samples and plasma samples spiked with 10 µg/mL of each peptide after precipitation with acetonitrile followed by centrifugal filtration with nonpassivated 10 K filters and

PEG-passivated 10 K filters at 25°C. As can be observed, while only Met was detected with nonpassivated 10 K filters due to the poor recovery values (Fig. 1), all peptides were detected with PEG-passivated 10 K filters.

3.2 SPE coupled online to CE-MS

Plasma samples spiked with 100 ng/mL of each peptide were subjected to the double-step pretreatment with nonpassivated and PEG-passivated 10 K filters before C₁₈-SPE-CE-MS. As can be observed in Fig. 3, according to the recoveries of Fig. 1, Dyn A was not detected and peak height of End 1 and Met were significantly lower with nonpassivated 10 K filters. This was especially significant for End 1 that as Dyn A was strongly affected by PEG-passivation. Using the PEG-passivated 10 K filters, the LODs for the studied peptides by C₁₈-SPE-CE-MS in plasma samples could be again decreased down to 10 ng/mL for Dyn A and 1 ng/mL for End 1 and Met as in our previous work with YM-10 filters [23, 24].

4 Concluding remarks

We can conclude that Amicon® Ultra-0.5 centrifugal devices need to be passivated with BSA or PEG to obtain optimum recoveries and repeatability in the analysis of opioid peptides in human blood plasma. The paper highlights how changes in the manufacturing process of a device routinely used for sample pretreatment may be critical to achieve an accurate and reliable characterization of the low M_r peptidome of biological fluids using high-performance separation techniques. Furthermore, results point to the continuous revision and update of the sample pretreatment procedures typically used in proteome and peptidome research, which may be extremely sensitive to subtle and unexpected changes in reagents and materials.

Laura Pont acknowledges the Spanish Ministry of Economy and Competitiveness for a FPI fellowship. This study was supported by a grant from the Spanish Ministry of Education and Science (CTQ2011–27130).

The authors have declared no conflict of interest.

5 References

- [1] Luque-Garcia, J. L., Neubert, T. A., *J. Chromatogr. A* 2007, 1153, 259–276.
- [2] Vogelsang, M. S., Kani, K., Katz, J. E., *Methods Mol. Biol.* 2011, 728, 87–107.
- [3] Kawashima, Y., Fukutomi, T., Tomonaga, T., Takahashi, H., Nomura, F., Maeda, T., Kodera, Y., *J. Proteome Res.* 2010, 9, 1694–1705.
- [4] Issaq, H. J., Fox, S. D., Chan, K. C., Veenstra, T. D., *J. Sep. Sci.* 2011, 34, 3484–3492.
- [5] Chlu, C., Chang, C., Chen, S., *J. Sep. Sci.* 2012, 35, 3293–3301.
- [6] Zheng, X., Baker, H., Hancock, W. S., *J. Chromatogr. A* 2006, 1120, 173–184.
- [7] Jiang, X., Ye, M., Zou, H., *Proteomics* 2008, 8, 686–705.
- [8] Visser, N. F. C., Lingeman, H., Irth, H., *Anal. Bioanal. Chem.* 2005, 382, 535–558.
- [9] Hernández-Borges, J., Borges-Miquel, T. M., Rodríguez-Delgado, M. A., Cifuentes, A., *J. Chromatogr. A* 2007, 1153, 214–226.
- [10] Macdonald, W. H., *Dis. Markers* 2002, 18, 99–105.
- [11] Sandra, K., Moshir, M., D'hondt, F., Verleysen, K., Kas, K., Sandra, P., *J. Chromatogr. B* 2008, 866, 48–63.
- [12] Sandra, K., Moshir, M., D'hondt, F., Tuytten, R., Verleysen, K., Kas, K., Francois, I., Sandra, P., *J. Chromatogr. B* 2009, 877, 1019–1039.
- [13] Fonslow, B. R., Yates, J. R., *J. Sep. Sci.* 2009, 32, 1175–1188.
- [14] Herrero, M., Ibanez, E., Cifuentes, A., *Electrophoresis* 2008, 29, 2148–2160.
- [15] Simo, C., Cifuentes, A., Kasicka, V., *Methods Mol. Biol.* 2013, 984, 139–151.
- [16] Kasicka, V., *Electrophoresis* 2012, 33, 48–73.
- [17] Bioss, G., Addis, M. F., Tanca, A., Pisanu, S., Roggio, T., Uzzau, S., Pagnozzi, D., *J. Proteomics* 2011, 75, 93–99.
- [18] Borujeni, E., Zydney, A. L., *Biotechniques* 2012, 53, 49–56.
- [19] Ralston P. B., Strein, T. G., *Microchem. J.* 1997, 55, 270–283.
- [20] Daykin, C. A., Foxall, J. D., Connor, S. C., Lindon, J. C., Nicholson, J. K., *Anal. Biochem.* 2002, 304, 220–230.
- [21] Ban, E., Choi, O. K., Chung, W. Y., Park, C. S., Yoo, E. A., Chung, B. C., Yoo, Y. S., *Electrophoresis* 2001, 22, 2173–2178.
- [22] Hernández, E., Benavente, F., Sanz-Nebot, V., Barbosa, J., *Electrophoresis* 2007, 28, 3957–3965.
- [23] Hernández, E., Benavente, F., Sanz-Nebot, V., Barbosa, J., *Electrophoresis* 2008, 29, 3366–3376.
- [24] Benavente, F., Medina-Casanellas, S., Barbosa, J., Sanz-Nebot, V., *J. Sep. Sci.* 2010, 33, 1294–1304.
- [25] Georgiou, H. M., Rice, G. E., Baker, M. S., *Proteomics* 2001, 1, 1503–1506.
- [26] Eppler, A., Markus Weigandt, M., Schulze, S., Hanefeld, A., Bunjes, H., *Int. J. Pharm.* 2011, 421, 120–129.
- [27] Jones, T. H., Brassard, J., Johns, M. W., Gagné, M. J., *J. Virol. Methods* 2009, 161, 199–204.
- [28] Amicon® Ultra-0.5 Centrifugal Filter Devices, PR03711, Rev. A, 11/1 User Guide, Millipore Corporation.
- [29] Passivation of amicon microcon microconcentrators for improved recovery, PC1003EN00 Protocol Note, 1999, Millipore Corporation.
- [30] Treatment of Vivaspin concentrators for improved recovery of low-concentrated protein samples, SLL4007-e06111 Application Note, Sartorius AG.
- [31] Fujimoto, T., Miya, M., Machida, M., Takechi, S., Kakinoki, S., Kanda, K., Nomura, A., *Glob. J. Health Sci.* 2006, 52, 718–723.

3902 L. Pont et al.

J. Sep. Sci. 2013, 36, 3896–3902

- [32] Guzman, N. A., Phillips, T. M., *Electrophoresis* 2011, 32, 1565–1578.
- [33] Guzman, N. A., Blanc, T., Philips, T. M., *Electrophoresis* 2008, 29, 3259–3278.
- [34] Puig, P., Borrull, F., Calull, M., Aguilar, C., *TrAC Trends Anal. Chem.* 2007, 26, 664–678.
- [35] Ramautar, R., Somsen, G. W., Jong, G. J., *Electrophoresis* 2010, 31, 44–54.
- [36] Tak, Y. H., Torano, J. S., Somsen, G. W., de Jong, G. J., *J. Chromatogr. A* 2012, 1267, 138–143.
- [37] Ramautar, R., de Jong, G. J., Somsen, G. W., *Electrophoresis* 2012, 33, 243–250.
- [38] Macià, A., Borrull, F., Calull, M., Benavente, F., Hernández, F., Sanz-Nebot, V., Barbosa, J., Aguilar, C., *J. Sep. Sci.* 2008, 31, 872–880.
- [39] Heiger, D., *High Performance Capillary Electrophoresis. An Introduction*, Agilent Technologies, Germany 2000, pp. 83–86.
- [40] Ziganshin, R., Arapidi, G., Azarkin, I., Zaryadieva, E., Alexeev, D., Govorun, V., Ivanov, V., *J. Proteomics* 2011, 74, 595–606.

Laura Pont¹
 Fernando Benavente¹
 Joaquim Jaumot²
 Romà Tauler²
 Jordi Alberch^{3,4,5}
 Silvia Ginés^{3,4,5}
 José Barbosa¹
 Victoria Sanz-Nebot¹

¹Departament de Química Analítica, Facultat de Química, Universitat de Barcelona, Barcelona, Spain

²Department of Environmental Chemistry, IDAEA-CSIC, Barcelona, Spain

³Departament de Biologia Cel·lular, Immunologia i Neurociències, Facultat de Medicina, Universitat de Barcelona, Barcelona, Spain

⁴Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, Spain

⁵Centro de Investigación Biomédica en Red sobre Enfermedades Neurodegenerativas (CIBERNED), Madrid, Spain

Received August 12, 2015

Revised November 10, 2015

Accepted December 7, 2015

Research Article

Metabolic profiling for the identification of Huntington biomarkers by on-line solid-phase extraction capillary electrophoresis mass spectrometry combined with advanced data analysis tools

In this work, an untargeted metabolomic approach based on sensitive analysis by on-line solid-phase extraction capillary electrophoresis mass spectrometry (SPE-CE-MS) in combination with multivariate data analysis is proposed as an efficient method for the identification of biomarkers of Huntington's disease (HD) progression in plasma. For this purpose, plasma samples from wild-type (wt) and HD (R6/1) mice of different ages (8, 12, and 30 weeks), were analyzed by C₁₈-SPE-CE-MS in order to obtain the characteristic electrophoretic profiles of low molecular mass compounds. Then, multivariate curve resolution alternating least squares (MCR-ALS) was applied to the multiple full scan MS datasets. This strategy permitted the resolution of a large number of metabolites being characterized by their electrophoretic peaks and their corresponding mass spectra. A total number of 29 compounds were relevant to discriminate between wt and HD plasma samples, as well as to follow-up the HD progression. The intracellular signaling was found to be the most affected metabolic pathway in HD mice after 12 weeks of birth, when mice already showed motor coordination deficiencies and cognitive decline. This fact agreed with the atrophy and dysfunction of specific neurons, loss of several types of receptors, and changed expression of neurotransmitters.

Keywords:

Biomarker / Huntington / Metabolomics / Multivariate data analysis / SPE-CE-MS
 DOI 10.1002/elps.201500378

1 Introduction

Huntington's disease (HD) is an inherited neurodegenerative disorder, which is characterized by progressive motor and cognitive disturbances. HD is caused by an expansion of the cytosine-adenine-guanine (CAG) repeat in the exon 1 of the Huntingtin gene (*HTT*), which encodes a stretch of glutamines in the Huntingtin protein [1–8]. Although the *HTT* gene is ubiquitously expressed as the Huntingtin protein in most tissues, HD pathology has primarily been located to the basal ganglia and to the neocortex. The pathology involves

atrophy and dysfunction of specific neurons, loss of several types of receptors, changed expression of neurotransmitters and key proteins, as well as formation of ubiquitin positive aggregates [1–8]. HD is a fatal disease, and the median interval between clinical diagnosis and death is typically given as 15–20 years [2, 4, 6, 8].

By use of predictive genetic testing, it is possible to identify individuals who carry the *HTT* gene defect before the onset of symptoms, providing a unique window of opportunity for intervention aimed at preventing or delaying disease onset [4, 7]. However, without robust and practical measures of disease progression, the efficacy of therapeutic interventions in this premanifest HD cannot be readily assessed. Neuroimaging and biochemical biomarkers are being investigated for their potential in clinical use and their value in the development of future treatments [4, 7]. Modern neuroimaging techniques such as magnetic resonance imaging (MRI) enable high-quality images of brain structure and function to be obtained [9, 10]. However, metabolites that can be quantified in biofluids, such as blood or urine, are appealing due to the improved selectivity, the minimal requirement for

Correspondence: Dr. Fernando Benavente, Departament de Química Analítica, Facultat de Química, Universitat de Barcelona, Gran Via de les Corts Catalanes, 585, 08007 Barcelona, Spain
E-mail: fbenavente@ub.edu
Fax: +34-934021233

Abbreviations: CAG, cytosine-adenine-guanine; HD, Huntington's disease; *HTT*, Huntingtin gene; LVs, latent variables; MCR-ALS, multivariate curve resolution alternating least squares; MRI, magnetic resonance imaging; PLS-DA, partial least squares discriminant analysis; VIP, variable importance in the projection; wt, wild-type

Colour Online: See the article online to view Figs. 1–4 in colour.

patient involvement, opportunity for rapid bulk processing of specimens, availability of reliable assays, and possibility of carrying out multiple analyses on a single sample [5, 7, 11].

Metabolomics aims to obtain a comprehensive coverage of low molecular mass compounds from biological systems [12–14]. Metabolomics studies can be approached using targeted or untargeted analysis [15–17]. In targeted analysis, a specified list of metabolites is analyzed. In contrast, untargeted analysis requires comprehensive metabolite measurements. Furthermore, it can implicate previously unrecognized metabolites or pathways with a unique phenotype and, therefore, is a powerful platform to elucidate novel biomarkers and gain insight into disease pathogenesis.

Different techniques are currently used for untargeted metabolomics, including NMR, GC-MS, LC-MS, and CE-MS [17–21]. For the first time to our knowledge, the use of on-line C_{18} -SPE-CE-MS is proposed as an alternative sensitive method for metabolomic studies of biological fluids, which are complex diluted samples. CE is a versatile, high-performance separation technique with many desirable characteristics such as instrumental simplicity, full automation, high efficiency, low consumption of sample and reagents, and reduced analysis times. However, like many other microanalytical techniques, it has poor concentration sensitivity for most analytes, from low molecular mass compounds to biopolymers such as proteins [22, 23]. Several strategies have been proposed to improve CE sensitivity. Today, SPE-CE is becoming widely recognized as a powerful approach that overcomes this major drawback [22–27]. In SPE-CE, a microcartridge placed inside and near the inlet of the separation capillary contains an appropriate extraction sorbent (in our case, C_{18}). This sorbent selectively retains the target analyte, enabling large volumes of sample to be introduced (≈ 50 – $100\ \mu\text{L}$). The retained analyte is eluted in a small volume of an appropriate solution (≈ 25 – $50\ \text{nL}$), which results in sample cleanup and concentration enhancement with minimum sample handling before separation and detection, for example, by on-line MS (SPE-CE-MS) [22–27].

Chemometric methods play a crucial role in data processing, exploration, and classification of the massive datasets generated in metabolomic studies [28–34]. If the goal of the study is the compound detection, the use of resolution methods such as multivariate curve resolution alternating least squares (MCR-ALS) can be an excellent alternative. MCR-ALS can resolve overlapped electrophoretic/chromatographic peaks from the collected data and provide the separation profiles and mass spectra of the constituents in the analyzed samples. This approach allows overcoming problems such as retention time shifts, background noise contributions, and differences in S/Ns among different injections. Several published articles focus on the application of MCR-ALS to solve similar problems in LC-MS [32, 33] and GC-MS [34]; but only a few studies have been previously reported combining CE-MS and MCR-ALS in metabolomic applications [35].

In this paper, we evaluate the capacity of SPE-CE-MS combined with advanced multivariate data analysis to pre-concentrate, separate, detect, and identify low molecular mass

metabolites in plasma samples from wild-type (wt) and HD (R6/1) mice of different ages (8, 12, and 30 weeks). A comparison between the different untargeted metabolomic profiles allows us to propose novel potential biomarker candidates involved in the progression of HD, which could be useful for prediction of disease onset or response to treatment.

2 Materials and methods

2.1 Chemicals and reagents

All the chemicals used in the preparation of buffers and solutions were of analytical reagent grade. ACN (99.9%), methanol (99.9%), 2-propanol ($\geq 99.9\%$), formic acid (HFor) (99.0%), glacial acetic acid (HAc), ammonia (25%), and sodium hydroxide ($\geq 99.0\%$, pellets) were purchased from Merck (Darmstadt, Germany). Methionine enkephalin (Met, $\geq 97.0\%$) and endomorphin 1 (End 1, $\geq 98.0\%$) were provided by Sigma (St. Louis, MO, USA). Dynorphin A (1–7) (Dyn A, $\geq 98.0\%$) was supplied by Bachem (Bubendorf, Switzerland). PEG 8000 M_r ($\sim 50\%$ in water) was purchased from Fluka (Buchs, Switzerland). Water with a conductivity value $< 0.05\ \mu\text{S}/\text{cm}$ was obtained using a Milli-Q water purification system (Millipore, Molsheim, France).

2.2 Electrolyte solutions, sheath liquid, and standard solutions

Aqueous standard solutions ($2500\ \mu\text{g}/\text{mL}$) of Dyn A, End 1, and Met peptides were prepared and stored in a freezer at -20°C when not in use. A $10\ \text{ng}/\text{mL}$ standard mixture of the three peptides was prepared and analyzed at the beginning and at the end of each SPE-CE-MS sequence, in order to check the proper functioning of the on-line SPE microcartridges. The BGE contained $50\ \text{mM}$ of HAc and $50\ \text{mM}$ of HFor and was adjusted to pH 3.50 with ammonia. The sheath liquid solution consisted of a hydroorganic mixture of 2-propanol/water (60:40, v/v) with $0.05\ \text{v}/\text{v}$ of HFor. All solutions were passed through a $0.45\ \mu\text{m}$ nylon filter (MSI, Westboro, MA, USA) before analysis and were stored at 4°C when not in use. The sheath liquid was degassed for 10 min by sonication before use.

2.3 Mice blood plasma and sample preparation

Plasma samples from male wt mice and R6/1 transgenic mice (B6CBA background) expressing exon 1 of mutant Huntingtin with 145 repeats (HD, R6/1) of different ages (8, 12, and 30 weeks; early, middle, and late disease stage, respectively), were kindly supplied by the Department of Cellular Biology, Immunology and Neurosciences (Faculty of Medicine, University of Barcelona) [36]. Blood from mice was collected by cardiac puncture in standard clinical vials and placed on ice. Plasma was separated from the blood cells, pooled,

deposited into polyethylene tubes, and frozen at -20°C . It is worth mentioning that due to the small amount of blood that was possible to extract from a single mouse (between 1 and 2 mL), each set of samples corresponded to the combination of the plasma obtained from four or five mice. All animal procedures were approved by the CEEA committee of the University of Barcelona and were in accordance with the European Communities Council Directive (2010/63/EU).

The sample pretreatment used for the analysis of low molecular mass compounds in plasma samples was described elsewhere [22, 37]. The off-line double-step pretreatment of plasma samples consisted of protein precipitation with cold ACN (plasma/ACN, 200:1200 μL) followed by centrifugal filtration with 10 000 M_r cutoff cellulose acetate filters (Amicon® Ultra-0.5, Millipore). Centrifugal filters were passivated before the first use with 5% v/v of PEG in water [37].

2.4 Apparatus and procedures

Measurements of pH were made with a Crison 2002 potentiometer and a Crison electrode 52-03 (Crison Instruments, Barcelona, Spain). Centrifugal filtration was carried out in a cooled Rotanta 460 centrifuge (Hettich Zentrifugen, Tuttlingen, Germany) for centrifugation at controlled temperature (25°C).

2.4.1 On-line SPE-CE-MS

The construction of the microcartridge or analyte concentrator for C_{18} -SPE-CE-MS was carried out as described elsewhere [22, 37]. All fused silica capillaries were supplied by Polymicro Technologies (Phoenix, AZ, USA). The microcartridge (7 mm $L_T \times 250 \mu\text{m}$ id $\times 360 \mu\text{m}$ od) was inserted inside the separation capillary (72 cm $L_T \times 75 \mu\text{m}$ id $\times 360 \mu\text{m}$ od), at 7.5 cm from the inlet, using two plastic sleeves. Previously, it was filled with the sorbent found in C_{18} Sep-pak cartridges (Waters, Milford, MA, USA). The sorbent particles were retained in the microcartridge between two frits (0.1 cm).

All capillary rinses were performed at high pressure (930 mbar). New separation capillaries were flushed with 1.0 M NaOH (20 min) and water (15 min) before inserting the microcartridge. This activation procedure was performed off-line to avoid the unnecessary entrance of NaOH into the MS system. Once inserted the microcartridge, the SPE-CE-MS capillaries were first conditioned by consecutive flushes of water (1 min), methanol (1 min), water (1 min), and BGE (3 min) at 930 mbar. Standard peptide mixture (Dyn A, End 1, and Met) or mice plasma samples were then introduced at 930 mbar for 10 min (approximately 60 μL using the Hagen-Poiseuille equation [38]). A final rinse with the BGE (2 min at 930 mbar) eliminated nonretained molecules and equilibrated the capillary before the elution. Retained compounds were eluted by injecting a solution of methanol/water (60:40, v/v) with 50 mM HAc and 50 mM HFor at 50 mbar for 10 s (approximately 50 nL [38]). Separation was carried out at 25°C by applying a voltage of 17 kV (normal polarity, cath-

ode in the outlet). Between runs, the capillary was rinsed for 2 min with water and 2 min with ACN, in order to avoid carry-over between consecutive analyses. In general, the different plasma samples (i.e. 8wt, 12wt, 30wt and 8HD, 12HD, and 30HD) were analyzed in triplicate (with the exception of 12wt, 12HD, and 30HD, for which only two replicates were analyzed due to the small volume of plasma sample available). Each series of replicate analyses was performed in a new SPE-CE-MS capillary due to the limited durability of the SPE microcartridges (≈ 10 analyses) because of the complexity of the plasma matrix and the limited selectivity of the C_{18} sorbent. After these analyses, the extraction efficiency decreased and the microcartridge was packed until it was completely clogged [39]. At the beginning and at the end of each sequence, a 10 ng/mL standard peptide mixture was analyzed as a quality control of the system.

All CE-MS experiments were performed in an HP^{3D} CE system coupled with an orthogonal G1603A sheath-flow interface to a 6220 oa-TOF LC/MS spectrometer (Agilent Technologies, Waldbronn, Germany). The sheath liquid was delivered at a flow rate of 3.3 $\mu\text{L}/\text{min}$ by a KD Scientific 100 series infusion pump (Holliston, MA, USA). ChemStation C.01.06 software (Agilent Technologies) was used for CE control and separation data acquisition (e.g. voltage, temperature, and current), and was run in combination with MassHunter B.04.00 workstation software (Agilent Technologies) for control of the mass spectrometer and MS data acquisition.

The TOF mass spectrometer was operated under optimum conditions in positive mode using the following parameters: capillary voltage 4000 V, drying gas temperature 200°C , drying gas flow rate 4 L/min, nebulizer gas 7 psig, fragmentor voltage 215 V, skimmer voltage 60 V, and OCT 1 RF Vpp voltage 300 V. Data were collected in profile at one spectrum/s between 40 and 1250 m/z , with the mass range set to high-resolution mode (4 GHz). A standard tune and an external mass calibration were performed daily at the beginning of the day following the manufacturer's instructions using the typical LC-MS sprayer and ESI-L tuning mix (Agilent Technologies).

2.5 Data analysis

SPE-CE-MS data were analyzed by a combination of advanced chemometric tools to evaluate the most significant metabolic changes involved in HD. Figure 1 shows a summary of the data analysis workflow, which is explained in detail in this section.

2.5.1 Data preprocessing of dataset

First, SPE-CE-MS raw data were converted to .txt format using the ProteoWizard software [40] and, then imported into the MATLAB environment (The Mathworks, Natick, MA, USA) using in-house made routines. During this import process, MS information was compressed to 0.01 Da/e resolution. Every sample provided a data matrix with 2490 rows (migration

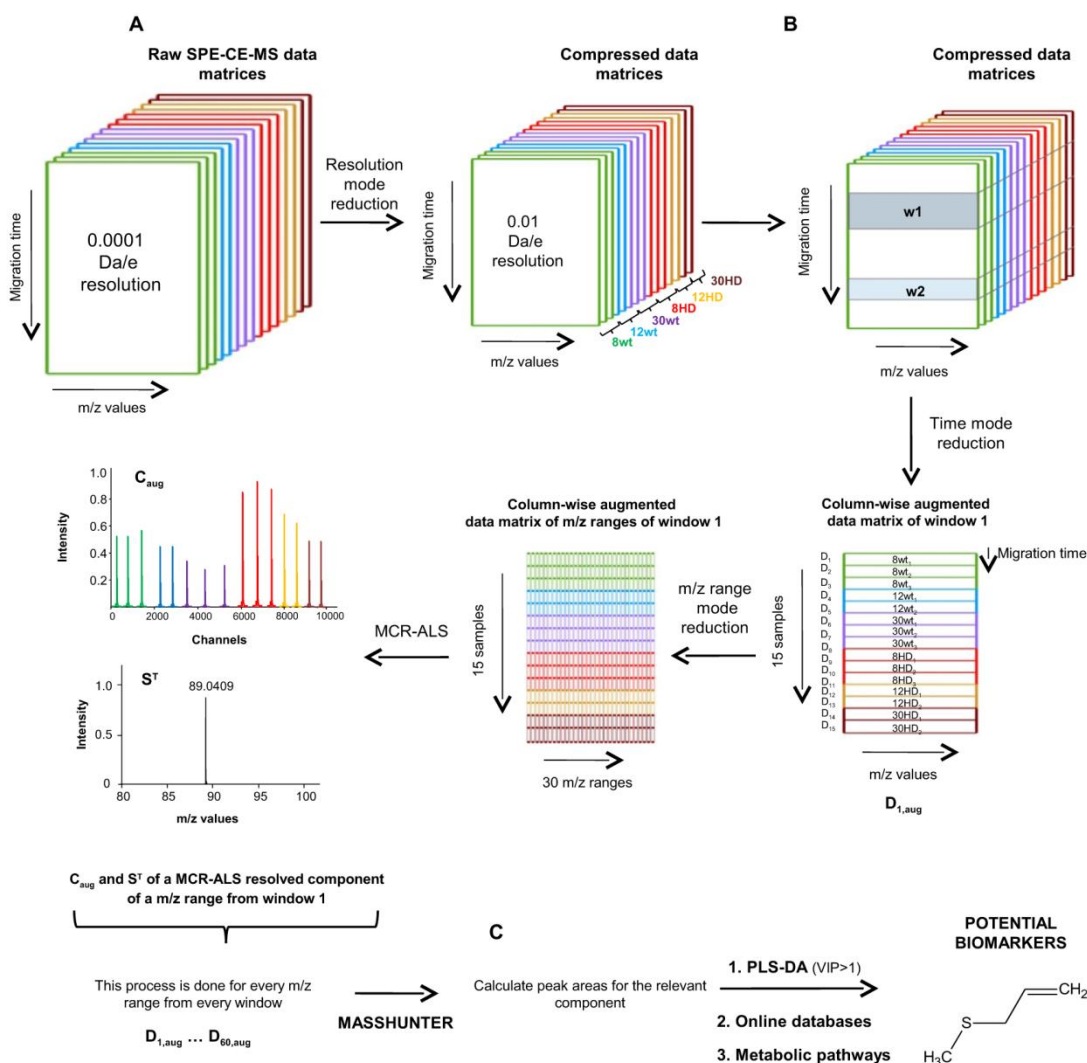


Figure 1. Workflow of C_{18} -SPE-CE-MS data analysis. (A) Preprocessing of dataset, (B) data arrangement and MCR-ALS analysis in order to detect metabolites, and (C) tentative identification of relevant metabolites and metabolic pathways.

times, 40 min of electrophoretic run) and 121 001 columns (m/z values, from 40 to 1250; see Fig. 1A). An automatic weighted least squares baseline correction was applied before to the MCR-ALS analysis.

2.5.2 Full scan MS data arrangement and MCR-ALS analysis

MCR-ALS is a chemometric method especially useful to analyze multicomponent systems with strongly overlapping contributions, such as those present in CE separations, where the electrophoretic behavior of metabolites is rather similar [41]. In the case of SPE-CE-MS, full scan MS data matrix D contains the experimental mass spectra at all retention times in their rows and the electropherograms at all m/z values in their columns. MCR-ALS analysis of the data matrix D ,

following a bilinear model, gives two factor matrices, C and S^T , as in Eq. (1):

$$D = CS^T + E, \quad (1)$$

where matrix C contains the electrophoretic profiles of the resolved contributions (components), matrix S^T contains the corresponding mass spectra of the resolved contributions, and matrix E contains the residuals unexplained by the model.

The different samples can be simultaneously analyzed and compared by MCR-ALS using a column-wise augmented data matrix configuration (see matrix D_{aug} in Eq. (2) and Fig. 1B), following the strategy described in the work of Ortiz-Villanueva [35]:

$$D_{aug} = \begin{bmatrix} D_1 \\ \vdots \\ D_{15} \end{bmatrix} = \begin{bmatrix} C_1 \\ \vdots \\ C_{15} \end{bmatrix} S^T + \begin{bmatrix} E_1 \\ \vdots \\ E_{15} \end{bmatrix} = C_{aug} S^T + E_{aug}. \quad (2)$$

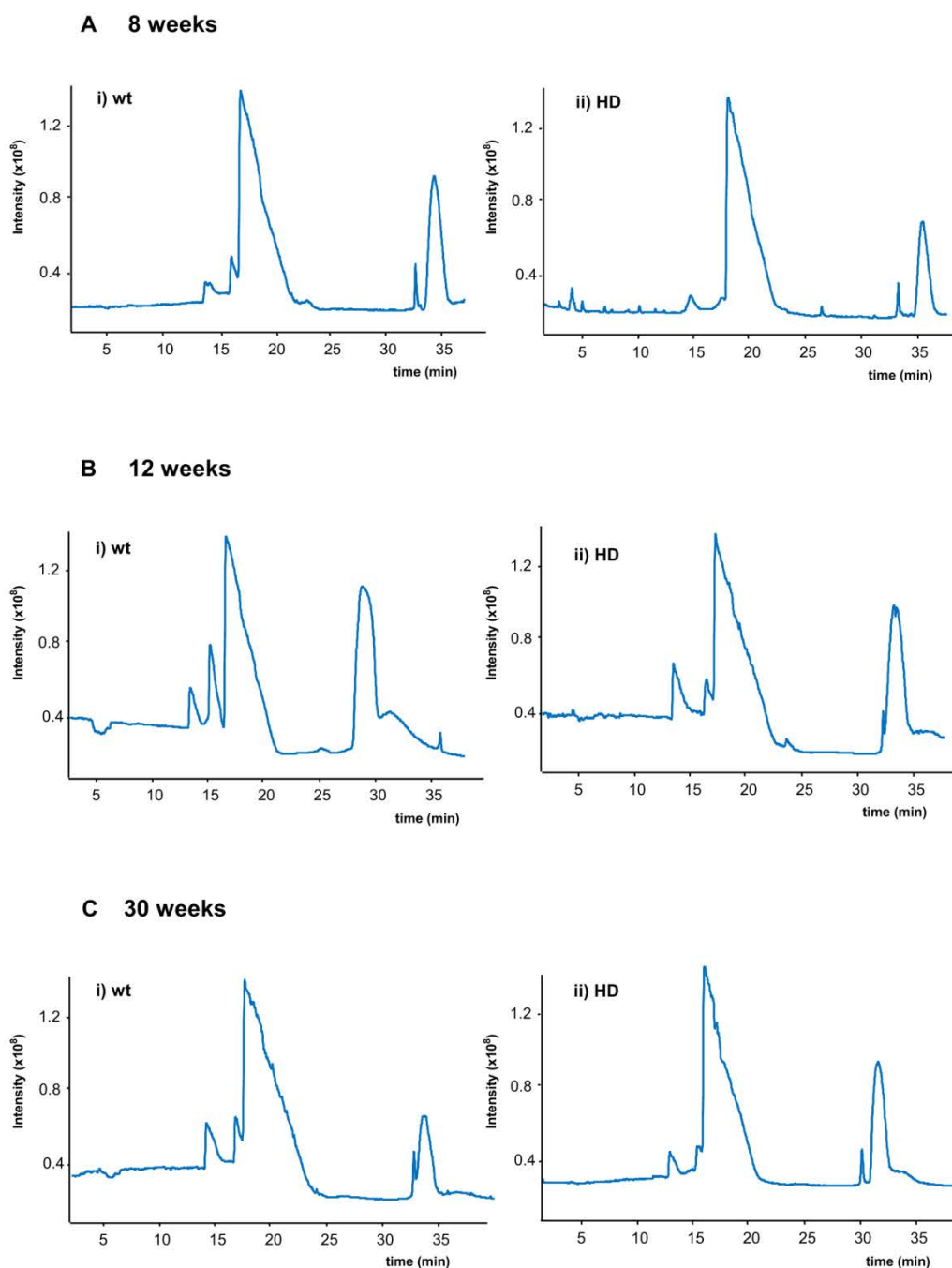


Figure 2. Total ion electropherograms obtained by C_{18} -SPE-CE-MS for (i) wt and (ii) HD plasma samples at (A) 8, (B) 12, and (C) 30 weeks.

This approach allowed obtaining a common matrix of the mass spectra of the resolved components (S^T) for all samples, and a set of matrices describing the resolved electrophoretic profiles (C_{aug}) in every sample. These electrophoretic peaks resolved in matrix C_{aug} are allowed to vary in position (shifts) and shape among samples because the only requirement for a proper resolution is that the resolved spectra are the same for the common constituents in the different samples [42]. This

aspect is especially useful in the case of CE data where migration shifts among samples occur and, hence, the alignment of electrophoretic peaks before analysis is not needed.

In this study, the electropherograms were partitioned in two time windows corresponding to the two regions with the most intense peaks (selected regions, depending on the sample, varied approximately from 10 to 25 min and from 30 to 40 min, respectively, Fig. 2). Then, the resulting data

matrices were further reduced in their m/z mode dimension in 30 different m/z ranges (m/z widths for reduction were 20, 50, and 100 m/z in the m/z ranges 40–400, 400–800, and 800–1250 m/z , respectively; Fig. 1B) [35].

MCR-ALS analysis was carried out following standard procedures for the determination of the number of components (SVD, [43]) and initial estimates (SIMPLISMA, [44]). ALS optimization was performed under nonnegativity constraints for electrophoretic (C_{aug}) and spectral (S^T) profiles, and spectral normalization (equal height) [45,46].

2.5.3 Detection and identification of potential metabolites

For every resolved MCR-ALS component, electropherogram (peak) profiles of the six sample sets (i.e. 8wt, 12wt, 30wt and 8HD, 12HD, and 30HD) were compared. Only resolved components of C_{aug} that showed S/Ns higher than 10% of the abundance of the most intense component were selected. Next, their corresponding mass spectra profiles (S^T) were used to identify the m/z values causing the differentiation between wt and HD plasma samples at 8, 12, and 30 weeks. Finally, peak areas of these candidate m/z values were recovered from the full scan SPE-CE-MS data using the MassHunter workstation software, taking as a reference the m/z value and the migration time of the MCR-ALS resolved components (Fig. 1C). Areas were finally normalized considering the peak area corresponding to a compound present in all the samples that was not discriminant between control and HD samples (m/z of 72.9858, in the first time window).

These areas were used to build a data matrix containing the area of each candidate (feature) in every sample. This data matrix was autoscaled in order to give equal weighting to all candidates in the measured samples. Finally, partial least squares discriminant analysis (PLS-DA) models were applied to the autoscaled data matrix to evaluate sample discrimination and to identify the most important features. There are numerous methods for feature selection when considering PLS models. In this work, the variable importance in the projection (VIP) method was used [47], because it is one of the preferred methods to deal with metabolomic data due to its ability for handling multicollinear data [48]. For each model, VIP scores estimate the importance of each feature in the projection. Only features with a VIP score over a particular threshold value (usually 1) are considered important and selected for further analysis. In all the cases, leave-one-out cross-validation was used to assess the performance of the built models. Thereafter, the accurate experimental molecular mass values of the finally VIP selected metabolites were searched in on-line databases resources, such as METLIN Metabolite Database [49] and Human Metabolome Database [50]. A small error from the calculated (theoretical) molecular mass (M_r) was used to evaluate the accuracy of possible molecular formulas ($E_r \leq 20$ ppm, $|M_{r \text{ experimental}} - M_{r \text{ theoretical}}| / M_{r \text{ theoretical}} \times 10^6$). Finally, the list of the tentatively identified metabolites was used to investigate the

possible metabolic pathways and mechanisms involved in HD according to the KEGG database [51] (Fig. 1C).

2.5.4 Software

Most of the calculations and data analysis were performed under MATLAB R2013a (The Mathworks). PLS Toolbox 7.3.1 (Eigenvector Research, Wenatchee, WA, USA) was used for PLS-DA and VIP calculations; and MCR-ALS toolbox [42] was used for resolution of electrophoretic and mass spectral metabolite profiles from full MS scan augmented data matrices.

3 Results and discussion

3.1 Analysis of mice plasma by C_{18} -SPE-CE-MS

Untargeted metabolomics analysis requires a comprehensive coverage of low molecular mass compounds from biological samples. However, very often sample amount limitations, matrix complexity, and metabolite concentration preclude direct analysis with CE-MS. With the aim of solving these issues, plasma samples from wt (control) and HD mice were analyzed by C_{18} -SPE-CE-MS in order to preconcentrate, separate, detect and identify low molecular mass compounds, and establish significant differences between the global metabolite profiles from different groups of samples. In order to evaluate HD progression in individuals at the premanifest motor stage of the disease, plasma samples from wt and HD mice were analyzed at 8, 12, and 30 weeks of age. In HD mice, these samples corresponded to early (asymptomatic), middle (symptomatic), and late (terminal) disease stage mice, respectively, although this classification is only based on motor coordination deficiencies [52].

The applied C_{18} -SPE-CE-MS method in positive ESI mode was developed for the analysis of peptides in human plasma in previous works [22, 37], but preliminary experiments showed that it was also useful to obtain a rich fingerprint of low molecular mass compounds in mouse plasma. As shown in those studies, all the plasma samples were subjected to an off-line sample pretreatment before C_{18} -SPE-CE-MS in order to prevent microcartridge saturation due to the limited selectivity of the C_{18} sorbent. A double-step pretreatment based on solvent precipitation and centrifugal filtration with M_r cutoff filters was applied to eliminate salts and high molecular mass compounds (i.e. proteins). This pretreatment allowed excellent recoveries for low molecular mass opioid peptides (>70%) [37]. Furthermore, LODs were improved by C_{18} -SPE-CE-MS between 1000 and 10 000 times compared to CE-MS, depending on the peptides and the sample [39]. Figure 2 shows the total ion electropherograms obtained for the mice plasma samples by C_{18} -SPE-CE-MS. As can be observed, separation resolution is not high because of the complexity of the sample. All the electropherograms present a characteristic profile with two time regions with the most

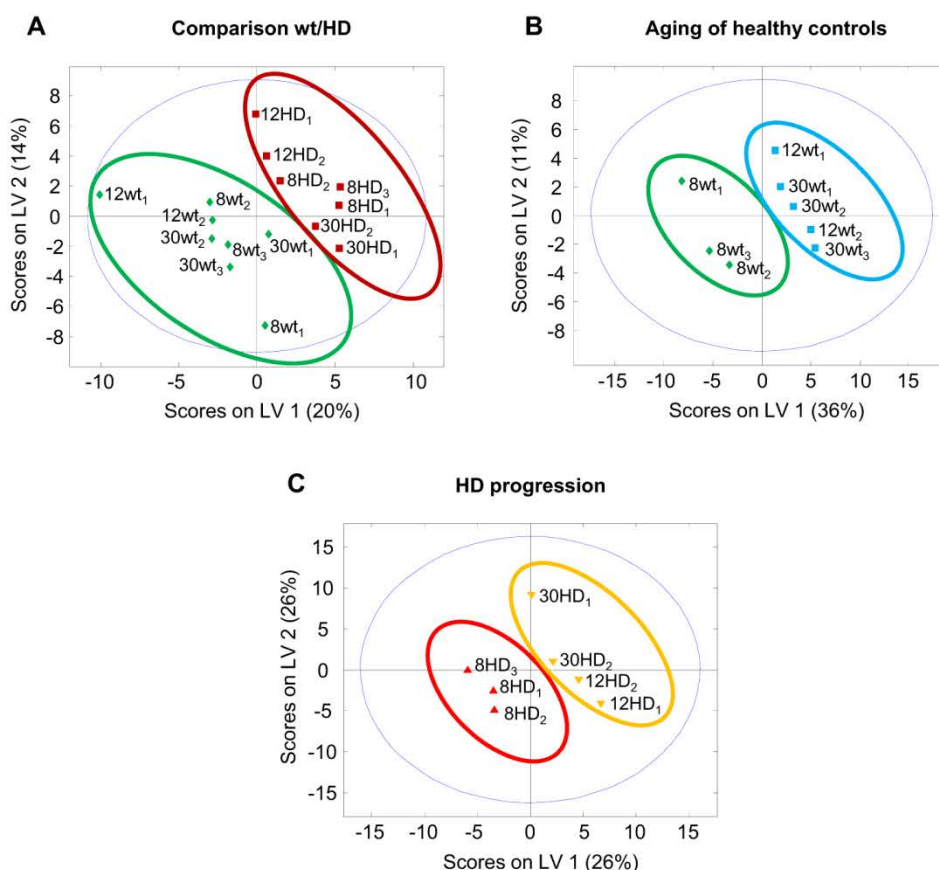


Figure 3. PLS-DA scores plot for the mice samples in order to differentiate between (A) all wt and HD plasma samples, (B) wt samples at 8, and from 12 to 30 weeks, and (C) HD samples at 8, and from 12 to 30 weeks. Every sample was analyzed in triplicate (with the exception of 12wt, 12HD, and 30HD, for which only two replicates were analyzed due to the small volume of plasma sample provided).

intense peaks (approximately at 10–25 and 30–40 min, respectively), and advanced chemometrics methods are necessary for high throughput and reliable comparison between the different sets of plasma samples.

3.2 MCR-ALS analysis and detection of the most relevant metabolites

MCR-ALS was applied using a column-wise augmented data matrix containing simultaneously the information of the 15 samples (wt and HD, both at 8, 12, and 30 weeks) and allowed the resolution of the electropherogram profiles and corresponding mass spectra of the plasma metabolites.

MCR-ALS analysis was performed separately on column-wise augmented data matrices of different m/z ranges (at the resolution of 0.01 Da/e), corresponding to the two selected time windows. A total number of 60 column-wise augmented matrices (two time windows \times 30 m/z intervals) were separately analyzed. The number of components selected was related to the number of electrophoretic peaks, despite the fact that some of these resolved components could be due to contributions such as solvent background or instrumental noise. In most of the cases, MCR-ALS models showed an explained variance (R^2) of almost 100%. The electropherogram profiles for the resolved MCR-ALS components in the six sample sets

(i.e. 8wt, 12wt, 30wt and 8HD, 12HD, and 30HD) were compared and only resolved components of C_{aug} that showed S/Ns higher than 10% of the abundance of the most intense component were finally selected (in order to remove contributions such as solvent background or instrumental noise). The mass spectra of these components (from S^T) were used to identify the m/z values causing the discrimination between samples. After the resolution and analysis of the 60 augmented data matrices, a total number of 74 features were detected. Finally, peak areas of these candidate m/z values were recovered from the full scan raw C_{18} -SPE-CE-MS data using the MassHunter workstation software, taking as a reference the m/z value and the migration time of features obtained from the MCR-ALS resolved components.

PLS-DA was then applied to identify the most important metabolites responsible for the sample discrimination considering the raw peak areas for the selected 74 candidate metabolites. In order to identify potential Huntington biomarkers that could be useful to discriminate between wt and HD samples, as well as to follow-up the HD progression, three different PLS-DA models were built. Figure 3 shows the PLS-DA scores plot for the mice plasma samples taking into account the three mentioned models. As can be observed in Fig. 3A, the first PLS-DA model was applied to discriminate between control and HD samples. This model permitted us to propose possible biomarkers involved in HD. Two latent

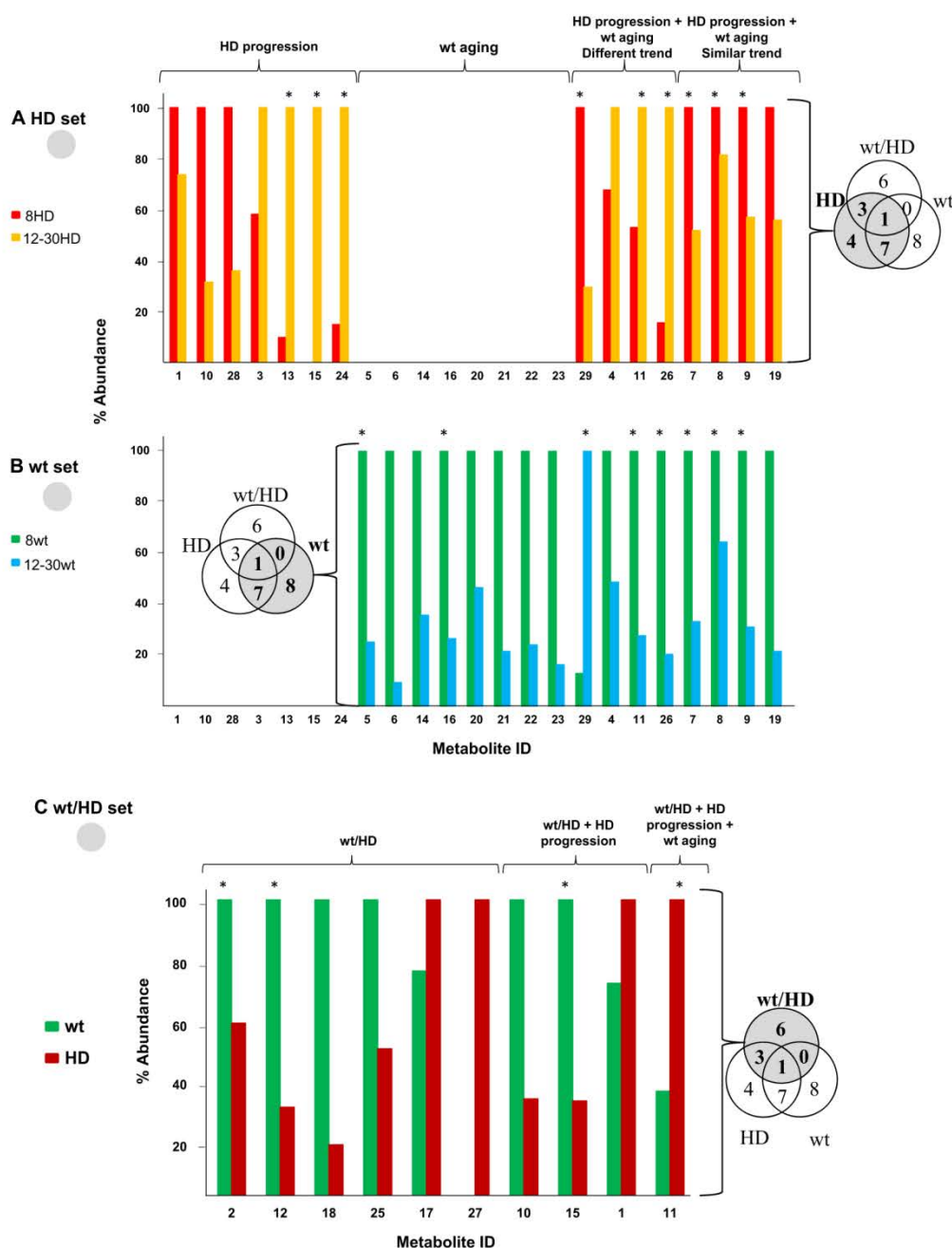


Figure 4. Bar graphs with the folding trends for the identified metabolites (Table 1) that were used to explain (A) HD progression (set HD), (B) aging of healthy controls (set wt), and (C) differences between wt and HD plasma samples (set wt/HD). The percent of Abundance of each metabolite was calculated normalizing to the metabolite presenting the highest abundance. (* See Table 2 for the related metabolic pathways.)

variables (LVs) explained 34 and 89% of the X and Y variances, respectively. The second PLS-DA model was applied to differentiate between wt samples of different ages and identify metabolites involved in aging of healthy controls. In order to improve the reliability of the PLS-DA model due to the limited amount of samples, a two-class model was used, which presented at least three samples in each class (8 and

12–30 weeks). These two sets of samples were also the best option to differentiate later between aging and early HD progression. A PLS-DA model with two LVs explained 47% of the X-variance and the 99% of the Y-variance (Fig. 3B). Finally, the third PLS-DA model was applied to distinguish between HD samples of different ages and identify possible biomarkers that could be useful to follow-up the disease progression.

Again, the same two sets of samples were defined (8 and 12–30 weeks). In this case, two LVs explained 52 and 98% of the X and Y variances, respectively, (see Fig. 3C). All PLS-DA models allowed class discrimination and the detection of the most relevant components for the differentiation of the samples. It is worth mentioning that HD samples of 12 and 30 weeks were slightly separated in the scores plot (Fig. 3C), whereas this separation was not observed for wt samples (Fig. 3B). Anyway, a three-class PLS-DA model was not recommended because of the limited amount of samples. VIP scores values higher than 1 were used as a feature selection tool in order to choose only the most relevant candidate metabolites for each PLS-DA model (33 out of 74).

3.3 Tentative metabolite identification and biological meaning

The most contributing metabolites to sample discrimination (33) were tentatively identified, taking advantage of the highly accurate experimental molecular mass values provided by the oa-TOF mass spectrometer. Only 29 features of the total of 33 were tentatively identified with an error ≤ 20 ppm (Table 1; the four nonidentified features were discarded for further discussion). As can be observed in Table 1, there were some ambiguities on the metabolite identities because this tentative identification was solely based on the agreement between the experimental and the theoretical molecular mass values. For example, in some cases several isobaric metabolites were proposed for a certain molecular formula and experimental molecular mass value (i.e. identification number (ID) 2, 3, 4, 6, 7, 9, 13, 15, 16, 17, 18, 19, and 29). In other cases, it was not possible to differentiate between metabolites with very close theoretical molecular mass values because $E_r \leq 20$ ppm (i.e. ID 6 [$E_r = 7$ and 14 ppm], ID 8 [$E_r = 16$ and 18 ppm], ID 9 [$E_r = 2$ and 15 ppm], ID 10 [$E_r = 13$ and 17 ppm], and ID 25 [$E_r = 7$ and 16 ppm]). In the future, analysis of standard samples and MS/MS measurements for structure characterization would be necessary to improve reliability of these identity assignments.

The Venn diagrams that appear as insets in Fig. 4A and B show the relations between the identified metabolites that explain HD progression and aging of healthy controls. As can be observed, seven metabolites (4 + 3) were useful to specifically explain HD progression (HD set, ID 1, 10, 28, 3, 13, 15, and 24 in Table 1 and Fig. 4). Similarly, eight metabolites (8 + 0, ID 5, 6, 14, 16, 20, 21, 22, and 23 in Table 1 and Fig. 4) were useful to specifically explain aging of healthy controls (wt set). The concentration trends of these specific metabolites were varied (Fig. 4A and B), some of them decrease, while others increase after 12 weeks of birth. Finally, there were eight metabolites (7 + 1, ID 29, 4, 11, 26, 7, 8, 9, and 19 in Table 1 and Fig. 4) that were explaining both progression of HD and aging. Four of them showed a clear different concentration trend in HD and wt plasma samples, but for the other four metabolites the trend was similar, indicating that differences were found on their absolute concentration (e.g.

ID 8 normalized areas in HD and wt plasma samples were 466.4072 and 431.2187, respectively). With regard to differentiation in general of wt and HD samples (wt/HD set), there were six metabolites that were useful to specifically distinguish between control and HD samples, four downregulated and two upregulated in HD samples, as shown in Fig. 4C (ID 2, 12, 18, 25, 17, and 27 in Table 1). For metabolites explaining also HD progression and/or aging (ID 10, 15, 1, and 11 in Table 1 and Fig. 4C), the concentration trends were varied (two were downregulated and two upregulated).

The identified metabolites were searched against different on-line databases to identify the potential metabolic pathways that could be involved in HD pathology. Different metabolic pathways were found to be related to 13 of the 29 identified metabolites (see Table 2). It is well known that HD could affect different metabolic pathways. Huntingtin is ubiquitously expressed and, in addition to neurological features, the peripheral phenotype of HD could include weight loss, energy disturbances, and alteration of endocrine function.

As is shown in Fig. 4A, concentrations of phenylalanyl-arginine and arginyl-phenylalanine (ID 13, Table 1) were found increased in HD mice after 12 weeks of birth (Fig. 4A). These metabolites, which were specific to explain HD progression, are incomplete breakdown products of protein digestion or protein catabolism known to have physiological or cell-signaling effects (Table 2) [53]. Similarly, prostaglandins, thromboxanes, lipoxins, and leukotrienes (ID 15, Table 1) were found upregulated after 12 weeks (Fig. 4A), but downregulated when all HD samples were compared to all controls (Fig. 4C), thus indicating a change of trend after 12 weeks. These metabolites are related with regulation of inflammatory processes and signaling pathways, mainly the arachidonic acid metabolism, the neuroactive ligand-receptor interaction, the serotonergic synapse, the cAMP signaling pathway, and the oxytocin signaling pathway (see Table 2). The arachidonic acid metabolism has also been related with the synthesis of cytochromes involved in the mitochondrial oxidative phosphorylation, and altered mitochondrial function has been associated to HD [54,55]. Furthermore, cAMP levels have been found reduced in the striatum of several HD mouse models [56], while the oxytocin signaling pathway has been related with changes in the hypothalamic and limbic systems that take place at HD early stages [57]. Concentration of L-urobilinogen (ID 24, Table 1), which is related with the porphyrin metabolism (Table 2), was also found increased after 12 weeks of birth (Fig. 4A). All these changes in 12-weeks-old HD mice suggest an onset on specific neuronal dysfunction, altered expression of several types of receptors and changed expression of neurotransmitters and key proteins. Unbalanced activity within these pathways provides a potential mechanism for many of the pathological phenotypes associated with HD, such as transcriptional dysregulation, inflammation, and ultimately neurodegeneration [58–60].

With regard to metabolites explaining both progression of HD and aging (Fig. 4A and B), gangliosides (ID 29, Table 1), which are cell plasma membrane components that

Table 1. Tentative identification of the relevant metabolites detected by the combination of C₁₈-SPE-CE-MS and MCR-ALS ($E_r \leq 20$ ppm)

ID	Metabolite	Time window (min)	Molecular formula	Theoretical molecular mass (Da)	Adduct	Experimental m/z (Da/e)	Theoretical m/z (Da/e)	E_r (ppm)
1	3-(Methylthio)-1-propene	10–25	C ₄ H ₈ S	88.0347	[M+H] ⁺	89.0401	89.0419	20
2	m-Cresol	30–40	C ₇ H ₈ O	108.0575	[M+H] ⁺	109.0653	109.0648	5
	p-Cresol		C ₇ H ₈ O	108.0575	[M+H] ⁺	109.0653	109.0648	5
3	3-Methyl-1,2-cyclopentanedione	30–40	C ₆ H ₈ O ₂	112.0524	[M+H] ⁺	113.0602	113.0597	4
	2,5-Dimethyl-3(2H)-furanone		C ₆ H ₈ O ₂	112.0524	[M+H] ⁺	113.0602	113.0597	4
	(2E,4E)-2,4-Hexadienoic acid		C ₆ H ₈ O ₂	112.0524	[M+H] ⁺	113.0602	113.0597	4
	(E)-4-Oxo-2-hexen-1-al		C ₆ H ₈ O ₂	112.0524	[M+H] ⁺	113.0602	113.0597	4
	1,2-Cyclohexanedione		C ₆ H ₈ O ₂	112.0524	[M+H] ⁺	113.0602	113.0597	4
	trans-1,2-Dihydrobenzene-1,2-diol		C ₆ H ₈ O ₂	112.0524	[M+H] ⁺	113.0602	113.0597	4
4	2-Hepten-4-one	30–40	C ₇ H ₁₂ O	112.0888	[M+H] ⁺	113.0968	113.0961	6
	(Z)-4-Heptenal		C ₇ H ₁₂ O	112.0888	[M+H] ⁺	113.0968	113.0961	6
	3-Hepten-2-one		C ₇ H ₁₂ O	112.0888	[M+H] ⁺	113.0968	113.0961	6
	2-Methylcyclohexanone		C ₇ H ₁₂ O	112.0888	[M+H] ⁺	113.0968	113.0961	6
	trans-2-trans-4-Heptadien-1-ol		C ₇ H ₁₂ O	112.0888	[M+H] ⁺	113.0968	113.0961	6
	5-Methyl-5-hexen-2-one		C ₇ H ₁₂ O	112.0888	[M+H] ⁺	113.0968	113.0961	6
	4-Methylcyclohexanone		C ₇ H ₁₂ O	112.0888	[M+H] ⁺	113.0968	113.0961	6
	5-Methyl-3-hexen-2-one		C ₇ H ₁₂ O	112.0888	[M+H] ⁺	113.0968	113.0961	6
5	Dimethylbenzimidazole	30–40	C ₉ H ₁₀ N ₂	146.0844	[M+H] ⁺	147.0919	147.0917	2
6	Hydantoin-5-propionic acid	30–40	C ₆ H ₈ N ₂ O ₄	172.0484	[M+Na] ⁺	195.0389	195.0376	7
	Menadione		C ₁₁ H ₈ O ₂	172.0524	[M+Na] ⁺	195.0389	195.0416	14
	Methyl (Z)-2-decene-4,6,8-triynoate		C ₁₁ H ₈ O ₂	172.0524	[M+Na] ⁺	195.0389	195.0416	14
7	(–)-Epinephrine	10–25	C ₉ H ₁₃ NO ₃	183.0895	[M+Na] ⁺	206.0747	206.0788	20
	Normetanephrine		C ₉ H ₁₃ NO ₃	183.0895	[M+Na] ⁺	206.0747	206.0788	20
8	Vanylglycol	10–25	C ₉ H ₁₂ O ₄	184.0736	[M+Na] ⁺	207.0594	207.0628	16
	Phosphorylcholine		C ₅ H ₁₅ NO ₄ P	184.0739	[M+Na] ⁺	207.0594	207.0631	18
9	3-Indolebutyric acid	10–25	C ₁₂ H ₁₃ NO ₂	203.0946	[M+Na] ⁺	226.0833	226.0838	2
	Glycyl-glutamine		C ₇ H ₁₃ N ₃ O ₄	203.0906	[M+Na] ⁺	226.0833	226.0798	15
	Glycyl-gamma-glutamete		C ₇ H ₁₃ N ₃ O ₄	203.0906	[M+Na] ⁺	226.0833	226.0798	15
	Asparaginy-alanine		C ₇ H ₁₃ N ₃ O ₄	203.0906	[M+Na] ⁺	226.0833	226.0798	15
	Glutaminy-glycine		C ₇ H ₁₃ N ₃ O ₄	203.0906	[M+Na] ⁺	226.0833	226.0798	15
	Alanyl-asparagine		C ₇ H ₁₃ N ₃ O ₄	203.0906	[M+Na] ⁺	226.0833	226.0798	15
	Gamma-glutamyl-glycine		C ₇ H ₁₃ N ₃ O ₄	203.0906	[M+Na] ⁺	226.0833	226.0798	15
10	Propyl propane thiosulfonate	30–40	C ₁₁ H ₁₆ O ₅	228.0998	[M+Na] ⁺	251.0857	251.0890	13
	5-Methoxycanthin-6-one		C ₁₅ H ₁₀ N ₂ O ₂	250.0742	[M+H] ⁺	251.0857	251.0815	17
11	L-Hexanoylcarnitine	30–40	C ₁₃ H ₂₅ NO ₄	259.1784	[M+H] ⁺	260.1858	260.1858	0
12	Histidiny-histidine	30–40	C ₁₂ H ₁₆ N ₆ O ₃	292.1284	[M+H] ⁺	293.1327	293.1357	10
13	Phenylalanyl-arginine	10–25	C ₁₅ H ₂₃ N ₅ O ₃	321.1801	[M+H] ⁺	322.1920	322.1873	14
	Arginyl-phenylalanine		C ₁₅ H ₂₃ N ₅ O ₃	321.1801	[M+H] ⁺	322.1920	322.1873	14
14	2-[Octahydro-4,7-dimethyl-1-oxocyclopenta[c]pyran-3-yl]nepetalactam	30–40	C ₂₀ H ₂₉ NO ₃	331.2147	[M+H] ⁺	332.2186	332.2220	10
15	Prostaglandin D2	10–25	C ₂₀ H ₃₂ O ₅	352.2250	[M+Na] ⁺	375.2097	375.2142	12
	Prostaglandin E2		C ₂₀ H ₃₂ O ₅	352.2250	[M+Na] ⁺	375.2097	375.2142	12
	Prostaglandin H2		C ₂₀ H ₃₂ O ₅	352.2250	[M+Na] ⁺	375.2097	375.2142	12
	Prostaglandin I2		C ₂₀ H ₃₂ O ₅	352.2250	[M+Na] ⁺	375.2097	375.2142	12
	Thromboxane A2		C ₂₀ H ₃₂ O ₅	352.2250	[M+Na] ⁺	375.2097	375.2142	12

(continued)

Table 1. Continued

ID	Metabolite	Time window (min)	Molecular formula	Theoretical molecular mass (Da)	Adduct	Experimental m/z (Da/e)	Theoretical m/z (Da/e)	E_r (ppm)
	Lipoxin A4		C ₂₂ H ₂₈ N ₂ O ₂	352.2250	[M+Na] ⁺	375.2097	375.2142	12
	Lipoxin B4		C ₂₀ H ₃₂ O ₅	352.2250	[M+Na] ⁺	375.2097	375.2142	12
	20-Hydroxy-leukotriene B4		C ₂₀ H ₃₂ O ₅	352.2250	[M+Na] ⁺	375.2097	375.2142	12
	15-Keto-prostaglandin F2a		C ₂₀ H ₃₂ O ₅	352.2250	[M+Na] ⁺	375.2097	375.2142	12
16	18-Hydroxycorticosterone	10–25	C ₂₁ H ₃₀ O ₅	362.2093	[M+H] ⁺	363.2100	363.2166	18
	Cortisol		C ₂₁ H ₃₀ O ₅	362.2093	[M+H] ⁺	363.2100	363.2166	18
17	alpha-Peroxyachifolide	10–25	C ₂₀ H ₂₄ O ₇	376.1522	[M+H] ⁺	377.1531	377.1595	17
	Diosbulbin F		C ₂₀ H ₂₄ O ₇	376.1522	[M+H] ⁺	377.1531	377.1595	17
	Diosbulbin A		C ₂₀ H ₂₄ O ₇	376.1522	[M+H] ⁺	377.1531	377.1595	17
	Carissanol		C ₂₀ H ₂₄ O ₇	376.1522	[M+H] ⁺	377.1531	377.1595	17
18	PGD2-	10–25	C ₂₃ H ₃₉ NO ₆	425.2778	[M+Na] ⁺	448.2600	448.2670	16
	dihydroxypropylamine							
	PGE2-		C ₂₃ H ₃₉ NO ₆	425.2778	[M+Na] ⁺	448.2600	448.2670	16
	dihydroxypropylamine							
19	Artonol E	10–25	C ₂₆ H ₂₄ O ₇	448.1522	[M+H] ⁺	449.1605	449.1595	2
	Cycloartomunoxanthone		C ₂₆ H ₂₄ O ₇	448.1522	[M+H] ⁺	449.1605	449.1595	2
20	6-Deoxohomodolichosterone	10–25	C ₂₉ H ₅₀ O ₄	462.3709	[M+Na] ⁺	485.3584	485.3601	4
21	Limocitrin 3-glucoside	10–25	C ₂₃ H ₂₄ O ₁₃	508.1217	[M+Na] ⁺	531.1156	531.1109	9
22	Biochanin A	10–25	C ₂₅ H ₂₄ O ₁₃	532.1217	[M+H] ⁺	533.1183	533.1290	20
	7-(6-malonylglucoside)							
23	Maclurin 3-C-(6''-p-hydroxybenzoyl-glucoside)	10–25	C ₂₆ H ₂₄ O ₁₃	544.1217	[M+H] ⁺	545.1295	545.1289	1
24	L-Urobilinogen	10–25	C ₃₃ H ₄₈ N ₄ O ₆	596.3574	[M+H] ⁺	597.3724	597.3647	13
25	Tsugarioside B	30–40	C ₃₇ H ₆₀ O ₇	616.4339	[M+2Na] ²⁺	331.2084	331.2062	7
	Oryzanol C		C ₄₁ H ₆₀ O ₄	616.4492	[M+2Na] ²⁺	331.2084	331.2138	16
26	PC(14:1(9Z)/14:1(9Z))	10–25	C ₃₆ H ₆₈ NO ₈ P	673.4683	[M+2Na] ²⁺	359.7223	359.7233	3
27	Nostoxanthin sulfate	10–25	C ₄₀ H ₅₅ NaO ₇ S	702.3566	[M+H] ⁺	703.3700	703.3639	10
28	PI(13:0/12:0)	10–25	C ₃₄ H ₆₅ O ₁₃ P	712.4164	[M+2Na] ²⁺	379.1987	379.1974	3
29	Ganglioside GD1b (d18:1/12:0)	30–40	C ₇₉ H ₁₃₇ N ₃ O ₃₉	1751.8829	[M+2Na] ²⁺	898.9184	898.9307	14
	Ganglioside GD1a (d18:1/12:0)		C ₇₉ H ₁₃₇ N ₃ O ₃₉	1751.8829	[M+2Na] ²⁺	898.9184	898.9307	14

These are the experimental m/z (Da/e) of the four nonidentified features: 63.0249 (10–25 min), 179.0635 (10–25 min), 224.0858 (10–25 min), and 347.7141 (30–40 min).

modulate cell signal transduction events, showed a different concentration trend on HD progression compared to aging. Gangliosides levels decreased after 12 weeks of birth in HD progression (Fig. 4A), while increased in controls (Fig. 4B). Decreased ganglioside concentration has also been found in the cerebellum of R6/1 (HD) mice at 35–40 weeks [61], and in fibroblasts, cortex, and striatum of YAC128 mice [62]. Similarly, L-hexanoylcarnitine levels (ID 11, Table 1), which decreased with aging in healthy controls (Fig. 4B), were found to increase with HD progression (Fig. 4A), and also when all HD samples were compared to all controls (Fig. 4C), suggesting that the disease involves disturbances in energy production, which are characterized by production and excretion of unusual acylcarnitines [63]. Concentration of PC(14:1(9Z)/14:1(9Z)) (ID 26, Table 1), which is related with signaling pathways (the arachidonic acid metabolism and the retrograde endocannabinoid signaling), the glycerophospholipid metabolism and the linoleic acid metabolism, was also found increased with HD progression and decreased with wt aging (Table 2, Fig. 4A and B). In contrast, changes on the concentration trend with HD progression or aging of (–)-epinephrine and normetanephrine (ID 7, Table 1), which are

metabolites related with tyrosine metabolism and signaling pathways (cAMP signaling pathway, adrenergic signaling in cardiomyocytes, and neuroactive ligand-receptor interaction; Table 2) were not observed (Fig. 4A and B). These metabolites were found decreased after 12 weeks in HD progression and aging (Fig. 4A and B). The same trend was observed for vanlyglycol and phosphorylcholine (ID 8, Table 1, Fig. 4A and B), which are related with the tyrosine and the glycerophospholipid metabolisms, respectively (Table 2). Finally, metabolites with ID 9 (Table 1), presented again a decreasing trend in both HD progression and wt aging (Fig. 4A and B). In this case, 3-indolebutyric acid is related with the tryptophan metabolism, while the other metabolites are incomplete products of protein digestion or protein catabolism associated with cell-signaling effects (Table 2) [64].

With regard to metabolites explaining only wt aging, dimethylbenzimidazole (ID 5, Table 1), which is related with the riboflavin and porphyrin metabolisms (Table 2), was found reduced after 12 weeks of birth (Fig. 4B). The same concentration trend was observed for 18-hydroxycorticosterone and cortisol (ID 16, Table 1, Fig. 4B), which are metabolites associated with the steroid hormone biosynthesis (Table 2).

Table 2. Metabolic pathways related to some of the identified metabolites

ID	Compound	Metabolic pathways
2	m-Cresol	Degradation of aromatic compounds
	p-Cresol	Degradation of aromatic compounds/protein digestion and absorption
5	Dimethylbenzimidazole	Riboflavin metabolism / Porphyrin metabolism
7	(–)-Epinephrine	cAMP signaling pathway*/adrenergic signaling in cardiomyocytes*/neuroactive ligand-receptor interaction*/tyrosine metabolism
	Normetanephrine	Tyrosine metabolism
8	Vanylglycol	Tyrosine metabolism
	Phosphorylcholine	Glycerophospholipid metabolism
9	3-Indolebutyric acid	Tryptophan metabolism
	Glycyl-glutamine	Incomplete breakdown product of protein digestion or protein catabolism with cell-signaling effects*
	Glycyl-gamma-glutamate	Incomplete breakdown product of protein digestion or protein catabolism with cell-signaling effects*
	Asparaginy-alanine	Incomplete breakdown product of protein digestion or protein catabolism with cell-signaling effects*
	Glutaminy-glycine	Incomplete breakdown product of protein digestion or protein catabolism with cell-signaling effects*
	Alanyl-asparagine	Incomplete breakdown product of protein digestion or protein catabolism with cell-signaling effects*
	Gamma-glutamyl-glycine	Incomplete breakdown product of protein digestion or protein catabolism with cell-signaling effects*
11	L-Hexanoylcarnitine	Energy production
12	Histidinyl-histidine	Incomplete breakdown product of protein digestion or protein catabolism with cell-signaling effects*
13	Phenylalanyl-arginine	Incomplete breakdown product of protein digestion or protein catabolism with cell-signaling effects*
	Arginyl-phenylalanine	Incomplete breakdown product of protein digestion or protein catabolism with cell-signaling effects*
15	Prostaglandin D2	Arachidonic acid metabolism*/neuroactive ligand-receptor interaction*/serotonergic synapse*/Fc epsilon RI signaling pathway*
	Prostaglandin E2	Arachidonic acid metabolism*/neuroactive ligand-receptor interaction*/serotonergic synapse*/cAMP signaling pathway*/oxytocin signaling pathway*/inflammatory mediator regulation of TRP channels*
	Prostaglandin H2	Arachidonic acid metabolism*/serotonergic synapse*/retrograde endocannabinoid signaling*/oxytocin signaling pathway*
	Prostaglandin I2	Arachidonic acid metabolism*/neuroactive ligand-receptor interaction*/cAMP signaling pathway*/VEGF signaling pathway*
	Thromboxane A2	Arachidonic acid metabolism*/neuroactive ligand-receptor interaction*/serotonergic synapse*
	Lipoxin A4	Arachidonic acid metabolism*/neuroactive ligand-receptor interaction*
15	Lipoxin B4	Arachidonic acid metabolism*
	20-Hydroxy-leukotriene B4	Arachidonic acid metabolism*
	15-Keto-prostaglandin F2a	Arachidonic acid metabolism*
16	18-Hydroxycorticosterone	Steroid hormone biosynthesis
	Cortisol	Steroid hormone biosynthesis
24	L-Urobilinogen	Porphyrin metabolism
26	PC(14:1(9Z)/14:1(9Z))	Arachidonic acid metabolism*/retrograde endocannabinoid signaling*/glycerophospholipid metabolism/linoleic acid metabolism
29	Ganglioside GD1b (d18:1/12:0)	Signal transduction*
	Ganglioside GD1a (d18:1/12:0)	Signal transduction*

*All these metabolic pathways are included in the signaling pathway.

Comparing all HD samples with all controls, concentration levels of m-cresol and p-cresol (ID 2, Table 1), which are involved in protein digestion and absorption, as well as in degradation of aromatic compounds, were found downregu-

lated in HD samples (Fig. 4C). The same concentration trend was observed for histidinyl-histidine (ID 12, Table 1, Fig. 4C), an incomplete breakdown product of protein digestion or catabolism with cell-signaling effects [65, 66].

4 Concluding remarks

An optimized sample pretreatment was applied to wt and R6/1 mice plasma samples (of 8, 12, and 30 weeks) prior to the analysis by C₁₈-SPE-CE-MS. The proposed methodology demonstrated to be suitable to ensure a reliable and comprehensive metabolite profiling of the plasma samples. The combination of MCR-ALS with other chemometric tools, such as PLS-DA, allowed the comprehensive analysis of the C₁₈-SPE-CE-MS metabolomic data, resolving electrophoretic peaks and mass spectra of a large number of metabolites. Finally, a list of potential metabolites useful to discriminate between control and HD plasma samples, as well as to follow-up the HD progression, were tentatively identified, and the most affected metabolic pathways were discussed. Although different pathways were found altered in HD, the intracellular signaling was observed to be the most affected, especially after 12 weeks of birth, thus suggesting that the pathology involves dysfunction of specific neurons, altered expression of several types of receptors, and changed expression of neurotransmitters. In addition, although some of the identified metabolites have been previously described in the striatum of R6/1 (HD) mice or other rat models, attempts to find such biomarkers in plasma have hitherto been unsuccessful. In the present work, we propose direct brain-striatal metabolites as good biomarkers that can be found in periphery (plasma samples). Therefore, we provide a window of opportunity for prediction of disease onset, evaluation of HD early progression, or response to treatment.

L. P. acknowledges the Spanish Ministry of Economy and Competitiveness for an FPI fellowship. This study was supported by a grant from the Spanish Ministry of Education and Science (CTQ2011-27130). Part of the study was supported by the (FP/2007-2013)/ERC Grant Agreement No. 320737. We also thank Josep Maria Marimón for the blood sample collection.

The authors have declared no conflict of interest.

5 References

- [1] Mastrolakos, A., Ariyurek, Y., Goeman, J. J., van Duijn, E., Roos, R. A., van der Mast, R. C., van Ommen, G. B., den Dunnen, J. T., 't Hoen, P. A., van Roon-Mom, W. M., *Eur. J. Hum. Genet.* 2015, 23, 1349–1356.
- [2] Zielonka, D., Mielcarek, M., Landwehrmeyer, G. B., *Parkinsonism Rel. Disord.* 2015, 21, 169–178.
- [3] van den Bogaard, S. J., Dumas, E. M., Teeuwisse, W. M., Kan, H. E., Webb, A., van Buchem, M. A., Roos, R. A., van der Grond, J., *J. Huntingtons Dis.* 2014, 3, 377–386.
- [4] Kim, S. D., Fung, V. S., *Mov. Disord.* 2014, 27, 477–483.
- [5] Andre, R., Scahill, R. I., Haider, S., Tabrizi, S. J., *Drug Discov. Today* 2014, 19, 972–979.
- [6] Ross, C. A., Aylward, E. H., Wild, E. J., Langbehn, D. R., Long, J. D., Warner, J. H., Scahill, R. I., Leavitt, B. R., Stout, J. C., Paulsen, J. S., Reilmann, R., Unschuld, P. G., Wexler, A., Margolis, R. L., Tabrizi, S. J., *Nat. Rev. Neurol.* 2014, 10, 204–216.
- [7] Weir, D. W., Sturrock, A., Leavitt, B. R., *Lancet. Neurol.* 2011, 10, 573–590.
- [8] Ross, C. A., Shoulson, I., *Parkinsonism Rel. Disord.* 2009, 15S3, S135–S138.
- [9] Niccolini, F., Politis, M., *World J. Radiol.* 2014, 6, 301–312.
- [10] Rees, E. M., Scahill, R. I., Hobbs, N. Z., *J. Huntingtons Dis.* 2013, 2, 21–39.
- [11] Krzysztoń-Russjan, J., Zielonka, D., Jackiewicz, J., Kuśmirek, S., Bubko, I., Klimberg, A., Marcinkowski, J. T., Anuszczyńska, E. L., *J. Bioenerg. Biomembr.* 2013, 45, 71–85.
- [12] Hirayama, A., Wakayama, M., Soga, T., *Trends Anal. Chem.* 2014, 61, 215–222.
- [13] Villas-Bôas, S. G., Mas, S., Akesson, M., Smedsgaard, J., Nielsen, J., *Mass Spectrom. Rev.* 2005, 24, 613–646.
- [14] Ramautar, R., Shyti, R., Schoenmaker, B., De Groote, L., Derks, R. J., Ferrari, M. D., Van Den Maagdenberg, A. M., Deelder, A. M., Mayboroda, O. A., *Anal. Bioanal. Chem.* 2012, 404, 2895–2900.
- [15] Meier, F., Garrard, K. P., Muddiman, D. C., *Rapid Commun. Mass Spectrom.* 2014, 28, 2461–2470.
- [16] Bignardi, C., Cavazza, A., Corradini, C., Salvadeo, P., *J. Chromatogr. A* 2014, 1372, 133–144.
- [17] Rao, J. U., Engelke, U. F., Sweep, F. C., Pacak, K., Kusters, B., Goudswaard, A. G., Hermus, A. R., Mensenkamp, A. R., Eisenhofer, G., Qin, N., Richter, S., Kunst, H. P., Timmers, H. J., Wevers, R. A., *J. Clin. Endocrinol. Metab.* 2015, 100, E214–E222.
- [18] Gertsman, I., Gangoiti, J. A., Barshop, B. A., *Metabolomics* 2014, 10, 312–323.
- [19] Arbulu, M., Sampedro, M. C., Gómez-Caballero, A., Goicolea, M. A., Barrio, R. J., *Anal. Chim. Acta* 2014, 858, 32–41.
- [20] Wehrens, R., Weingart, G., Mattivi, F., *J. Chromatogr. B* 2014, 966, 109–116.
- [21] Carty, D. M., Siwy, J., Brennand, J. E., Zürlbig, P., Mullen, W., Franke, J., McCulloch, J. W., North, R. A., Chappell, L. C., Mischak, H., Poston, L., Dominiczak, A. F., Delles, C., *Hypertension* 2011, 57, 561–569.
- [22] Benavente, F., Medina-Casanellas, S., Barbosa, J., Sanz-Nebot, V., *J. Sep. Sci.* 2010, 33, 1294–1304.
- [23] Ramautar, R., Somsen, G. W., de Jong, G. J., *Electrophoresis* 2015, in press.
- [24] Medina-Casanellas, S., Tak, Y. H., Benavente, F., Sanz-Nebot, V., Sastre Toraño, J., Somsen, G. W., de Jong, G. J., *Electrophoresis* 2014, 35, 2996–3002.
- [25] Guzman, N. A., Blanc, T., Phillips, T. M., *Electrophoresis* 2008, 29, 3259–3278.
- [26] Guzman, N. A., Phillips, T. M., *Electrophoresis* 2011, 32, 1565–1578.
- [27] Breadmore, M. C., Tubaon, R. M., Shallen, A. I., Phung, S. C., Keyon, A. S., Gstoettenmayr, D., Prapatpong, P., Alhusban, A. A., Ranjbar, L., See, H. H., Dawod, M., Quirino, J. P., *Electrophoresis* 2015, 36, 36–61.
- [28] Madsen, R., Lundstedt, T., Trygg, J., *Anal. Chim. Acta* 2010, 659, 23–33.
- [29] Kotłowska, A., *Drug Dev. Res.* 2014, 75, 283–290.

- [30] González-Domínguez, R., García, A., García-Barrera, T., Barbas, C., Gómez-Ariza, J. L., *Electrophoresis* 2014, **35**, 3321–3330.
- [31] Tseng, Y. J., Kuo, C. T., Wang, S. Y., Liao, H. W., Chen, G. Y., Ku, Y. L., Shao, W. C., Kuo, C. H., *Electrophoresis* 2013, **34**, 2918–2927.
- [32] Farrés, M., Piña, B., Tauler, R., *Metabolomics* 2015, **11**, 210–224.
- [33] Siano, G. G., Pérez, I. S., García, M. D., Galera, M. M., Goicoechea, H. C., *Talanta* 2011, **85**, 264–275.
- [34] Parastar, H., Jalali-Heravi, M., Sereshti, H., Mani-Varnosfaderani, A., *J. Chromatogr. A* 2012, **1251**, 176–187.
- [35] Ortiz-Villanueva, E., Jaumot, J., Benavente, F., Piña, B., Sanz-Nebot, V., Tauler, R., *Electrophoresis* 2015, in press.
- [36] Brito, V., Giral, A., Enriquez-barreto, L., Puigdelívol, M., Suelves, N., Zamora-moratalla, A., Ballesteros, J. J., Martín, E. D., Dominguez-iturza, N., Morales, M., Alberch, J., Ginés, S., *J. Clin. Invest.* 2014, **124**, 4411–4428.
- [37] Pont, L., Benavente, F., Barbosa, J., Sanz-Nebot, V., *J. Sep. Sci.* 2013, **36**, 3896–3902.
- [38] Heiger, D., *High Performance Capillary Electrophoresis. An Introduction*, Agilent Technologies Germany 2000, pp. 82–89.
- [39] Ortiz-Villanueva, E., Benavente, F., Giménez, E., Yilmaz, F., Sanz-Nebot, V., *Anal. Chim. Acta* 2014, **846**, 51–59.
- [40] Kessner, D., Chambers, M., Burke, R., Agus, D., Mallick, P., *Bioinformatics* 2008, **24**, 2534–2536.
- [41] Saurina, J., *Chemometr. Methods Capill. Electrophor.* 2009, 199–226.
- [42] Jaumot, J., Gargallo, R., de Juan, A., Tauler, R., *Chemometr. Intell. Lab.* 2005, **76**, 101–110.
- [43] Golub, G., Solna, K., Dooren, P. V., *SIAM J. Matrix Anal. Appl.* 2000, **22**, 1–19.
- [44] Windig, W., Guilment, J., *Anal. Chem.* 1991, **63**, 1425–1432.
- [45] Tauler, R., Smilde, A., Kowalski, B., *J. Chemometr.* 1995, **9**, 31–58.
- [46] de Juan, A., Jaumot, J., Tauler, R., *Anal. Methods* 2014, **6**, 4964–4976.
- [47] Wold, S., Sjöström, M., Eriksson, L., *Chemometr. Intell. Lab.* 2001, **58**, 109–130.
- [48] Palermo, G., Piraino, P., Zucht, H. D., *Adv. Appl. Bioinform. Chem.* 2009, **2**, 57–70.
- [49] Smith, C. A., Maille, G. O., Want, E. J., Qin, C., Trauger, S. A., Brandon, T. R., Custodio, D. E., Abagyan, R., Siuzdak, G., *Ther. Drug Monit.* 2005, **27**, 747–751.
- [50] Wishart, D. S., Knox, C., Guo, A. C., Eisner, R., Young, N., Gautam, B., Hau, D. D., Psychogios, N., Dong, E., Boua-tra, S., Mandal, R., Sinelnikov, I., Xia, J., Jia, L., Cruz, J. A., Lim, E., Sobsey, C. A., Shrivastava, S., Huang, P., Liu, P., Fang, L., Peng, J., Fradette, R., Cheng, D., Tzur, D., Clements, M., Lewis, A., De Souza, A., Zuniga, A., Dawe, M., Xiong, Y., Clive, D., Greiner, R., Nazzyrova, A., Shaykhutdinov, R., Li, L., Vogel, H. J., Forsythe, I., *Nucleic Acids Res.* 2009, **37**, 603–610.
- [51] Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., Tanabe, M., *Nucleic Acids Res.* 2012, **40**, 109–114.
- [52] Anglada-Huguet, M., Xifró, X., Giral, A., Zamora-Moratalla, A., Martín, E. D., Alberch, J., *Mol. Neurobiol.* 2014, **49**, 784–795.
- [53] Schug, K. A., Lindner, W., *J. Am. Soc. Mass Spectrom.* 2004, **15**, 840–847.
- [54] Wang, H., Lim, P. J., Karbowski, M., Monteiro, M. J., *Hum. Mol. Genet.* 2009, **18**, 737–752.
- [55] Shirendeb, U., Reddy, A. P., Manczak, M., Calkins, M. J., Mao, P., Tagle, D. A., Reddy, P. H., *Hum. Mol. Genet.* 2011, **20**, 1438–1455.
- [56] Gines, S., Seong, I. S., Fossale, E., Ivanova, E., Trettel, F., Gusella, J. F., Wheeler, V. C., Persichetti, F., MacDonald, M. E., *Hum. Mol. Genet.* 2003, **12**, 497–508.
- [57] Petersén, A., Gabery, S., *J. Hunt.* 2012, **1**, 5–16.
- [58] Bowles, K. R., Jones, L., *J. Huntingtons Dis.* 2014, **3**, 89–123.
- [59] Carrasco, E., Casper, D., Werner, P., *J. Neurosci. Res.* 2007, **85**, 3109–3117.
- [60] Gabery, S., Halliday, G., Kirik, D., Englund, E., Petersén, A., *Neuropathol. Appl. Neurobiol.* 2015, **41**, 843–848.
- [61] Denny, C. A., Desplats, P. A., Thomas, E. A., Seyfried, T. N., *J. Neurochem.* 2010, **115**, 748–758.
- [62] Maglione, V., Marchi, P., Di Pardo, A., Lingrell, S., Horkey, M., Tidmarsh, E., Sipione, S., *J. Neurosci.* 2010, **30**, 4072–4080.
- [63] Chang, K. L., New, L. S., Mal, M., Goh, C. W., Aw, C. C., Browne, E. R., Chan, E. C., *J. Proteome Res.* 2011, **10**, 2079–2087.
- [64] Kee, A. J., Smith, R. C., Gross, A. S., Madsen, D. C., Rowe, B., *Metabolism* 1994, **43**, 1373–1378.
- [65] Kardys, I., de Maat, M. P., Klaver, C. C., Despriet, D. D., Uitterlinden, A. G., Hofman, A., de Jong, P. T., Witteman, J. C., *Am. J. Cardiol.* 2007, **100**, 646–648.
- [66] Alí-Torres, J., Rodríguez-Santiago, L., Sodupe, M., Rauk, A., *J. Phys. Chem. A* 2011, **115**, 12523–12530.

Capítulo 5. Resultados y Discusión

5.1. Análisis dirigido de biomarcadores proteicos. La transtiretina en la polineuropatía amiloidótica familiar tipo I

Este capítulo se centra en el estudio de la transtiretina (TTR), una proteína homotetramérica compuesta por cuatro subunidades enlazadas mediante interacciones no covalentes y responsable del transporte de tiroxina y retinol en sangre. Algunas modificaciones en la secuencia aminoacídica de la TTR (isoformas) están asociadas a variantes genéticas causantes de diferentes tipos de polineuropatías amiloidóticas familiares (FAP), las cuales se caracterizan por la formación de agregados de tipo amiloide en ciertos tejidos. Esta tesis en concreto se centra en el estudio de la TTR(Met30), que implica la sustitución de una valina por una metionina en la posición 30 del monómero de la TTR, como biomarcadora de la polineuropatía amiloidótica familiar tipo I (FAP-I). Este capítulo trata la separación, identificación y caracterización de las diferentes proteoformas de la TTR, las cuales juegan un papel importante en el diagnóstico de la enfermedad y podrían ser relevantes en el estudio de los mecanismos involucrados en su agregación.

La TTR es una proteína que se encuentra en el suero y en el líquido cefalorraquídeo (CSF). En suero, sus concentraciones varían entre individuos y oscilan entre 200 y 400 $\mu\text{g/mL}$. Actualmente, el análisis de TTR en muestras de suero está basado en la utilización de técnicas de inmunoprecipitación (IP) y espectrometría de masas (MS) con ionización por desorción mediante láser asistido por matriz, MALDI, y electrospray, ESI. En esta tesis doctoral se han optimizado diferentes métodos de IP *off-line* y *on-line* (en condiciones desnaturalizantes y no desnaturalizantes) para aislar la TTR (el monómero o sus oligómeros) del resto de componentes del suero, y se han desarrollado y validado nuevas metodologías analíticas para la separación, detección y caracterización de las isoformas y modificaciones post-traduccionales (PTMs) de la TTR mediante electroforesis capilar, cromatografía de líquidos capilar y espectrometría de

movilidad iónica, todas ellas acopladas *on-line* a la espectrometría de masas (CE-MS, CapLC-MS e IM-MS, respectivamente).

5.1.1. Análisis de patrones de TTR por CE-MS y CapLC-MS

En el primer trabajo de esta tesis doctoral (**artículo 2.1**), se evaluó la capacidad de la CE-MS para analizar la TTR con dos electrolitos de separación (BGEs): ácido acético (HAc) 1,0 M (pH 2,30) y acetato de amonio (NH₄Ac) 10 mM (pH 7,00). En estudios anteriores de nuestro grupo de investigación con la enzima homodimérica superóxido dismutasa I (SOD-I) [234], se demostró que el BGE ácido proporcionaba una mayor sensibilidad que el BGE neutro, probablemente debido a su mayor volatilidad y a la protonación de los grupos básicos de los aminoácidos ionizables presentes en la proteína. Sin embargo, la estructura dimérica nativa de la SOD-I se veía alterada como consecuencia del BGE ácido utilizado para la separación por CE, permitiendo únicamente el estudio de su estructura monomérica. A continuación se presenta lo que ocurrió con la TTR.

En la **Figura 5.1** se muestran los electroferogramas totales de iones (TIEs) (i), los espectros de masas (ii) y los espectros de masas deconvolucionados (iii) obtenidos por CE-MS para un patrón de TTR de 1000 µg/mL usando (A) HAc 1,0 M y (B) NH₄Ac 10 mM como BGEs. Tal y como se puede observar en esta figura, la utilización del BGE ácido, a pesar de la disminución del flujo electroosmótico (EOF), permitió detectar la TTR a un tiempo de migración inferior (6-8 min vs. 12-18 min para los BGEs ácido y neutro, respectivamente), probablemente debido a su mayor movilidad electroforética hacia el cátodo a pH ácido (punto isoelectrico (pI) de la TTR=5,4). Un análisis exhaustivo del clúster de iones multicargados obtenidos en el espectro de masas de ambos picos electroforéticos, demostró la presencia de monómero (MO) con el BGE ácido (los iones moleculares más abundantes corresponden a las cargas +16, +15 y +14), y la presencia de MO, y de una pequeña cantidad de dímero (DI), con el BGE neutro. Tal y como se puede observar en el espectro de masas de la **Figura 5.1 B-ii** y en la **Tabla 5.1**, los iones

moleculares con m/z nominales de 2314 y 2777 corresponderían a la suma de la contribución del MO (cargas +6 y +5, respectivamente) y del DI (cargas +12 y +10, respectivamente), mientras que el ión molecular con m/z nominal de 2525 se detectaría exclusivamente para el DI (carga +11).

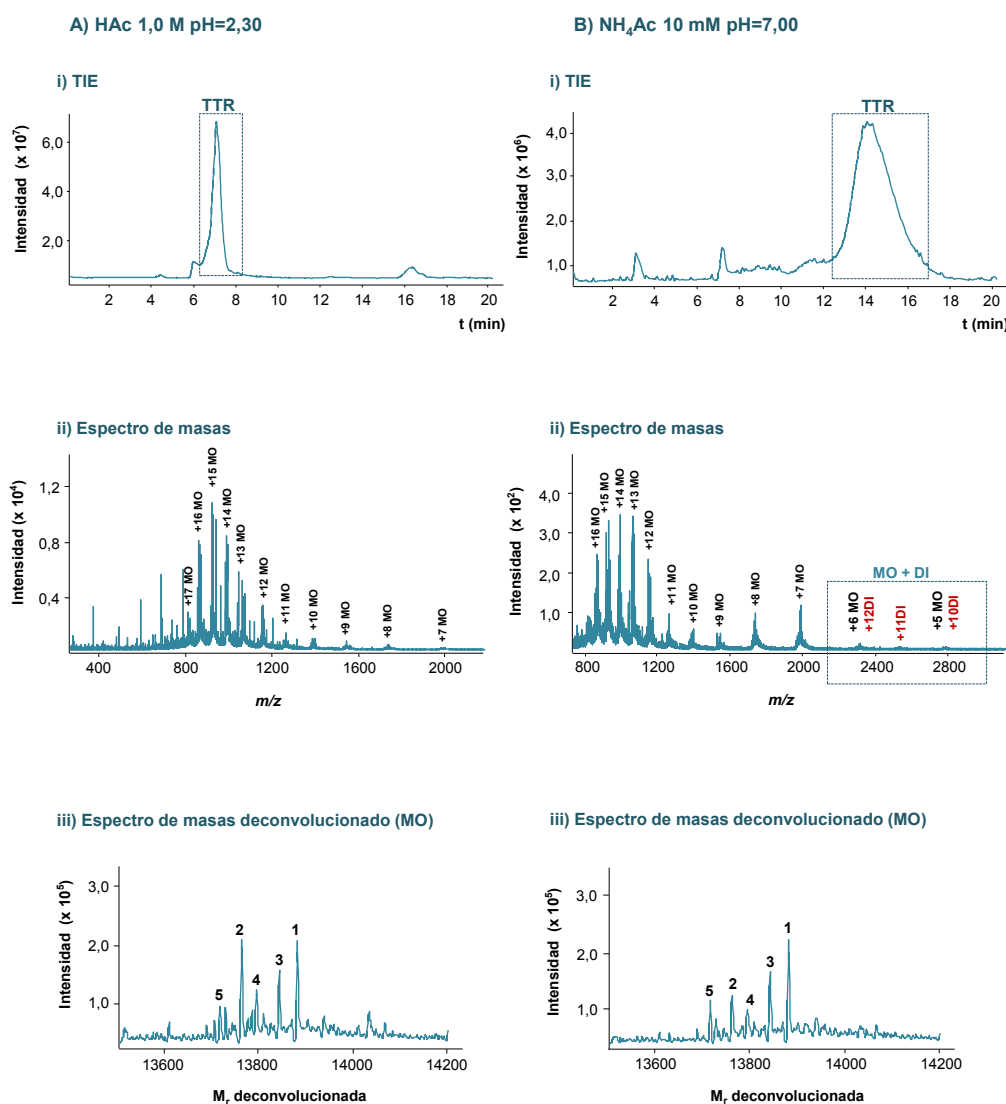


Figura 5.1. TIEs (i), espectros de masas (ii) y espectros de masas deconvolucionados (iii) obtenidos por CE-MS para un patrón de TTR de 1000 µg/mL usando BGEs de (A) HAc 1,0 M (pH=2,30) y (B) NH₄Ac 10 mM (pH=7,00). Las proteoformas del monómero (MO) en los espectros de masas deconvolucionados corresponden a (1): TTR-Cys, (2): TTR-Libre, (3): TTR-Fosforilada o TTR-Sulfonada, (4): TTR-Dihidroxilada o TTR-Sulfínico, (5): (10) C-G.

Tabla 5.1. Solapamientos de m/z nominal (teórica) entre los iones moleculares del monómero (MO), dímero (DI), trímero (TRI) y tetramero (TE) de la TTR en la región comprendida entre 1730 y 3100 m/z (zona en la que se suelen observar los iones moleculares de la TTR por CE-MS) para la TTR-Cys (proteoforma más abundante). Rojo: solapamientos MO-DI-TRI-TE. Verde: solapamientos DI-TE.

MO		DI		TRI		TE	
Ión molecular	m/z teórica	Ión molecular	m/z teórica	Ión molecular	m/z teórica	Ión molecular	m/z teórica
+8	1736	+16	1736	+24	1736	+32	1736
-		-		-		+31	1792
				+23	1811	-	
		+15	1852	-		+30	1852
		-		+22	1894	-	
+7	1984	+14	1984	-		+29	1916
-		-		+21	1984	+28	1984
				-		+27	2057
		+13	2136	+20	2083	-	
		-		-		+26	2136
+6	2314	+12	2314	+19	2193	-	
-		-		-		+25	2222
		+11	2525	+18	2314	+24	2314
		-		-		+23	2415
				+17	2450	-	
+5	2777	+10	2777	-		+22	2525
-		-		+16	2604	-	
				-		+21	2645
		+9	3085	+15	2777	+20	2777
		-		-		+19	2923
				+14	2975	-	
				-		+18	3085

Desafortunadamente, con el BGE neutro, no se pudieron detectar los iones moleculares correspondientes a la estructura tetramérica de la TTR, probablemente debido a la utilización en CE-MS de un líquido auxiliar coaxial ácido (*sheath liquid*) (0,05% (v/v) y 0,25% (v/v) de ácido fórmico (HFor) en el caso de utilizar el BGE ácido o el neutro, respectivamente) y al elevado vacío presente en el primer tramo recorrido por los iones en el interior del espectrómetro de masas que, en el modelo de instrumento empleado, no se podía manipular. En general, el uso de un BGE neutro es recomendable cuando se pretende trabajar con proteínas lábiles en condiciones menos desnaturalizantes. Sin embargo, en el caso de la TTR, debido a la mayor sensibilidad obtenida a pH ácido (comparar la magnitud de la señal en los ejes de ordenadas, tanto para los TIEs como para los espectros de masas), y al no ser posible la detección de los iones moleculares correspondientes a la proteína nativa en forma de tetrámero (TE), se utilizó el BGE de HAc para el resto de estudios. En estas condiciones, se detectaron hasta cinco proteoformas de la TTR: la TTR que forma un puente disulfuro con una cisteína libre en la posición 10 de la secuencia monomérica (TTR-Cys), la TTR libre (TTR-Libre), la TTR fosforilada o sulfonada (TTR-Fosforilada o TTR-Sulfonada), la TTR dihidroxilada o la TTR conjugada con el ácido sulfinico (TTR-Dihidroxilada o TTR-Sulfinico,) y la isoforma causada por la sustitución de una cisteína por una glicina en la posición 10 de la secuencia monomérica ((10) C-G). La exactitud y resolución del espectrómetro de masas no permitieron diferenciar entre la TTR-Fosforilada y la TTR-Sulfonada. Tampoco fue posible distinguir entre la TTR-Dihidroxilada y la TTR-Sulfinico, dos proteoformas con la misma masa molecular relativa (M_r).

En el segundo trabajo de esta tesis doctoral (**artículo 2.2**), se compararon los resultados de CE-MS y los obtenidos mediante cromatografía de líquidos capilar de fase reversa (C_{18}) acoplada a la espectrometría de masas (CapLC-MS), con el fin de evaluar la capacidad de ambas técnicas para proporcionar información relevante sobre las proteoformas de la TTR. Para ello, en CapLC-MS se utilizaron fases móviles de acetonitrilo (ACN):agua con un 0,1% (v/v) de HFor.

En la **Figura 5.2** se muestran los electroferogramas/cromatogramas totales de iones (TIEs/TICs) (i), los espectros de masas (ii) y los espectros de masas deconvolucionados (iii) obtenidos por (A) CE-MS (con el BGE ácido) y (B) CapLC-MS para un patrón de TTR de 1000 µg/mL. Tal y como se puede observar en esta figura, ambas técnicas proporcionaron una sensibilidad similar (comparar los ejes de ordenadas de los TIEs/TICs y de los espectros de masas). Sin embargo, la CE permitió detectar la TTR a un tiempo de migración inferior (7-9 min vs. 9-14 min para CE-MS y CapLC-MS, respectivamente).

De nuevo, un análisis exhaustivo del clúster de iones multicargados obtenidos en el espectro de masas de ambos picos electroforéticos, demostró la presencia de MO con CE-MS (los iones moleculares más abundantes corresponden a las cargas +16, +15 y +14), y la presencia de MO, y de una pequeña cantidad de DI, con CapLC-MS. En este caso, tal y como se puede observar en el espectro de masas de la **Figura 5.2 B-ii** y en la **Tabla 5.1**, los iones moleculares con m/z nominales de 1736, 1984, 2314 y 2777 corresponderían a la suma de la contribución del MO (cargas +8, +7, +6 y +5, respectivamente) y del DI (cargas +16, +14, +12 y +10, respectivamente), mientras que los iones moleculares con m/z nominal de 1852, 2136 y 2525 se detectarían exclusivamente para el DI (cargas +15, +13 y +11, respectivamente). La detección de esta pequeña cantidad de DI por CapLC-MS podría ser debida a la menor concentración de ácido en la fase móvil (0,1% (v/v) de HFor) si comparamos con el BGE utilizado en CE-MS (HAc 1,0 M, que equivale aproximadamente a un 6% (v/v)).

Es conveniente destacar que en este segundo trabajo (**artículo 2.2**) se consiguieron detectar más proteoformas monoméricas de la TTR que en el primer trabajo (**artículo 2.1**) (TTR-CysGly, TTR-Glutatión y TTR-CysGlu, **Figura 5.2 A-iii**) bajo las mismas condiciones experimentales por CE-MS (1000 µg/mL de concentración de TTR y HAc 1,0 M como BGE), probablemente debido a una mejora en las condiciones del espectrómetro de masas después de la sustitución de diversas piezas, entre ellas, el detector.

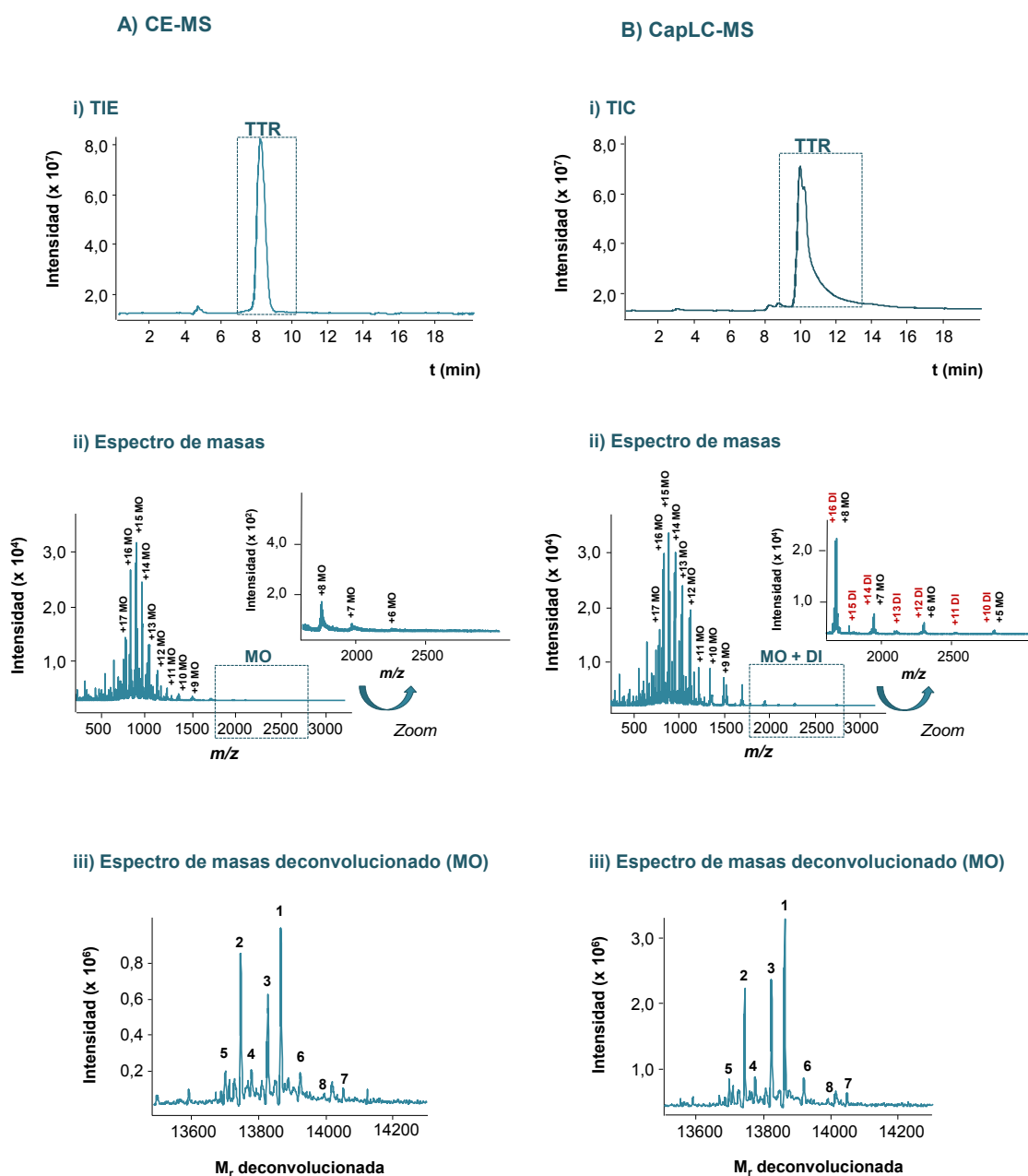


Figura 5.2. TIEs/TICs (i), espectros de masas (ii) y espectros de masas deconvolucionados (iii) para un patrón de TTR de 1000 µg/mL usando (A) CE-MS y (B) CapLC-MS. Las proteoformas del monómero (MO) en los espectros de masas deconvolucionados corresponden a (1): TTR-Cys, (2): TTR-Libre, (3): TTR-Fosforilada o TTR-Sulfonada, (4): TTR-Dihidroxilada o TTR-Sulfínico, (5): (10) C-G, (6): TTR-CysGly, (7): TTR-Glutatión, (8): TTR-CysGlu. BGE CE-MS: HAc 1,0 M (pH=2,30). Fase móvil CapLC-MS: ACN:agua con un 0,1% (v/v) de HFor.

La **Tabla 5.2** muestra las masas moleculares relativas (M_r) teóricas y experimentales, los errores de masas (E_r , en ppm) y las abundancias relativas (%A) para las proteoformas monoméricas de la TTR (1000 $\mu\text{g/mL}$) detectadas por (A) CE-MS y (B) CapLC-MS (BGE para CE-MS: HAc 1,0 M (pH=2,30). Fase móvil para CapLC-MS: ACN:agua con un 0,1% (v/v) de HFor). Ambas técnicas permitieron detectar las mismas isoformas y PTMs de la TTR con E_r y %A similares (por ejemplo, para la TTR-Libre, se calcularon unas %A del 91% y 88% por CE-MS y CapLC-MS, respectivamente).

Tabla 5.2. Masas moleculares relativas (M_r) teóricas y experimentales, errores de masas (E_r , en ppm) y abundancias relativas (%A) para las proteoformas del monómero (MO) de la TTR (1000 $\mu\text{g/mL}$) detectadas por (A) CE-MS y (B) CapLC-MS. Las %A se calcularon normalizando respecto a la proteoforma más abundante (TTR-Cys).

Patrón de TTR			(A) CE-MS			(B) CapLC-MS		
N	Proteoformas MO	M_r teórica	M_r experimental	E_r^a (ppm)	%A	M_r experimental	E_r^a (ppm)	%A
1	TTR-Cys	13880,4022	13880,9300	38	100	13880,7500	25	100
2	TTR-Libre	13761,2640	13761,7600	36	91	13761,4500	14	88
3	TTR-Fosforilada	13841,2439	13841,6400	29	67	13841,5700	24	67
	TTR-Sulfonada	13841,3283		23			17	
4	TTR-Dihidroxilada TTR-Sulfinico	13793,2628	13793,5000	17	27	13793,9300	48	25
5	(10) C-G	13715,1713	13715,6400	34	24	13715,3300	12	21
6	TTR-CysGly	13937,4590	13937,5000	3	11	13937,8900	31	12
7	TTR-Glutatión	14066,5732	14066,9600	27	10	14066,6700	7	10
8	TTR-CysGlu	14009,5218	14009,3700	11	9	14009,3500	12	7

^aEl E_r (en ppm) se calcula en valor absoluto como: $(M_r \text{ experimental} - M_r \text{ teórica}) / M_r \text{ teórica} * 10^6$.

Tal y como se observa en esta tabla, la principal PTM detectada correspondió a la TTR-Cys (M_r teórica=13880,4022). El resto de proteoformas correspondieron (en orden de %A) a la TTR-Libre (M_r teórica=13761,2640), la TTR-Fosforilada o TTR-Sulfonada (M_r teórica=13841,2439 y M_r teórica=13841,3283, respectivamente), la TTR-Dihidroxilada o TTR-Sulfinico (M_r teórica=13793,2628), la isoforma (10) C-G (M_r teórica=13715,1713) y las PTMs originadas cuando la cisteína de la posición 10 forma un puente disulfuro con el péptido cisteinil glicina (TTR-CysGly, M_r teórica=13937,4590), el péptido glutatión (TTR-Glutatión, M_r teórica=14066,5732) y el péptido cisteinil ácido glutámico (TTR-CysGlu, M_r teórica=14009,5218). En esta tabla es posible apreciar como la exactitud y resolución del espectrómetro de masas no fueron suficientes para diferenciar entre la TTR-Fosforilada y la TTR-Sulfonada (29 ppm y 23 ppm de E_r por CE-MS, y 24 ppm y 17 ppm de E_r por CapLC-MS, respectivamente). Además, tal y como se apuntó anteriormente, las proteoformas TTR-Dihidroxilada y la TTR-Sulfinico tienen la misma M_r , por lo que tampoco fue posible distinguirlas.

Por lo que hace referencia a los parámetros de calidad de ambos métodos, la **Tabla 5.3** muestra los valores de repetitividad (%RSD, $n=10$) para las áreas de pico (A_p) y los tiempos de migración/ retención (t_m/t_r), los intervalos de linealidad y los límites de detección (LODs) para cada una de las proteoformas monoméricas de la TTR detectadas por (A) CE-MS y (B) CapLC-MS. Tal y como se puede observar en esta tabla, ambas técnicas proporcionaron una repetitividad adecuada para las A_p (entre 0,2% y 7,1% para CE-MS, y entre 0,6% y 2,9% para CapLC-MS) y los t_m/t_r (entre 0,1% y 0,9% para CE-MS, y entre 0,1% y 0,3% para CapLC-MS). En CE-MS el método era lineal entre 50 y 1000 $\mu\text{g/mL}$ (concentración máxima a la que se ha trabajado) para todas las proteoformas, mientras que por CapLC-MS se observó un comportamiento lineal entre 100 y 1000 $\mu\text{g/mL}$ (con excepción de las PTMs TTR-CysGly, TTR-Glutatión y TTR-CysGlu, para las cuales se observó entre 250 y 1000 $\mu\text{g/mL}$). Finalmente, es importante destacar que, pese a la capacidad de ambas técnicas para proporcionar información relevante sobre las diferentes isoformas y PTMs de la TTR, la CE-

MS permitió detectar las diferentes proteoformas monoméricas de la TTR con LODs entre 2 y 2,5 veces más bajos que los obtenidos por CapLC-MS, aunque esta última técnica proporcionó cierta información sobre el DI en el caso de analizar patrones de TTR.

Tabla 5.3. Repetitividad (%RSD, $n=10$) para las áreas de pico (A_p) y los tiempos de migración/retención (t_m/t_r) (patrón de TTR de 250 $\mu\text{g/mL}$), intervalos de linealidad y límites de detección (LODs) para las proteoformas del monómero (MO) de la TTR detectadas por (A) CE-MS y (B) CapLC-MS. (S/N=relación señal ruido).

Patrón de TTR		(A) CE-MS					(B) CapLC-MS				
N	Proteoformas MO	A_p	t_m		Linealidad 50-1000 $\mu\text{g/mL}$ $r^2 > 0,998$	LODs ($\mu\text{g/mL}$)	A_p	t_r		Linealidad ^a 100-1000 $\mu\text{g/mL}$ $r^2 > 0,998$	LODs ($\mu\text{g/mL}$)
		%RSD ($n=10$)	Media ($n=10$)	%RSD ($n=10$)			%RSD ($n=10$)	Media ($n=10$)	%RSD ($n=10$)		
1	TTR-Cys	0,2	7,9	0,5	$A=1 \times 10^4 C - 2 \times 10^4$	10 (S/N=457)	1,8	10,3	0,1	$A=2 \times 10^4 C + 9 \times 10^4$	25 (S/N=46)
2	TTR-Libre	1,7	8,0	0,2	$A=1 \times 10^4 C - 2 \times 10^4$	10 (S/N=430)	2,0	10,1	0,1	$A=1 \times 10^4 C + 6 \times 10^4$	25 (S/N=31)
3	TTR-Fosforilada TTR-Sulfonada	1,0	8,1	0,3	$A=7 \times 10^3 C - 2 \times 10^4$	10 (S/N=301)	1,7	10,3	0,2	$A=1 \times 10^4 C + 4 \times 10^4$	25 (S/N=14)
4	TTR-Dihidroxilada TTR-Sulfínico	2,5	8,1	0,3	$A=3 \times 10^3 C - 1 \times 10^4$	25 (S/N=264)	2,2	10,2	0,3	$A=5 \times 10^3 C + 2 \times 10^4$	50 (S/N=18)
5	(10) C-G	0,6	7,9	0,6	$A=2 \times 10^3 C - 4 \times 10^4$	25 (S/N=181)	2,6	10,0	0,1	$A=4 \times 10^3 C + 2 \times 10^4$	50 (S/N=13)
6	TTR-CysGly	4,7	7,9	0,9	$A=2 \times 10^3 C - 8 \times 10^4$	100 (S/N=780)	2,2	10,0	0,1	$A=4 \times 10^3 C + 3 \times 10^4$	250 (S/N=220)
7	TTR-Glutatión	1,5	8,0	0,1	$A=1 \times 10^3 C - 2 \times 10^4$	100 (S/N=758)	0,6	10,2	0,2	$A=2 \times 10^3 C + 1 \times 10^4$	250 (S/N=184)
8	TTR-CysGlu	7,1	8,0	0,1	$A=9 \times 10^2 C - 2 \times 10^4$	100 (S/N=447)	2,9	10,2	0,3	$A=1 \times 10^3 C + 8 \times 10^4$	250 (S/N=118)

^aLa linealidad en CapLC-MS para las proteoformas TTR-CysGly, TTR-Glutatión y TTR-CysGlu se observó entre 250 y 1000 $\mu\text{g/mL}$.

5.1.2. Análisis de TTR en muestras de suero humano por CE-MS y CapLC-MS

Una vez optimizados y validados los métodos de CE-MS y CapLC-MS con patrones, se procedió a analizar la TTR en diferentes muestras de suero humano proporcionadas por el *Hospital Universitari de Bellvitge* (HUB, Hospitalet de Llobregat, España):

-Tres controles sanos.

-Tres pacientes de FAP-I en diferentes estadios de evolución:

- Asintomático: sin síntomas de FAP-I.
- Sintomático: con síntomas de FAP-I durante 4 años.
- Yatrogénico: paciente originalmente sin la mutación al que le trasplantaron el hígado de un paciente con FAP-I. Con síntomas de FAP-I durante 6 años.

-Un paciente de FAP-I trasplantado de hígado, estrategia terapéutica más prometedora actualmente para prevenir la enfermedad. Sin síntomas de FAP-I.

Para ello, fue necesario optimizar diferentes métodos de purificación de la TTR por inmunoprecipitación (IP) *off-line* (IP convencional en solución e IP con partículas magnéticas) para aislar la proteína del resto de componentes del suero previo al análisis por CE-MS y CapLC-MS. A continuación, se discuten en detalle los resultados obtenidos con los diferentes métodos de IP empleados.

5.1.2.1. IP convencional en solución

La IP convencional en solución consiste en incubar una solución del anticuerpo, específico contra un determinado antígeno, con la muestra que contiene la proteína de interés. Tras la incubación, se centrifuga la muestra y se elimina el sobrenadante, permitiendo de esta manera la recuperación del complejo antígeno-anticuerpo. Para los estudios de esta tesis, se utilizó un anticuerpo policlonal específico contra la TTR con el objetivo de aislar esta proteína del resto de

componentes del suero. Es importante destacar que, en estudios preliminares no recopilados en los artículos de esta tesis, la optimización de la IP convencional en solución se intentó llevar a cabo con patrones de TTR (disueltos en agua o en tampón fosfato salino (PBS)). Sin embargo, no se consiguieron recuperaciones adecuadas de la TTR, mientras que en presencia de la matriz compleja del suero, la IP proporcionó mejores resultados. Se evaluaron diferentes proporciones de suero:solución de anticuerpo, desde 50:20 hasta 50:100 μL , observándose una precipitación cuantitativa del complejo a relaciones de volumen igual o superiores a 50:20 μL . Primero, se analizó por CE-MS la mezcla de TTR y anticuerpo obtenida después de disociar el complejo antígeno-anticuerpo con HAc 1,0 M (pH 2,30), seguido de una etapa de ultrafiltración con filtros de corte molecular de M_r 10000 para eliminar sales y otros compuestos. Dichos filtros se pasivaron con una solución acuosa al 5% (v/v) de polietilenglicol (PEG, M_r 8000) con el objetivo de bloquear sus principales sitios activos, evitando de esta manera la adsorción de la proteína en las paredes (el efecto de la pasivación con PEG se discute en detalle en la sección 5.3.1 de esta tesis doctoral). En condiciones óptimas, tal y como se puede observar en la **Figura 5.3** para una muestra de suero control, los tres picos obtenidos por CE-MS correspondieron al gel polimérico utilizado en los tubos de recolección de sangre como activador de la coagulación (i), al anticuerpo policlonal contra la TTR (ii) y a la TTR monomérica (iii).

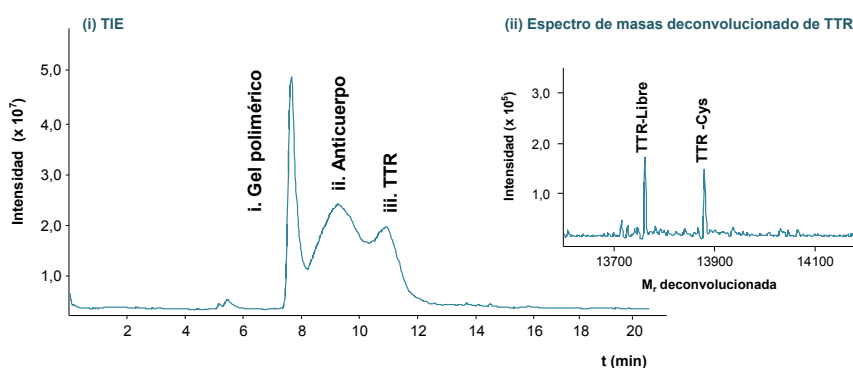


Figura 5.3. TIE (i) y espectro de masas deconvolucionado de la TTR (ii) obtenidos por CE-MS para una muestra de suero control después de llevar a cabo la IP y la ultrafiltración con filtros de corte molecular de M_r 10000 pasivados con PEG. BGE: HAc 1,0 M (pH=2,30).

Desafortunadamente, la presencia del anticuerpo y del gel polimérico producía caídas de corriente al inicio de la separación por CE. Por este motivo, fue necesario llevar a cabo una etapa de ultrafiltración con filtros de corte molecular de M_r 100000 con el fin de separar la TTR de éstos, que presentan unas M_r mucho mayores. Para ello, primero se evaluó por CE-MS la recuperación de la proteína en los filtrados, filtrando un patrón de TTR de 250 $\mu\text{g/mL}$ (similar a la concentración típica de TTR en el suero) con filtros de corte molecular de M_r 50000 y 100000 sin pasivar y pasivados con un 1% (m/v) de albúmina de suero bovino (BSA) en PBS y con una solución acuosa al 5% (v/v) de PEG. En el caso de los filtros de M_r 50000 sin pasivar, pasivados con PEG y pasivados con BSA, se obtuvieron recuperaciones de TTR (en los filtrados) del 43% ($n=3$, %RSD=5), 55% ($n=3$, %RSD=1) y 61% ($n=3$, %RSD=4), respectivamente. En el caso de los filtros de M_r 100000, se obtuvieron valores de recuperación más elevados, probablemente debido a una menor retención de la TTR en los filtros de mayor tamaño de poro (55% ($n=3$, %RSD=2), 76% ($n=3$, %RSD=3) y 62% ($n=3$, %RSD=5), para los filtros sin pasivar, pasivados con PEG y pasivados con BSA, respectivamente). Los filtros de corte molecular de M_r 100000 pasivados con PEG fueron escogidos para el resto de estudios, ya que fueron los que proporcionaron una mayor recuperación de TTR en los filtrados (76%).

La **Figura 5.4** muestra los TIEs (i) y los espectros de masas deconvolucionados (ii) obtenidos por CE-MS para la TTR procedente de una muestra de suero control y de un paciente de FAP-I asintomático después de llevar a cabo la IP convencional en solución seguida de una etapa de ultrafiltración con filtros de corte molecular de M_r 100000 pasivados con PEG. Si se comparan los espectros de masas deconvolucionados de la TTR para un control sano y para un paciente de FAP-I asintomático, se puede observar que sólo fue posible detectar las proteoformas normales más abundantes de la TTR para el control sano (TTR-Cys y TTR-Libre). En el caso del paciente asintomático de FAP-I, se detectaron, además de éstas, las mismas proteoformas mutantes: la TTR(Met30)-Cys y la TTR(Met30)-Libre. Sin embargo, en ninguno de los casos se pudieron detectar todas las proteoformas observadas en el patrón de TTR (TTR-Fosforilada o TTR-Sulfonada, TTR-Dihidroxilada o TTR-Sulfinico, TTR-CysGly, TTR-Glutatión y TTR-CysGlu,

discutidas en la sección anterior), probablemente debido a la baja concentración de TTR en las muestras de suero tratadas y a las recuperaciones globales del tratamiento al que se las sometió. Aun así, el método desarrollado proporcionó una repetitividad (%RSD, $n=10$) adecuada para todas las proteoformas monoméricas detectadas. Así, para el paciente de FAP-I asintomático, se obtuvieron para las proteoformas normales y mutantes, valores de entre 1,0-1,4% y 7-11% para los t_m y las A_p , respectivamente. La reproducibilidad en este caso (%RSD, $n=3$ muestras independientes) también fue buena, con valores de entre 3,6% y 3,7% para los t_m , y de entre 3% y 11% para las A_p .

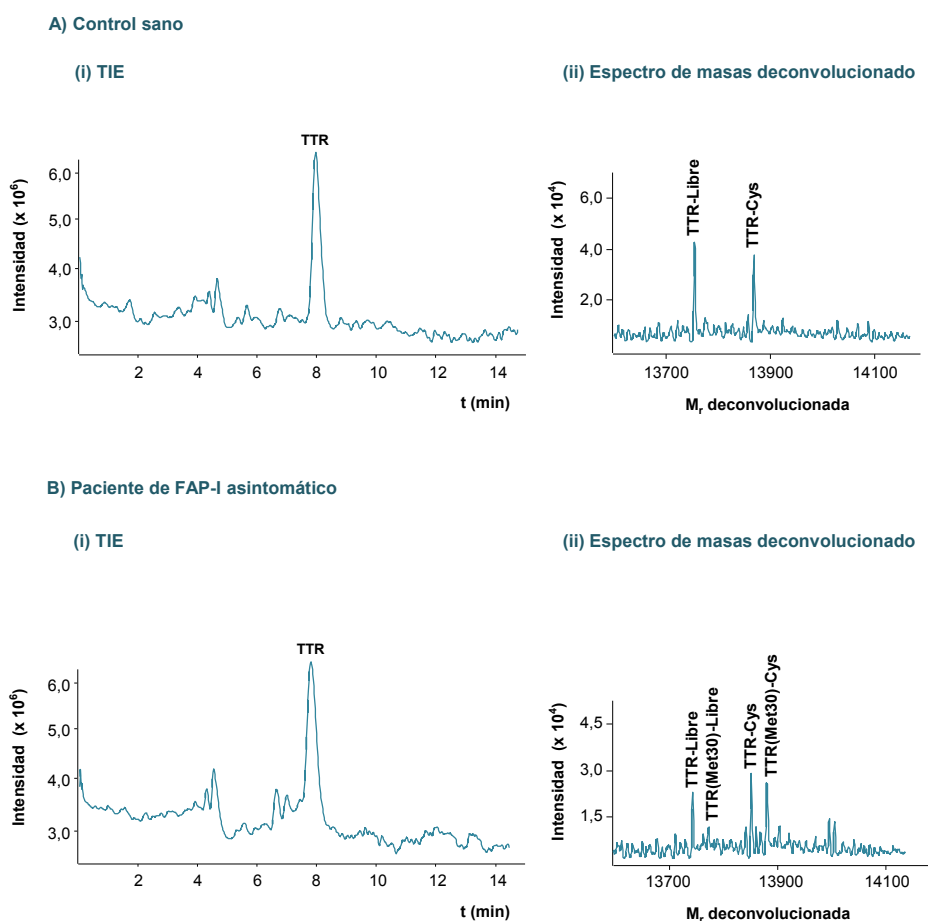


Figura 5.4. TIEs (i) y espectros de masas deconvolucionados (ii) de la TTR obtenidos por CE-MS para (A) una muestra de suero control y (B) una muestra de suero de un paciente de FAP-I asintomático después de llevar a cabo la IP y la ultrafiltración con filtros de corte molecular de M_r 100000 pasivados con PEG. BGE: HAc 1,0 M (pH=2,30).

Una vez optimizado el tratamiento de muestra para purificar la TTR del suero, se procedió a analizar por CE-MS todas las muestras de suero disponibles. El gráfico de la **Figura 5.5** muestra los valores medios ($n=3$) de las relaciones entre las A_p para las proteoformas mutantes y normales de la TTR (TTR(Met30)-Cys/TTR-Cys y TTR(Met30)-Libre/TTR-Libre), así como las relaciones entre las A_p para las proteoformas oxidadas y sin oxidar (TTR-Cys/TTR-Libre y TTR(Met30)-Cys/TTR(Met30)-Libre). A partir de este gráfico se pudieron extraer varias conclusiones:

- En el paciente de FAP-I trasplantado de hígado no se encontraron las proteoformas mutantes de la TTR(Met30), lo que confirmaría la efectividad de este tratamiento terapéutico.
- Las relaciones de A_p entre las proteoformas mutantes y normales de la TTR (TTR(Met30)-Cys/TTR-Cys y TTR(Met30)-Libre/TTR-Libre) eran similares en todos los pacientes (síntomáticos y asintomáticos), lo que indica que estas relaciones se mantendrían durante la progresión de la enfermedad.
- Las relaciones de A_p entre las proteoformas oxidadas y sin oxidar (TTR-Cys/TTR-Libre y, especialmente, TTR(Met30)-Cys/TTR(Met30)-Libre) podrían estar relacionadas con el inicio de los síntomas, ya que eran más bajas para el paciente asintomático, medias para el sintomático (con síntomas desde hace 4 años) y más altas para el paciente yatrogénico (con síntomas desde hace 6 años). Este hecho sugeriría una mayor cantidad de proteoformas oxidadas (TTR-Cys y, especialmente, TTR(Met30)-Cys) a medida que progresa la enfermedad.

Actualmente se cree que las PTMs asociadas a la cisteína de la posición 10 (Cys10), especialmente la TTR-Cys, juegan un papel importante en el inicio y proceso patológico de las diferentes amiloidosis ligadas a la TTR [67,74,235–237]. En esta tesis doctoral (**artículo 2.1**) se proponen las proteoformas oxidadas TTR-Cys y, especialmente, la TTR(Met30)-Cys, como biomarcadoras del inicio y progreso de la FAP-I. Sin embargo, se deberían analizar un mayor

número de muestras para poder confirmar la información proporcionada por el método propuesto, así como para profundizar en la relación entre las diferentes proteoformas y los mecanismos involucrados en la agregación de la TTR en FAP-I.

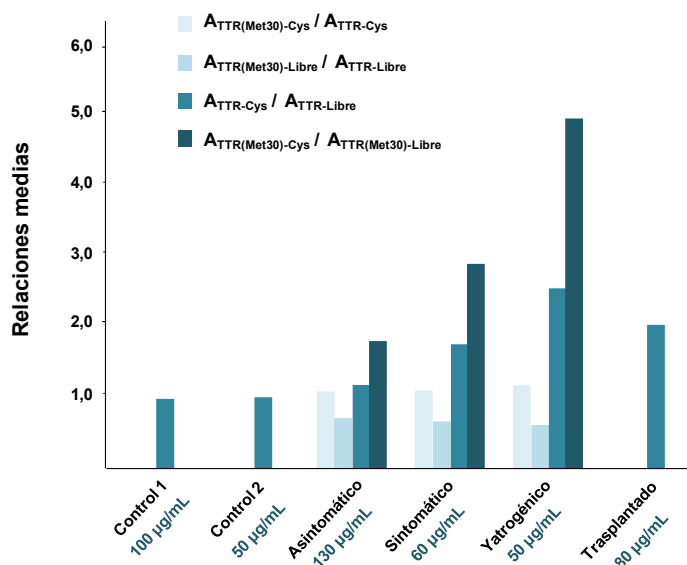


Figura 5.5. Relaciones medias ($n=3$) entre las A_p obtenidas por CE-MS para las proteoformas mutantes y normales de la TTR ($TTR(Met30)-Cys/TTR-Cys$ y $TTR(Met30)-Libre/TTR-Libre$) y entre las A_p para las proteoformas oxidadas y sin oxidar ($TTR-Cys/TTR-Libre$ y $TTR(Met30)-Cys/TTR(Met30)-Libre$). En el eje de las x se muestran también las concentraciones totales de la TTR cuantificadas en el **artículo 2.1** de esta tesis doctoral.

5.1.2.2. IP con partículas magnéticas

Actualmente, el uso de partículas magnéticas tiene un papel destacado en el desarrollo de metodologías de pretratamiento de muestra en diversos campos, hecho que ha llevado a una sustitución progresiva de los métodos de IP convencionales en solución. Las técnicas de IP con partículas magnéticas son ampliamente recomendadas para mejorar la reproducibilidad y sensibilidad de dichos métodos convencionales, así como para minimizar la unión no específica de otras proteínas. Tal y como se ha discutido anteriormente, en el **artículo 2.1** se propuso un

método *off-line* de IP convencional en solución para purificar la TTR del suero previo al análisis por CE-MS. Sin embargo, no se pudieron detectar todas las proteoformas observadas en los patrones de TTR (sólo se detectaron la TTR-Cys y la TTR-Libre, así como la TTR(Met30)-Cys y la TTR(Met30)-Libre, en el caso de pacientes de FAP-I), probablemente debido a la baja concentración de TTR en las muestras de suero tratadas y a las recuperaciones globales del tratamiento al que se las sometió.

En el segundo trabajo de esta tesis doctoral (**artículo 2.2**), se desarrolló un nuevo método *off-line* de IP con partículas magnéticas funcionalizadas con Proteína A (*Protein A Ultrarapid Agarose*TM, UAPA) para la purificación de TTR del suero previo al análisis por CE-MS y CapLC-MS. Es importante destacar que, en estudios preliminares, se evaluaron otros tipos de partículas magnéticas funcionalizadas con Proteína A con diferentes tamaños de partícula y afinidades por el anticuerpo (*Dynabeads*[®] *Protein A* (DyPA) y *SiMAG-Protein A* (SiPA)) [211]. Sin embargo, pese a la capacidad de éstas para aislar la TTR del resto de componentes del suero, las mejores recuperaciones se obtuvieron con las partículas UAPA, por lo que fueron seleccionadas para el resto de estudios. Con estas partículas magnéticas, la orientación del anticuerpo es óptima para la interacción con la TTR, ya que el anticuerpo se une a ellas a través de la región Fc. Sin embargo, dicha unión se produce mediante interacciones no covalentes, por lo que el anticuerpo se eluye junto a la TTR y puede ser detectado en el análisis por CE-MS y CapLC-MS.

Una vez optimizado el tratamiento de muestra con partículas magnéticas UAPA para la purificación de la TTR del suero, se procedió a analizar por CE-MS y CapLC-MS todas las muestras de suero disponibles. Cabe destacar que inmediatamente después de la IP con las partículas UAPA, fue necesario llevar a cabo una etapa de ultrafiltración con filtros de corte molecular de M_r 10000 pasivados con PEG, con el objetivo de cambiar el BGE de elución (que contenía glicina, que es poco volátil y, por tanto, poco recomendable para ESI-MS), por HAc 1,0 M antes de los análisis por CE-MS y CapLC-MS. En este caso, no fue necesaria la

ultrafiltración para separar el gel polimérico activador de la coagulación o el anticuerpo, tal y como sucedía anteriormente (**artículo 2.1**).

En la **Figura 5.6** se muestran los TIEs/TICs (i) y los espectros de masas deconvolucionados (ii) obtenidos por (A) CE-MS y (B) CapLC-MS para la TTR de una muestra de suero control y de un paciente de FAP-I sintomático después de llevar a cabo la IP con partículas magnéticas UAPA seguida de una etapa de ultrafiltración con filtros de corte molecular de M_r 10000 pasivados con PEG. Tal y como se puede observar en esta figura, los dos picos principales correspondieron a la TTR monomérica y al anticuerpo. Sin embargo, en el caso de la CE-MS, la TTR apareció a un t_m superior al del anticuerpo, mientras que por CapLC-MS, el anticuerpo quedó mucho más retenido en la fase estacionaria (C_{18}) y se eluyó con un porcentaje de fase móvil del 100% (v/v) de ACN. El espectro de masas deconvolucionado de la TTR para el suero control (**Figura 5.6**) muestra que las principales proteoformas monoméricas detectadas con ambas técnicas correspondieron (en orden de abundancia relativa) a la TTR-Cys, TTR-Libre, TTR-Fosforilada o TTR-Sulfonada, TTR-Dihidroxilada o TTR-Sulfinico, (10) C-G y TTR-CysGly. En este sentido, el método de IP con partículas magnéticas desarrollado en el segundo trabajo de esta tesis doctoral (**artículo 2.2**), permitió mejorar los LODs y detectar diferentes proteoformas de la TTR que se encuentran a baja concentración y que no habían podido ser estudiadas con el tratamiento anterior de IP convencional en solución (**artículo 2.1**). Sólo dos de las proteoformas detectadas en los patrones (TTR-Glutatión y TTR-CysGlu) no pudieron ser detectadas con esta metodología. En el caso del paciente sintomático de FAP-I, las principales proteoformas mutantes detectadas con ambas técnicas correspondieron a la TTR(Met30)-Cys y a la TTR(Met30)-Libre. Sin embargo, el E_r no permitió diferenciar entre la TTR(Met30)-Libre y la TTR-Dihidroxilada o TTR-Sulfinico (M_r similares), por lo que sólo la detección de la proteoforma TTR(Met30)-Cys permitiría confirmar de manera inequívoca la presencia de la TTR(Met30) en pacientes de FAP-I. Es importante destacar que este problema no existía con el método *off-line* de IP convencional en solución, ya que no era posible detectar las proteoformas TTR-Dihidroxilada o TTR-Sulfinico (sección 5.1.2.1). En condiciones óptimas, el método de IP

con partículas magnéticas proporcionó una reproducibilidad (%RSD, $n=3$ muestras independientes) adecuada para todas las proteoformas monoméricas detectadas usando ambas técnicas. Por ejemplo, en el caso del paciente sintomático de FAP-I, se obtuvieron unos valores de reproducibilidad para las A_p de entre 3,6-10,4% para CE-MS, y de entre 2,7-11,4% para CapLC-MS, y unos valores de reproducibilidad para los t_m/t_r de entre 0,1-1,0% para CE-MS, y de entre 0,2-0,7% para CapLC-MS.

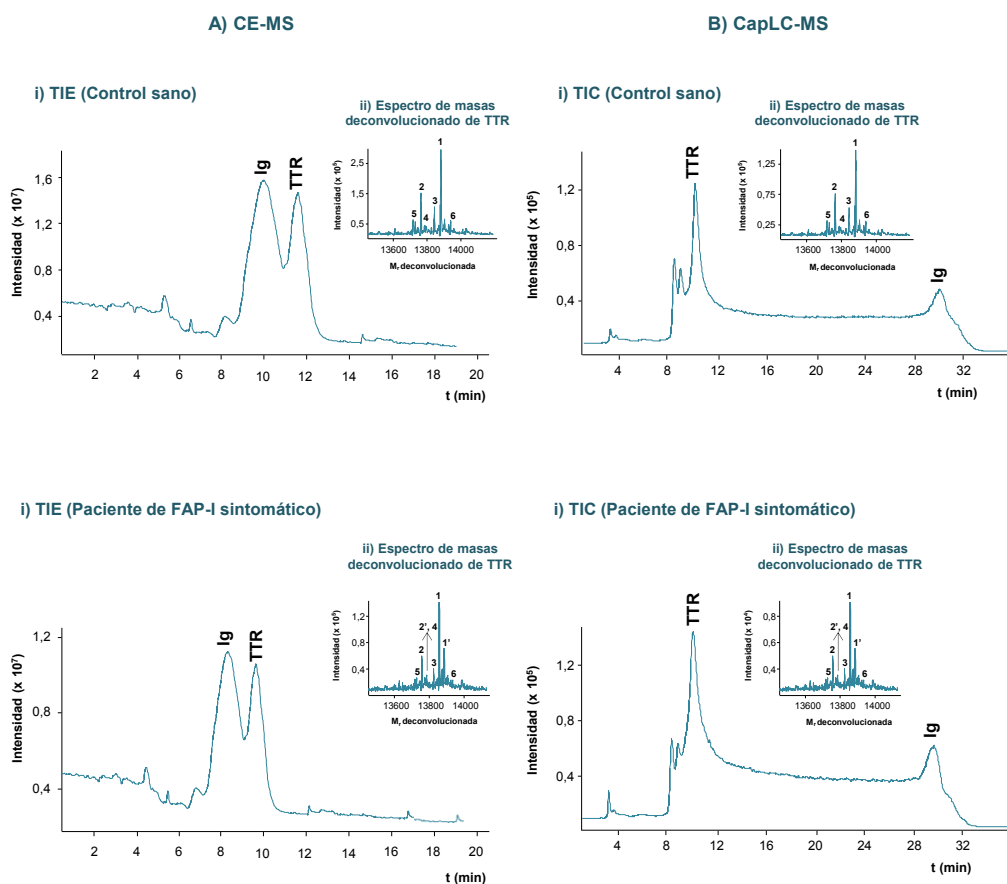


Figura 5.6. TIEs/TICs (i) y espectros de masas deconvolucionados de la TTR (ii) obtenidos por (A) CE-MS y (B) CapLC-MS para una muestra de suero control y de un paciente de FAP-I sintomático después de llevar a cabo la IP con partículas magnéticas UAPA seguida de una etapa de ultrafiltración con filtros de corte molecular de M_r 10000 pasivados con PEG. Las proteoformas del monómero (MO) corresponden a (1): TTR-Cys, (1'): TTR(Met30)-Cys, (2): TTR-Libre, (2'): TTR(Met30)-Libre, (3): TTR-Fosforilada o TTR-Sulfonada, (4): TTR-Dihidroxilada o TTR-Sulfinico, (5): (10) C-G, (6): TTR-CysGly. BGE CE-MS: HAc 1,0 M (pH=2,30). Fase móvil CapLC-MS: ACN:agua con un 0,1% (v/v) de HFor.

En el **artículo 2.1** se mostró, entre otras cosas, que las relaciones entre las proteoformas oxidadas y sin oxidar de la TTR podrían estar relacionadas con el inicio de los síntomas de la FAP-I, sugiriendo una mayor cantidad de proteoformas oxidadas a medida que progresa la enfermedad. Con el objetivo de estudiar la relación entre las diferentes proteoformas de la TTR y la FAP-I, la **Figura 5.7 A y B-i** muestra las relaciones medias ($n=3$) entre las A_p para la TTR-Cys/TTR-Libre y la TTR(Met30)-Cys/TTR(Met30)-Libre para todas las muestras de suero de controles sanos y de pacientes de FAP-I usando (A) CE-MS y (B) CapLC-MS. Tal y como se puede observar en este gráfico, ambas técnicas permitieron obtener la misma información, poniendo de manifiesto que, en pacientes de FAP-I, la proteoforma TTR(Met30)-Cys es mucho más abundante que la TTR(Met30)-Libre, tal y como se observó en el artículo descrito anteriormente (**artículo 2.1**). Respecto al resto de PTMs asociadas a la Cys10 (TTR-Sulfonada, TTR-Sulfinico y TTR-CysGly, **Figura 5.7 A y B-ii**), sólo la proteoforma TTR-CysGly incrementó ligeramente su abundancia respecto a la TTR-Libre en pacientes de FAP-I, confirmando, de esta manera, la importancia de las proteoformas oxidadas de la TTR (especialmente la TTR(Met30)-Cys) como biomarcadoras del estrés oxidativo implicado en la formación de agregados amiloides en FAP-I.

A pesar de que la IP con partículas magnéticas permite obtener más información sobre las diferentes proteoformas de la TTR que el método de IP convencional en solución, el hecho de que no exista enlace covalente entre el anticuerpo y las partículas magnéticas impide la reutilización de éstas y obliga a la separación del anticuerpo y la TTR por CE-MS o CapLC-MS. Para evitar estos inconvenientes, en un trabajo reciente no incluido en esta tesis doctoral, se exploró el uso de partículas magnéticas derivatizadas con grupos amino reactivos (*AffiAmino Ultrarapid Agarose*TM, UAAF) para purificar la TTR, las cuales permitieron enlazar covalentemente el anticuerpo y las partículas magnéticas a través de los grupos amino [211]. Como alternativa, en el **artículo 2.4** de esta tesis doctoral se utilizó un sorbente de base sílice con fragmentos Fab' (*antibody binding region*) de anticuerpo enlazados covalentemente a las partículas de sílice (previamente derivatizadas) a través de los grupos tiol.

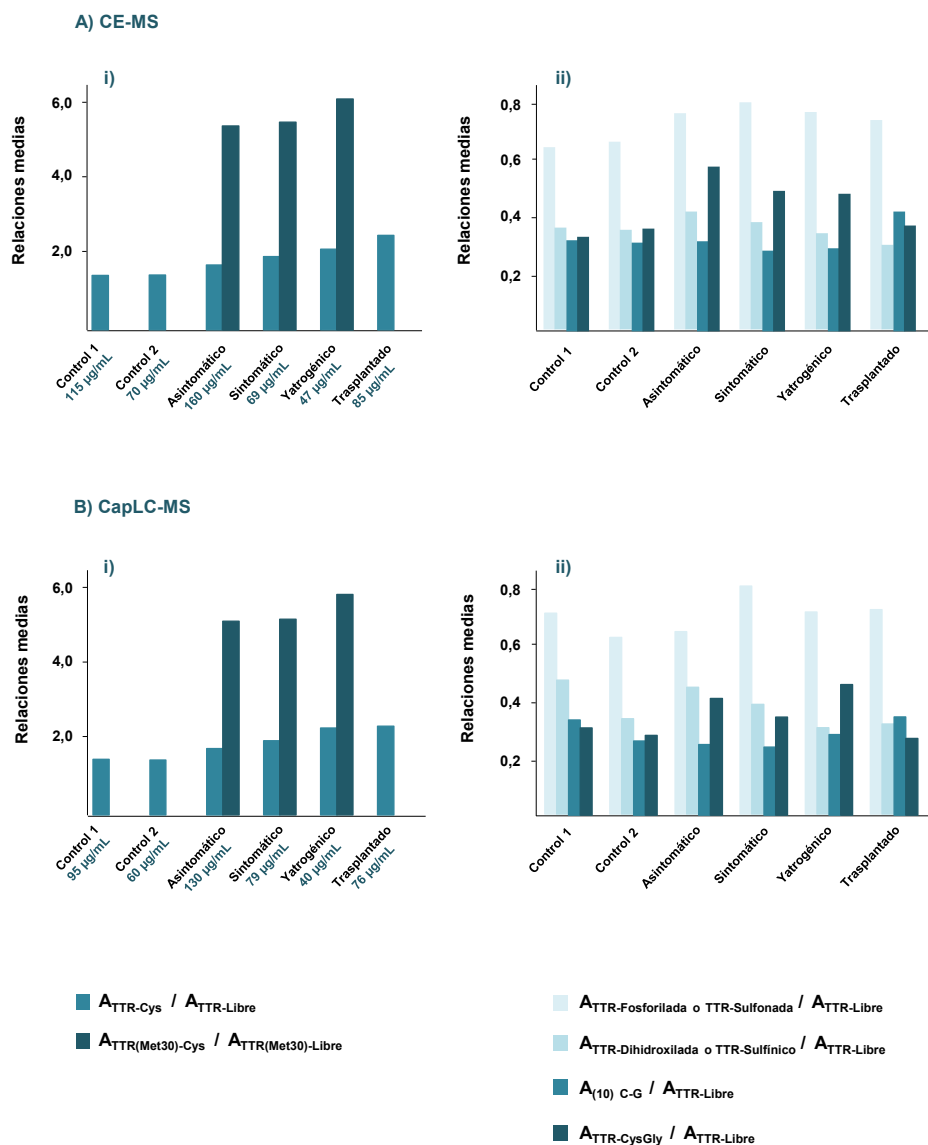


Figura 5.7. Relaciones medias ($n=3$) entre las A_p obtenidas por (A) CE-MS y (B) CapLC-MS para las proteoformas normales y mutantes de la TTR con respecto a la TTR-Libre y a la TTR(Met30)-Libre. (i: Proteoformas mayoritarias, ii: proteoformas minoritarias). En el eje de las x se muestran también las concentraciones totales de la TTR cuantificadas en el **artículo 2.2** de esta tesis doctoral por CE-MS y CapLC-MS.

5.1.3. Estudio de la estructura nativa de la TTR por IM-MS

El principal inconveniente que presentan las metodologías analíticas descritas anteriormente para el análisis de TTR en muestras de suero humano radica en que la estructura nativa

tetramérica de la proteína se ve alterada como consecuencia de la utilización de disoluciones ácidas para la disociación del complejo antígeno-anticuerpo durante la IP *off-line*, y al elevado vacío presente en el interior del espectrómetro de masas, lo que permite únicamente el estudio de su estructura monomérica. Con el objetivo de superar estas limitaciones y profundizar en el estudio de su estructura nativa, en el tercer trabajo de esta tesis doctoral (**artículo 2.3**), se desarrolló un método de IP *off-line* para la purificación de TTR en condiciones no desnaturalizantes, y se optimizó una metodología analítica para el análisis de TTR en suero mediante espectrometría de movilidad iónica acoplada a la espectrometría de masas con ionización por nanoelectrospray (nanoESI-IM-MS). Esta técnica permite la separación de iones gaseosos en función de su forma y tamaño, por lo que es capaz de distinguir entre iones con la misma m/z (si tienen diferente forma o tamaño) e interpretar el espectro de masas de mezclas de MO, DI y otros oligómeros que, tal y como se ha descrito anteriormente, se encuentran altamente solapados.

5.1.3.1. Análisis de patrones de TTR por IM-MS

Tal y como se ha comentado en la sección 1.4.4 de la introducción, la espectrometría de movilidad iónica (IMS) es una técnica que mide el tiempo (tiempo de deriva o *drift time*) que tarda un ion en fase gas en cruzar una región que contiene un gas inerte (normalmente N₂ o He) bajo la influencia de un pequeño campo eléctrico. El *drift time* depende de las colisiones de los iones en fase gas. Por consiguiente, los iones se separan en función de su sección transversal de colisión (Ω), la cual está relacionada con la forma y tamaño de dichos iones. Los iones pequeños cruzan antes la celda de IM debido a su menor Ω . Además, a medida que aumenta la carga del ión, aumenta la fuerza del campo eléctrico y, por tanto, los iones cruzan la celda de IM más rápidamente. Es por todo esto que se considera que el *drift time* de un ión es proporcional a su relación Ω/z (*drift time* $\propto \Omega/z$). Esta técnica ofrece una serie de ventajas que la hacen ideal para el estudio de la TTR, tales como la posibilidad de trabajar en condiciones experimentales suaves (BGEs acuosos volátiles con fuerzas iónicas bajas y pH neutro) y la capacidad de

distinguir entre iones con la misma m/z (si tienen diferentes Ω) e interpretar el espectro de masas de mezclas de MO, DI, trímero (TRI) y TE. Además, acoplada con la espectrometría de masas, la IM-MS permite obtener información analítica tridimensional: *drift time*, m/z e intensidad. La **Figura 5.8** muestra un ejemplo de la información obtenida por IM-MS para un patrón de TTR de 2000 $\mu\text{g/mL}$ (se muestra sólo la zona correspondiente al TE). Tal y como se observa en esta figura, los *extracted ion mobility profiles* (EIMs) muestran la separación e intensidad de los diferentes iones de la proteína tetramérica (iones con cargas +16, +15, +14 y +13). El *drift time* de los iones moleculares en cuestión decrece a medida que aumenta la carga, tal y como se deduce a partir de la aproximación $\text{drift time} \propto \Omega/z$. Además, el espectro de masas muestra las m/z y las intensidades correspondientes a dichos iones.

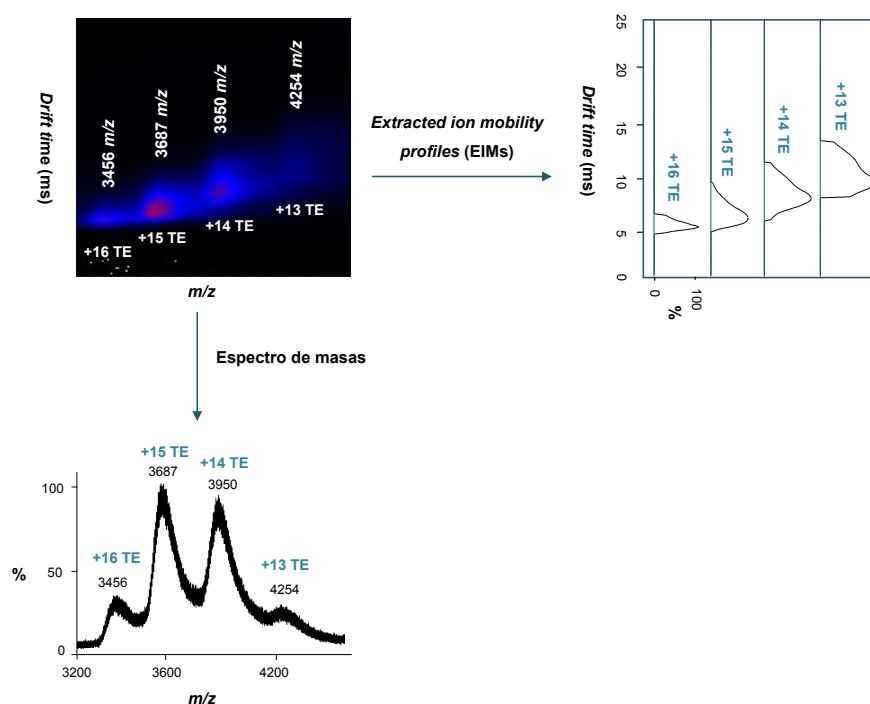


Figura 5.8. Información tridimensional (*drift time*, m/z e intensidad) obtenida por IM-MS para un patrón de TTR de 2000 $\mu\text{g/mL}$ (se muestra sólo la zona correspondiente al TE).

En el tercer trabajo de esta tesis doctoral (**artículo 2.3**), se desarrolló una metodología analítica para analizar la TTR mediante IM-MS empleando una interfase de nanoESI y experimentos de

infusión. Para ello, fue necesario optimizar diferentes condiciones experimentales y parámetros relacionados con la IM-MS. Primero, se estudió la influencia de la concentración del NH_4Ac (10 mM, 20 mM y 100 mM) a $\text{pH}=7,00$ en la calidad de los espectros de masas obtenidos para los patrones de TTR. En este sentido, se escogió una concentración 10 mM de NH_4Ac a $\text{pH}=7,00$ para el resto de estudios, ya que proporcionó una mayor estabilidad del spray y unas relaciones señal ruido (S/N) más elevadas. Una vez seleccionada la concentración de NH_4Ac más adecuada, se evaluó la influencia de algunos de los parámetros más importantes de la IM-MS, tales como el voltaje de cono, la energía de colisión de la trampa y la presión de vacío. Como ejemplo de la influencia del voltaje de cono, la **Figura 5.9** muestra (A) el espectro de masas y (B) los EIMs obtenidos por IM-MS para un patrón de TTR (2000 $\mu\text{g/mL}$) usando un voltaje de cono de 80 V. Tal y como se puede observar en esta figura, los espectros de masas muestran los clústeres de iones multicargados correspondientes al MO (iones con carga +7 y +6), al TE (iones con carga +16, +15, +14 y +13) y al octámero (OCT) (iones con carga +23, +22, +21 y +20). Estos iones se asignaron considerando las m/z experimentales de los iones moleculares del MO, TE y OCT, que coincidieron con las calculadas a partir de las M_r teóricas de la TTR teniendo en cuenta las proteoformas más abundantes descritas en el **artículo 2.1** para el patrón: TTR-Cys (52%) y TTR-Libre (48%). Tal y como se puede observar, el espectrómetro de masas empleado no permitió resolver las diferentes proteoformas de la TTR, probablemente debido a la dificultad para tener resoluciones adecuadas a valores de m/z tan elevados. Esta es una de las grandes limitaciones actuales para el análisis de oligómeros proteicos u otros complejos proteicos de elevada M_r mediante IM-MS, ya que se pierde información sobre las diferentes proteoformas, que debe ser obtenida mediante técnicas complementarias, como las que se han presentado en los **artículos 2.1 y 2.2**.

Es importante destacar que el OCT sólo sería estable en fase gas y que, en solución, está ampliamente demostrado que la TTR nativa es una proteína tetramérica que puede disociarse para producir MO, DI o TRI. Si comparamos con el espectro de masas de la **Figura 5.9 C** para un patrón de TTR de 2000 $\mu\text{g/mL}$ usando un voltaje de cono de 150 V, se puede observar que

los clústeres de iones multicargados correspondientes al OCT y al TE persistieron (aunque en menor proporción), mientras que la abundancia de los iones multicargados correspondientes al MO y al DI aumentó, sugiriendo la disociación del TE a este voltaje de cono tan elevado. Por este motivo, se seleccionó un voltaje de cono de 80 V para el resto de estudios con los patrones de TTR y las muestras de suero humano.

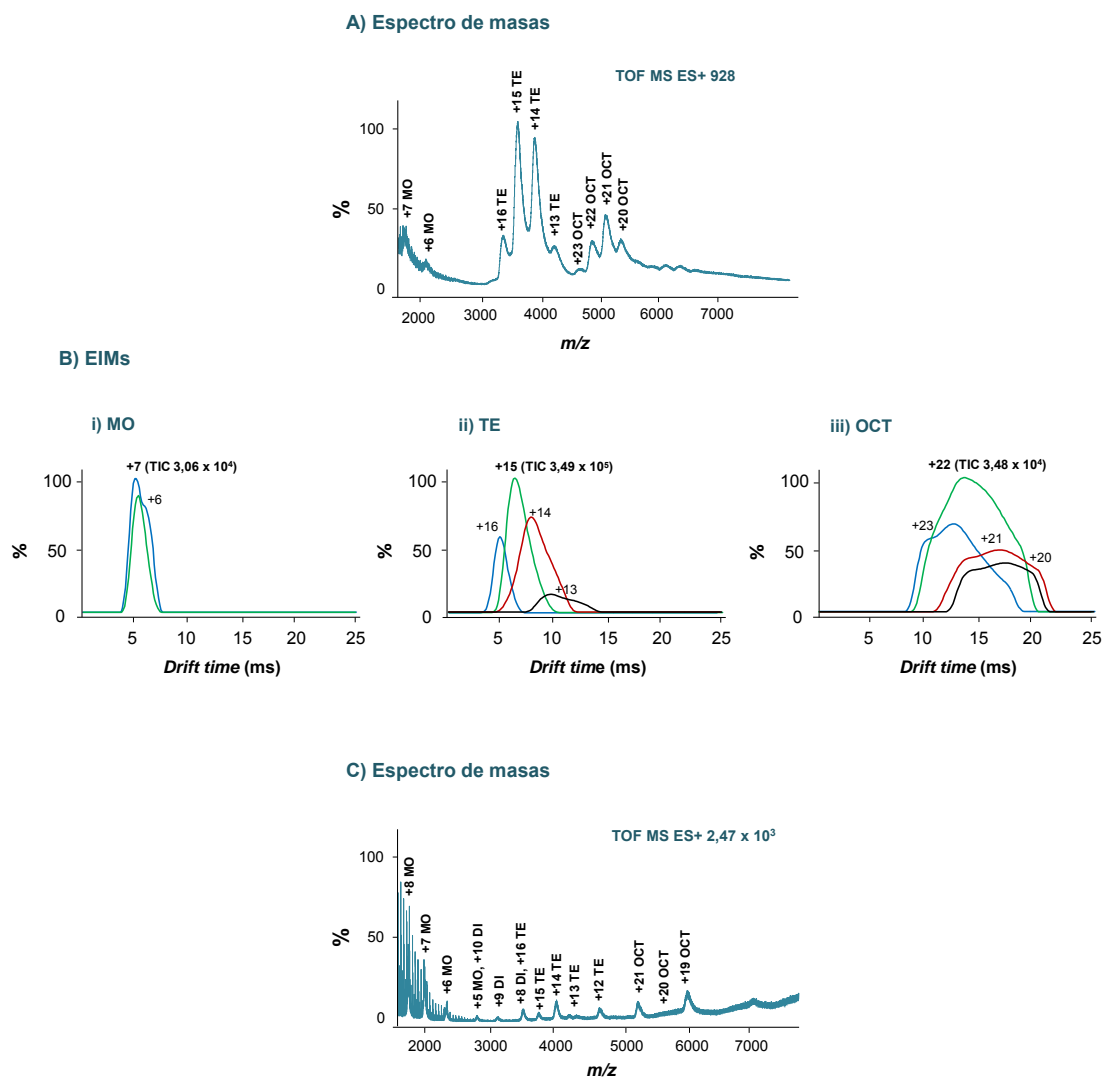


Figura 5.9. (A) Espectro de masas y (B) EIMs obtenidos por IM-MS para un patrón de TTR (2000 $\mu\text{g/mL}$) usando un voltaje de cono de 80 V. (C) Espectro de masas obtenido por IM-MS para un patrón de TTR de 2000 $\mu\text{g/mL}$ usando un voltaje de cono de 150 V.

En estas condiciones, la reproducibilidad (%RSD, $n=3$ muestras independientes) fue adecuada para los *drift times* de los iones moleculares más abundantes del MO (+7), del TE (+15) y del OCT (+22), con valores de 3,9%, 4,9% y 3,2%, respectivamente. Por lo que hace referencia a las A_p obtenidas de los diferentes EIMs, y con el objetivo de minimizar la variabilidad debida a las condiciones experimentales, se optó por evaluar la reproducibilidad, no de las A_p , sino de las relaciones entre ellas. Se obtuvieron, de esta manera, unos valores de reproducibilidad del 5,3%, 8,1% y 8,2% para las relaciones de las A_p de los iones +22OCT/+7MO, +22OCT/+15TE y +15TE/+7MO, respectivamente.

5.1.3.2. Optimización de un método de IP *off-line* en condiciones no desnaturalizantes

Los métodos de IP *off-line* de TTR que se describen en los **artículos 2.1 y 2.2** de esta tesis doctoral se basan, como muchos otros descritos en la bibliografía, en la utilización de disoluciones ácidas para disociar el complejo antígeno-anticuerpo, alterando de esta manera la estructura nativa tetramérica de la proteína. En el tercer trabajo de esta tesis doctoral (**artículo 2.3**), se optimizó un método de IP *off-line* (con muestras de suero control) con el objetivo de prevenir la desnaturalización y la disociación del complejo tetramérico de la TTR.

En primer lugar, se comprobó la reversibilidad del proceso de disociación del TE de la TTR después de disociar el complejo antígeno-anticuerpo con HAc 1,0 M. Para ello, antes del análisis por IM-MS, fue necesario cambiar el medio ácido por NH_4Ac 10 mM pH=7,00 mediante ultrafiltración con filtros de corte molecular de M_r 10000 pasivados con PEG. La **Figura 5.10** muestra el espectro de masas y los EIMs de la TTR para una muestra de suero control después de llevar a cabo la IP *off-line* en solución utilizando HAc 1,0 M (para disociar el complejo TTR-anticuerpo) y cambiándolo posteriormente por una solución NH_4Ac 10 mM pH=7,00. Tal y como se puede observar en esta figura, los espectros de masas muestran los clústeres de iones multicargados correspondientes al MO (iones con carga +9, +8 y +7), al TE (iones con carga +16, +15, +14 y +13) y al OCT (iones con carga +22, +21, +20 y +19). De

nuevo, tal y como ocurría con el patrón de TTR, los iones moleculares se asignaron considerando las M_r teóricas de las proteoformas más abundantes descritas en el **artículo 2.1**: TTR-Cys (52%) y TTR-Libre (48%). Tal y como se observa en esta figura, aunque el proceso pareció ser reversible en cierta medida (se pudo detectar cierta cantidad de TE), los iones correspondientes al TE y al OCT presentaban una abundancia baja si la comparamos con la intensidad del MO (mirar las intensidades máximas en las etiquetas de los EIMs para cada una de las estructuras, **Figura 5.10 B**), sugiriendo una reversibilidad parcial del TE después de disociar el complejo TTR-anticuerpo con HAc.

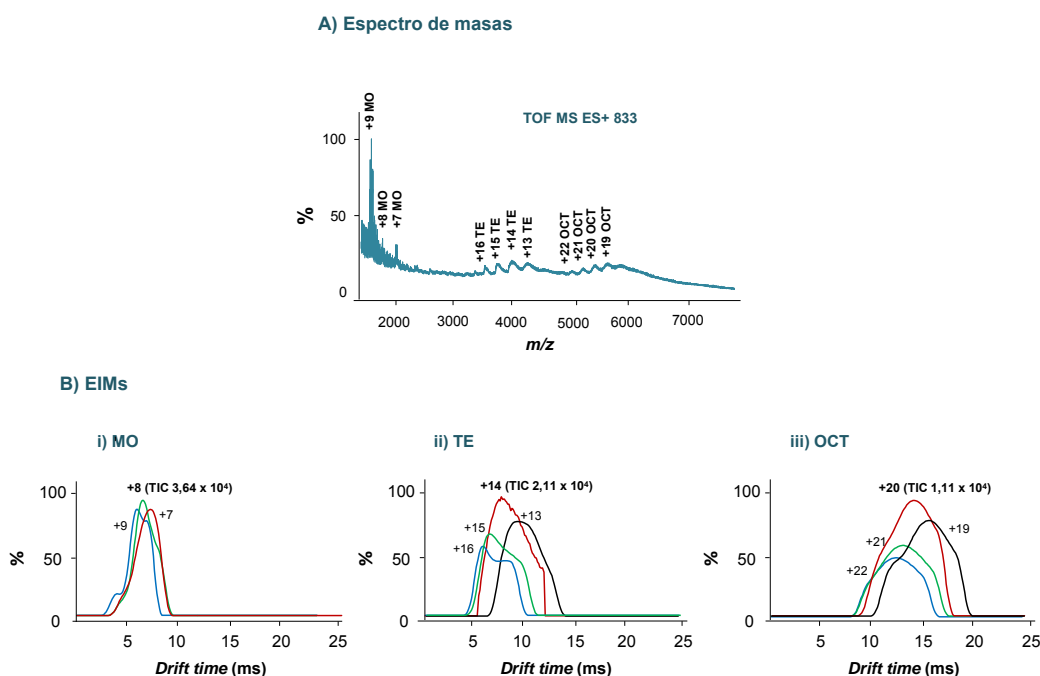


Figura 5.10. (A) Espectro de masas y (B) EIMs obtenidos por IM-MS para una muestra de suero control después de llevar a cabo la IP *off-line* en solución utilizando HAc 1,0 M para disociar el complejo TTR-anticuerpo y cambiándolo por una solución NH_4Ac 10 mM pH=7,00.

Con el objetivo de superar esta limitación, se desarrolló un método de IP *off-line* en solución en condiciones no desnaturizantes. En condiciones óptimas, este procedimiento consistió en utilizar NH_3 100 mM pH=11,30 para disociar el complejo TTR-anticuerpo, lo que permitió

conservar la estructura nativa tetramérica de la TTR. Con el objetivo de obtener una sensibilidad óptima, también fue necesario cambiar el NH_3 por NH_4Ac 10 mM pH=7,00 mediante ultrafiltración con filtros de corte molecular de M_r 10000 pasivados con PEG. A diferencia de lo que sucedía en el caso de la IP convencional en solución en condiciones desnaturalizantes (**artículo 2.1**), en este caso tampoco fue necesaria la ultrafiltración para separar el gel polimérico activador de la coagulación o el anticuerpo.

5.1.3.3. Análisis de TTR en muestras de suero humano por IM-MS

Una vez optimizado el tratamiento de muestra para la purificación de la TTR del suero, se procedió a analizar por IM-MS todas las muestras de suero disponibles. La **Figura 5.11** muestra los espectros de masas y los EIMS de la TTR obtenidos para muestras de suero de un control sano y de un paciente de FAP-I sintomático después de llevar a cabo el procedimiento de IP *off-line* en condiciones no desnaturalizantes. Tal y como se muestra en las **Figuras 5.11 A y C**, en ambos casos se pudo observar la presencia de los clústeres de iones multicargados correspondientes al DI (iones con carga +13, +12, +11 y +10), al TRI (iones con carga +18, +17, +16, +15 y +14) y al TE (iones con carga +16, +15, +14 y +13). Por el contrario, no se observaron iones moleculares correspondientes al MO, lo que indicó la efectividad del método de IP optimizado para preservar la estructura nativa de la TTR. Cabe destacar que los iones moleculares +12DI y +18TRI, así como los iones +10DI y +15TRI presentan las mismas m/z pero distintos *drift times* debido a sus diferencias en Ω/z , tal y como se puede observar en los EIMS de la **Figura 5.11 B**.

A diferencia de lo que ocurría en el suero de un paciente control, donde se tenían en cuenta las proteoformas más abundantes de la TTR normal (TTR-Cys (52%) y TTR-Libre (48%)), para un paciente sintomático de FAP-I, las m/z experimentales coincidieron con las calculadas considerando las proteoformas oxidadas y no oxidadas más abundantes de la TTR mutante,

TTR(Met30)-Cys (36%) y TTR(Met30)-Libre (12%) y de la TTR normal, TTR-Cys (33%) y TTR-Libre (19%) (artículo 2.1).

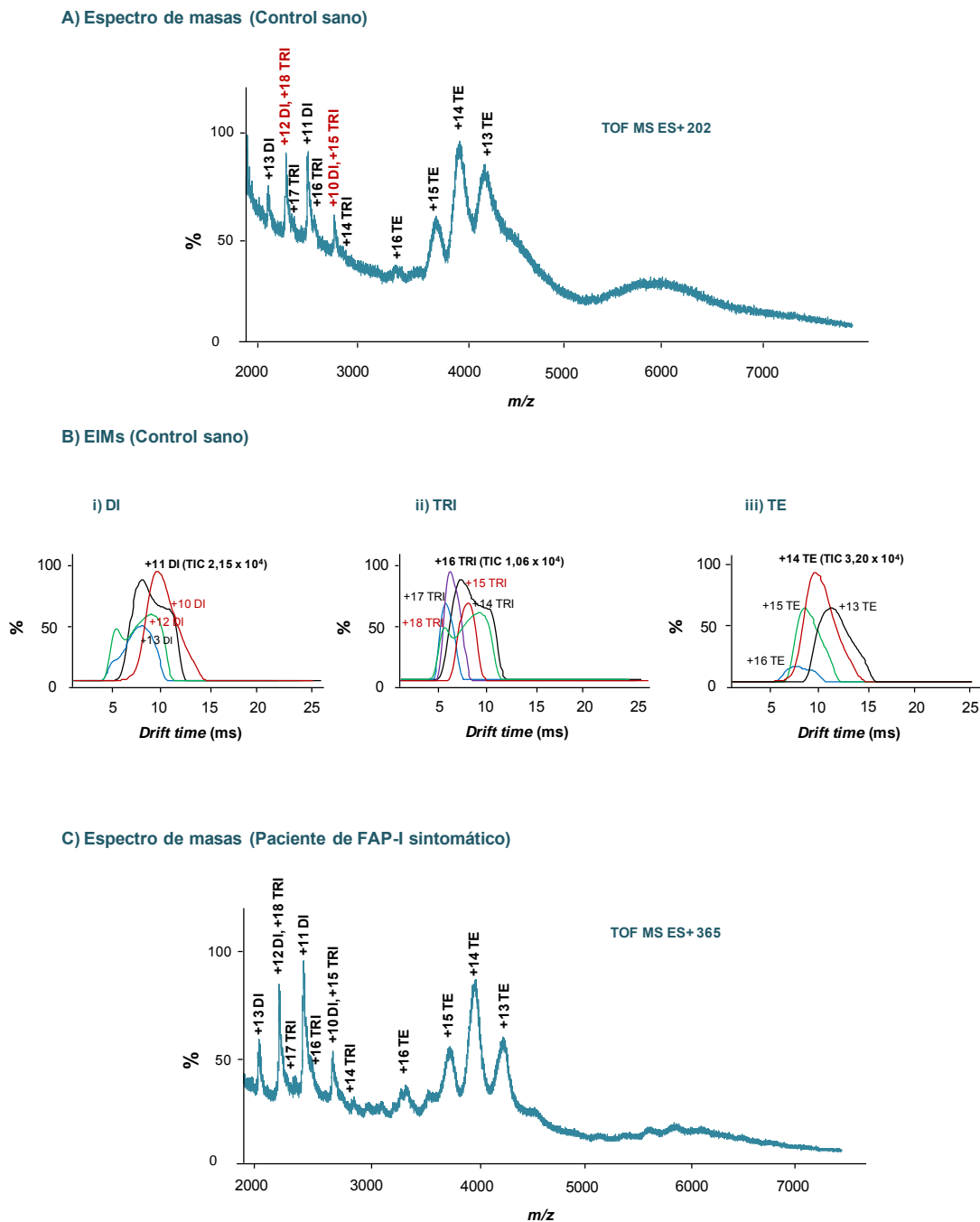


Figura 5.11. (A y C) Espectros de masas y (B) EIMs obtenidos por IM-MS para una muestra de suero control y de un paciente de FAP-I sintomático después de llevar a cabo la IP *off-line* en solución utilizando NH_3 100 mM pH=11,30 para disociar el complejo antígeno anticuerpo y cambiándolo por una solución NH_4Ac 10 mM pH=7,00.

El método desarrollado en el **artículo 2.3** para el análisis de TTR en muestras de suero por IM-MS proporcionó una reproducibilidad (%RSD, $n=3$ muestras independientes) adecuada para los *drift times* de los iones moleculares más abundantes del DI (+11), del TRI (+16) y del TE (+14), con valores de entre 0,8-1,2% y 1,6-4,9% para las muestras de suero de un control sano y de un paciente sintomático de FAP-I, respectivamente. Por lo que hace referencia a las relaciones entre las A_p de los iones moleculares más abundantes del TE, TRI y DI (+14TE/+11DI, +16TRI/+11DI, +14TE/+16TRI), se obtuvieron unos valores de reproducibilidad de entre 3,3-7,6% y 3,1-12,3% para las muestras de suero de un control sano y de un paciente sintomático de FAP-I, respectivamente.

Con el objetivo de profundizar en la comprensión de los mecanismos de agregación de la TTR en FAP-I, los cuales podrían estar iniciados por cambios en la estructura nativa tetramérica de la TTR (tales como cambios estructurales y conformacionales, disociación del complejo oligomérico, presencia de intermedios, etc.), el gráfico de la **Figura 5.12** muestra las relaciones medias ($n=3$ muestras independientes) para las A_p de los iones moleculares +14TE/+11DI, +16TRI/+11DI y +14TE/+16TRI para todas las muestras de suero disponibles. A partir de este gráfico se pudieron extraer las siguientes conclusiones:

- Las relaciones entre la A_p de los iones +14TE/+11DI y +14TE/+16TRI eran más altas en los pacientes de FAP-I (asintomático, sintomático y yatrogénico). Este hecho sugeriría una mayor estabilidad del TE en pacientes de FAP-I, lo que podría estar relacionado con la mayor abundancia de las proteoformas oxidadas TTR-Cys y TTR(Met30)-Cys descrita en los trabajos anteriores (**artículos 2.1 y 2.2**). Esto indicaría que la estabilización de TE, y no su disociación, sería un factor importante en la oligomerización y en la formación de agregados amiloides. Por el contrario, las relaciones entre las A_p de los iones +16TRI/+11DI se mantuvieron constantes en todos los casos.

- En el paciente de FAP-I trasplantado de hígado, las relaciones entre las A_p de los iones $+14TE/+11DI$ y $+14TE/+16TRI$ eran similares a las relaciones obtenidas en muestras de suero control, lo que confirmaría de nuevo la efectividad del tratamiento terapéutico. La ausencia de las proteoformas mutantes se confirmó con los valores de m/z obtenidos para los iones moleculares, que eran próximos a los valores observados en los controles sanos.

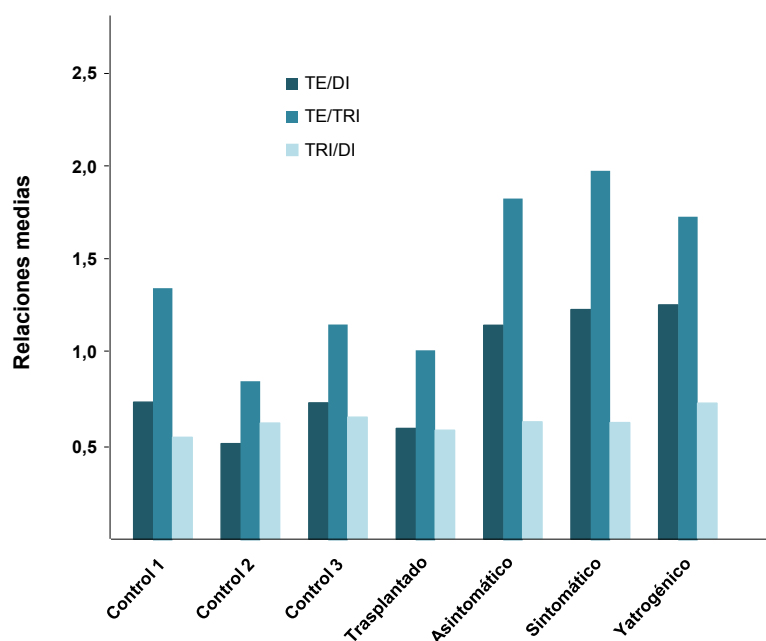


Figura 5.12. Relaciones medias ($n=3$) entre las A_p obtenidas por IM-MS para los iones moleculares $+14TE/+11DI$, $+16TRI/+11DI$ y $+14TE/+16TRI$ de la TTR en todas las muestras de suero disponibles. Las A_p se calcularon a partir de los EIMs obtenidos por IM-MS para los iones $+11DI$, $+16TRI$ y $+14TE$ que eran específicos del DI, TRI y TE.

La **Tabla 5.4** muestra las principales características de los tres métodos de IP *off-line* para la purificación de TTR desarrollados en esta tesis doctoral (**artículos 2.1, 2.2 y 2.3**), con sus ventajas y desventajas. A modo de resumen, la IP convencional en solución en condiciones desnaturizantes (usando HAc 1,0 M pH=2,30 para disociar el complejo TTR-anticuerpo) antes del análisis por CE-MS, permitió únicamente la detección de las

proteoformas más abundantes del MO para la TTR normal (TTR-Cys y TTR-Libre) y la TTR mutante (TTR(Met30)-Cys y TTR(Met30)-Libre). Por el contrario, la IP con partículas magnéticas en condiciones desnaturalizantes (usando glicina 50 mM pH=2,80 para eluir y disociar el complejo TTR-anticuerpo) previo al análisis por CE-MS y CapLC-MS, permitió detectar tres proteoformas monoméricas adicionales de la TTR normal (TTR-Fosforilada o TTR Sulfonada, TTR-Dihidroxilada o TTR-Sulfínico y TTR-CysGly). Finalmente, la IP convencional en solución en condiciones no desnaturalizantes (usando NH₃ 100 mM pH=11,30 para disociar el complejo TTR-anticuerpo) antes del análisis por IM-MS, permitió la detección de diferentes oligómeros de la TTR (DI, TRI y TE). Sin embargo, el espectrómetro de masas empleado no permitió resolver las diferentes proteoformas de la TTR, probablemente debido a la dificultad para tener resoluciones adecuadas a valores de m/z tan elevados ($>2000\ m/z$). Esta es una de las grandes limitaciones actuales para el análisis de oligómeros proteicos u otros complejos proteicos de elevada M_r mediante IM-MS, ya que se pierde información sobre las diferentes proteoformas, que debe ser obtenida mediante técnicas complementarias, tales como CE-MS y CapLC-MS.

Tabla 5.4. Características principales de los métodos de IP *off-line* desarrollados en esta tesis doctoral para el análisis de TTR en muestras de suero mediante CE-MS, CapLC-MS y IM-MS.

IP <i>off-line</i>	IP convencional en solución		IP con partículas magnéticas
Condiciones	Desnaturalizantes	No desnaturalizantes	Desnaturalizantes
Volumen (μL)/cantidad (μg) de anticuerpo	50/120	20/48	50/120
Volumen de suero (μL)	50	50	75
Volumen de partículas magnéticas (μL)	-	-	50
Tiempo de incubación suero:anticuerpo	18 h	18 h	20 min
Temperatura de incubación (°C)	4	4	Temperatura ambiente
Disociación/elución	HAc 1,0 M pH=2,30	NH ₃ 100 mM pH=11,30	Glicina 50 mM pH=2,80
Volumen de disociación/elución (μL)	50	50	50
Pretratamiento de muestra	Ultrafiltración con filtros de corte molecular de M _r 100000 pasivados con 5% PEG en agua para separar la TTR del anticuerpo y del gel polimérico	Ultrafiltración con filtros de corte molecular de M _r 10000 pasivados con 5% PEG en agua para cambiar el NH ₃ por NH ₄ Ac 10 mM pH=7,00	Ultrafiltración con filtros de corte molecular de M _r 10000 pasivados con 5% PEG en agua para cambiar el eluyente por HAc 1,0 M pH=2,30
Análisis realizado	CE-MS	IM-MS	CE-MS/CapLC-MS
BGEs/Fases móviles recomendadas	HAc 1,0 M pH=2,30	NH ₄ Ac 10 mM pH=7,00	HAc 1,0 M pH=2,30/ACN:agua 0,1% (v/v) HFor
Oligómeros detectados	MO	DI, TRI y TE	MO
Proteoformas detectadas	Normales: TTR-Cys, TTR-Libre Mutantes: TTR(Met30)-Cys, TTR(Met30)-Libre	No se pueden resolver ^a	Normales: TTR-Cys, TTR-Libre, TTR-Fosforilada/TTR-Sulfonada, TTR-Dihidroxilada/TTR-Sulfínico, (10) C-G, TTR-CysGly Mutantes: TTR(Met30)-Cys, TTR(Met30)-Libre

^aLa resolución del espectrómetro de masas no permitió resolver las diferentes proteoformas porque los iones moleculares detectados tienen m/z muy elevadas (>2000 m/z).

5.1.4. Análisis de TTR por IA-SPE-CE-MS

A pesar de los buenos resultados obtenidos en los trabajos anteriores (**artículos 2.1, 2.2 y 2.3**) para el análisis de TTR en muestras de suero humano, el tratamiento de muestra por *IP off-line* es laborioso y dificulta el análisis de un gran número de muestras por unidad de tiempo. Por otro lado, el principal inconveniente de la CE es la baja sensibilidad en unidades de concentración. Esto es debido principalmente a que las reducidas dimensiones de los capilares de separación limitan el volumen de muestra inyectado y el camino óptico en el caso de utilizar detectores de espectrofotometría ultravioleta (UV). Tal y como ya se ha comentado en la introducción de esta tesis doctoral (sección 1.5), existen diferentes alternativas para disminuir los LODs de la CE. Una de las mejores consiste en acoplar la extracción en fase sólida en línea a la CE (SPE-CE). La preconcentración por SPE-CE se basa en la interacción reversible de los analitos de la muestra con una fase estacionaria. La fase estacionaria se inmoviliza en un soporte sólido y se coloca el sorbente en un dispositivo, conocido como microcartucho de extracción o preconcentrador, lo más próximo posible al extremo de entrada del capilar de separación. Esta fase estacionaria permite retener los analitos presentes en un gran volumen de muestra (~100 μ L), que posteriormente se eluyen directamente dentro del capilar de separación con un pequeño volumen de una disolución adecuada (~50 nL). De esta manera, se consigue purificar el analito de interés, un aumento de su concentración y, globalmente, una disminución de los LODs. Las fases estacionarias utilizadas en SPE-CE pueden ser muy diversas: fases estacionarias hidrocarbonadas de base sílice (C_{18} , C_8 , C_2 , etc), poliméricas, de intercambio iónico, de inmunoafinidad (IA), etc. En esta tesis doctoral se evaluó el uso de sorbentes de IA con fragmentos Fab' de un anticuerpo para el análisis de TTR en muestras de suero mediante extracción en fase sólida de inmunoafinidad en línea a la electroforesis capilar acoplada a la espectrometría de masas (IA-SPE-CE-MS). Para ello, se optimizó y validó una metodología analítica mediante el uso de patrones de TTR y se aplicó al análisis de muestras de suero control y de pacientes sintomáticos de FAP-I. Además, esta metodología se comparó con una

previamente optimizada en nuestro grupo de investigación para el análisis de TTR mediante IA-SPE-CE-MS con un sorbente de IA con anticuerpo intacto [211].

5.1.4.1. Construcción del preconcentrador

A continuación se describe el procedimiento empleado para la construcción de los preconcentradores empleados en los **artículos 2.4** (IA-SPE-CE-MS), **4.1** y **4.2** (C₁₈-SPE-CE-MS). El cuerpo del preconcentrador consiste en un pequeño fragmento de capilar de sílice fundida de 7 mm de longitud (L_T), 250 μm (id) y 365 μm (od), relleno de sorbente. Para evitar la pérdida de las partículas empaquetadas, se colocan unas fritas en cada extremo del microcartucho. Estas fritas se extraen de cartuchos convencionales para SPE *off-line* (*Sep-Pack*, *Waters Corporation*) y se cortan de una medida adecuada para adaptarlos al diámetro interno del microcartucho. El proceso de construcción del microcartucho, el cual se supervisa a través de una lupa binocular, es el siguiente (**Figura 5.13**):

- El primer paso a realizar es la activación del capilar (20 min de NaOH 1,0 M y 20 min de agua) donde se conectará en línea el microcartucho. De esta manera se evita la exposición del sorbente del microcartucho a las condiciones extremas de la disolución de NaOH.
- Este capilar se corta en dos fragmentos. Para SPE-CE-MS, el fragmento del capilar de entrada medirá 7,5 cm, y el fragmento de salida, 64,5 cm. El corte del capilar se realiza con un cortador de capilar específico para que sea lo más limpio posible, y así minimizar los volúmenes muertos y pérdidas a través de las juntas.
- Se introduce una frita en uno de los extremos del capilar de 7 mm L_T que será el cuerpo del microcartucho (**Figura 5.13 A**).
- Se une el microcartucho al fragmento de entrada del capilar de separación con un tubo de tygon™ de diámetro adecuado (**Figura 5.13 B**).

- Se aspiran las partículas de fase estacionaria al interior del microcartucho mediante la aplicación de vacío (**Figura 5.13 B-C**).
- Se introduce una frita en el otro extremo del microcartucho (**Figura 5.13 D**).
- Se une el microcartucho al fragmento de salida del capilar de separación mediante otro tubo de tygon™ (**Figura 5.13 E**).
- Por último, antes de comenzar los análisis, se comprueba manualmente con una jeringa que el microcartucho empaquetado no ofrece demasiada restricción al flujo, ni inestabilidad en la corriente eléctrica durante un experimento de CE.

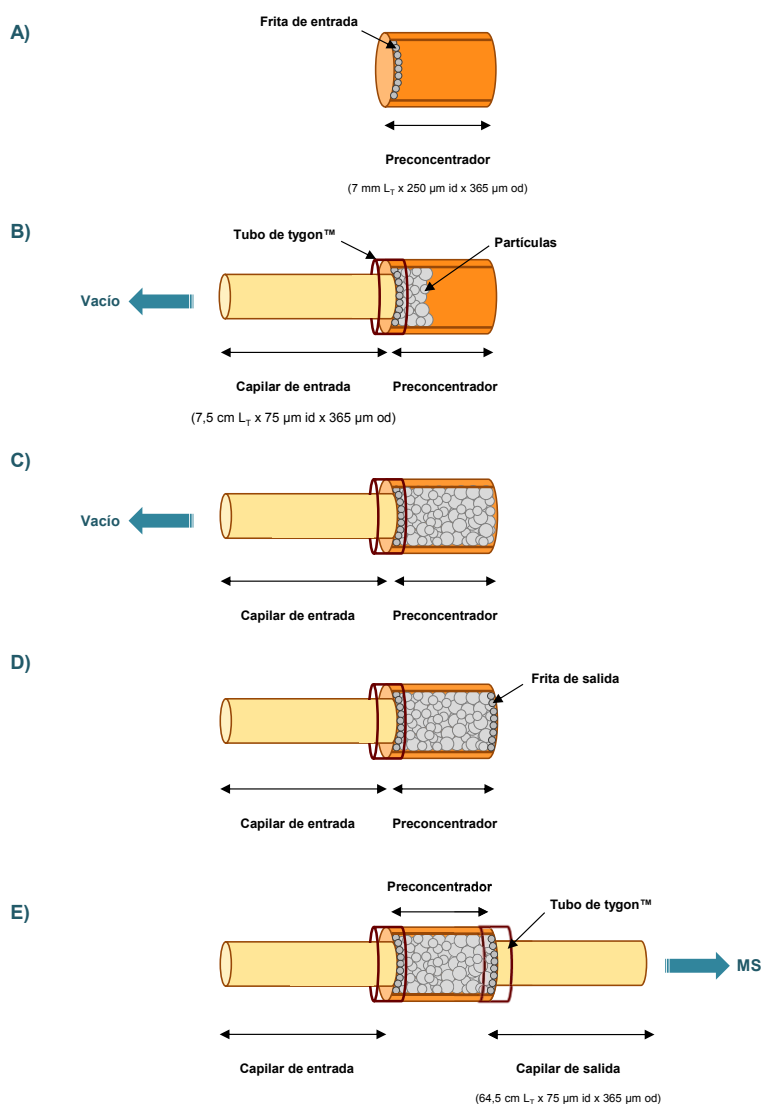


Figura 5.13. Proceso de construcción del microcartucho para SPE-CE-MS.

Una vez preparada la columna de SPE-CE-MS con el sorbente de IA adecuado, se debe desarrollar y optimizar la metodología analítica, lo que constituye todo un reto, pues, en este caso, se han de compatibilizar las particularidades de la inmunoextracción, la separación electroforética y la detección por MS.

En un trabajo anterior de nuestro grupo de investigación [211], se estudió la preparación de un sorbente de IA con un anticuerpo intacto contra la TTR usando, como soporte, partículas magnéticas derivatizadas con grupos amino reactivos (*AffiAmino Ultrarapid Agarose*TM, UAAF), las cuales permiten enlazar covalentemente el anticuerpo intacto a través de sus propios grupos amino. Sin embargo, este tipo de enlace no garantiza la orientación adecuada de los anticuerpos enlazados (**Figura 5.14**). Aun así, en este caso, los resultados para el análisis de TTR resultaron excelentes [211].

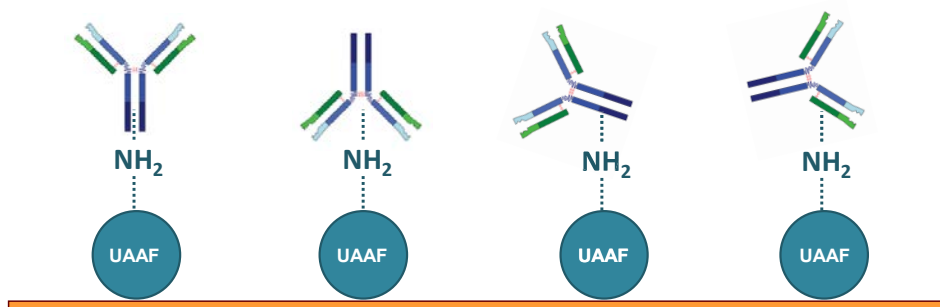


Figura 5.14. Posibles orientaciones del anticuerpo intacto inmovilizado en el soporte de partículas magnéticas UAAF cuando el anticuerpo se enlaza a través de los grupos amino disponibles.

Como alternativa al procedimiento anterior, en el **artículo 2.4** de esta tesis doctoral se evaluó la preparación de un sorbente de IA contra la TTR a partir de fragmentos Fab' del mismo anticuerpo que el utilizado en el artículo de Perú-Gascón et al. [211]. En este caso, se emplearon como soporte, partículas de sílice con grupos aminopropil derivatizados convenientemente para obtener grupos succinimidil. La metodología escogida, previamente optimizada en nuestro grupo de investigación para el análisis de neuropéptidos mediante IA-SPE-CE-MS [209],

consiste en enlazar covalentemente los fragmentos Fab' de anticuerpo a través de los grupos tiol. Previamente, se generan los fragmentos Fab' de anticuerpo mediante fragmentaciones controladas y sucesivas del anticuerpo intacto. Como los grupos tiol están distantes a los sitios activos antígeno-anticuerpo, la orientación de los fragmentos Fab' inmovilizados es siempre adecuada para la captura de los analitos y permiten ampliar la superficie activa del sorbente. El procedimiento experimental para la preparación de este sorbente de IA se detalla en la introducción de esta tesis doctoral (sección 1.5.2).

5.1.4.2. Análisis de patrones de TTR por IA-SPE-CE-MS

En el cuarto trabajo de esta tesis doctoral (**artículo 2.4**), se evaluó primero la capacidad de la IA-SPE-CE-MS con fragmentos Fab' de anticuerpo para analizar patrones de TTR. Las condiciones establecidas en el trabajo previo de nuestro grupo de investigación con anticuerpo intacto [211], fueron el punto de partida para optimizar la nueva metodología. De esta manera, se empleó un BGE de NH_4Ac 10 mM pH 7,00, ya que el uso de un BGE ácido provocaría la desnaturalización de los fragmentos Fab', así como la elución de la TTR durante los métodos de limpieza y acondicionamiento del capilar de separación. Así mismo, fue necesario disolver los patrones de TTR en PBS, ya que la proteína no se retenía una vez disuelta en agua o en el BGE neutro. El PBS facilitaría la interacción entre la TTR y los fragmentos Fab' de anticuerpo, probablemente debido a que presenta una osmolaridad y fuerza iónica similares a las de los fluidos biológicos.

Con el objetivo de obtener un método reproducible para el análisis de TTR, se evaluaron diferentes eluyentes ácidos y básicos: glicina 100 mM/HCl 50 mM (pH=2,50), ácido fosfórico 100 mM (pH=1,50), NH_3 100 mM (pH=11,30) y NH_3 100 mM (pH=11,30) con un 10% (v/v) de metanol (MeOH) o ACN. Sin embargo, ninguno de estos eluyentes permitió obtener resultados reproducibles, probablemente debido a la poca estabilidad de los fragmentos Fab' y las partículas con grupos succinimidil a pHs ácidos y fuertemente básicos [238]. Finalmente, para

el resto de estudios, se escogió un eluyente básico a un pH inferior que los anteriores consistente en glicina 100 mM/ NH_3 100 mM pH=9,50, ya que permitió eluir la TTR de una manera reproducible, así como realizar múltiples análisis en un mismo preconcentrador.

Una vez optimizada la composición del eluyente, se procedió a evaluar el efecto del tiempo de introducción de muestra (10, 20, 30 y 45 min) a 930 mbar y a 50 mbar. La **Figura 5.15** muestra el volumen (en μL) y la cantidad (en μg de proteína) de TTR introducida a cada presión y tiempo.

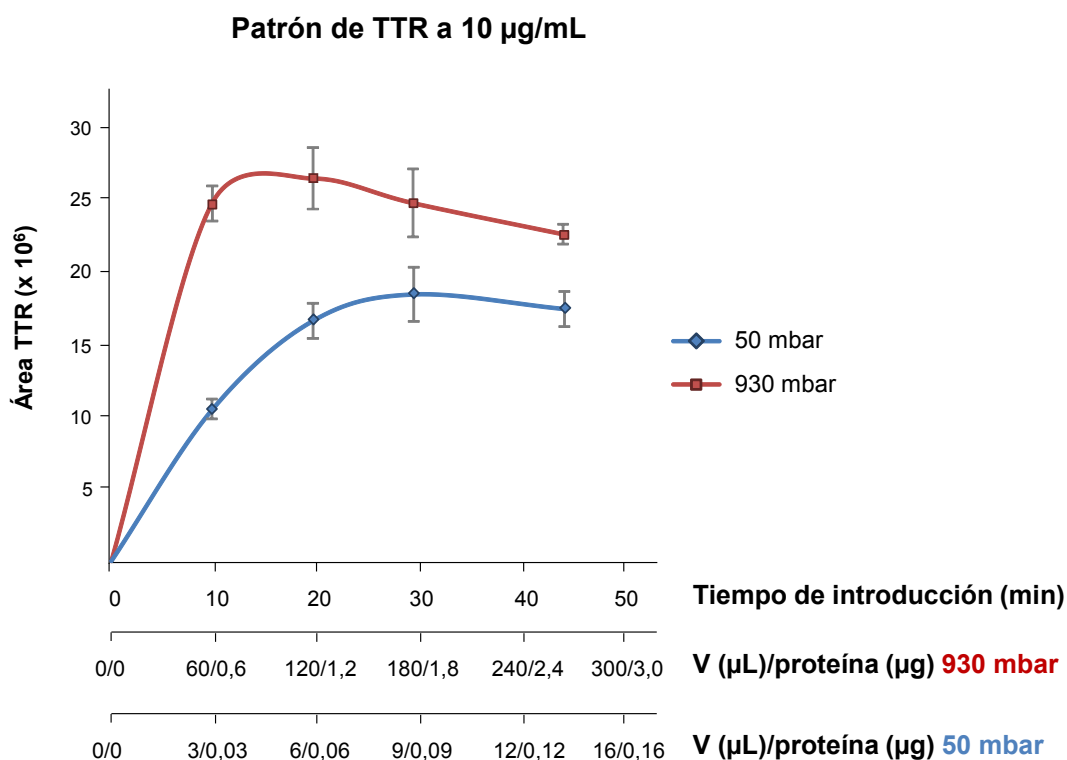


Figura 5.15. Estudio del efecto del tiempo de introducción de muestra (10, 20, 30 y 45 min) a 930 mbar y a 50 mbar por Fab'-IA-SPE-CE-MS (patrón de TTR de 10 $\mu\text{g/mL}$ de concentración). Las A_p corresponden a una media de tres análisis ($n=3$).

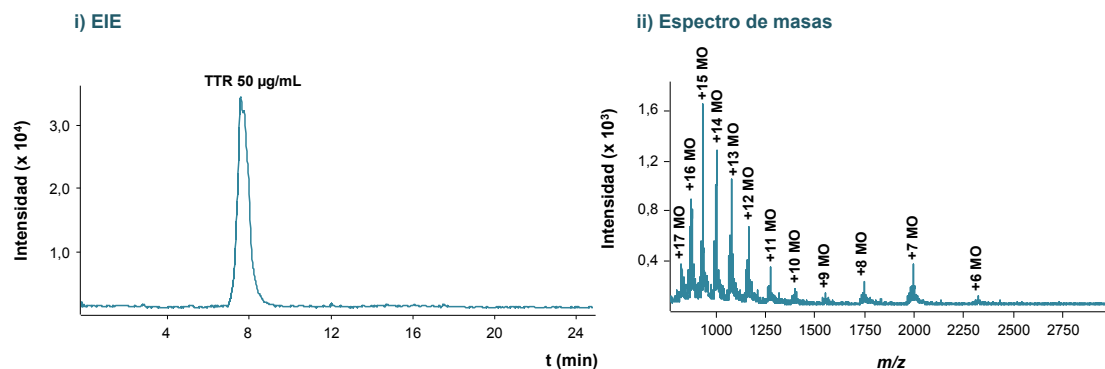
Tal y como se puede observar en esta figura, en todos los casos, las A_p fueron mayores a 930 mbar. A esta presión, los valores máximos de A_p se obtuvieron a los 20 min, aunque con poca

diferencia respecto a los obtenidos a los 10 min. En este caso, y debido a que los picos de TTR se ensanchaban a tiempos de introducción de muestra superiores a 10 min (llevando a inestabilidad en la corriente), se decidió introducir la muestra a 930 mbar durante 10 min.

Las **Figuras 5.16 A y B** muestran los EIEs (i) y los espectros de masas (ii e iii) obtenidos para un patrón de TTR empleando un BGE de NH_4Ac 10 mM pH=7,00 en (A) CE-MS (50 $\mu\text{g/mL}$) y (B) Fab'-IA-SPE-CE-MS (25 $\mu\text{g/mL}$). A diferencia de lo ocurrido en CE-MS, la TTR eluyó como dos picos con formas y espectros de masas diferentes en el caso de Fab'-IA-SPE-CE-MS. Esto es probablemente debido a la presencia de dos conformeros de la TTR, correspondientes a estructuras plegadas (*folded*) y desplegadas (*unfolded*) de la proteína. Está ampliamente reconocido que las estructuras plegadas y desplegadas de las proteínas producen espectros de masas con diferentes distribuciones de carga [239]. Así, tal y como se puede observar en el espectro de masas del conformero plegado (**Figura 5.16 B-ii**), éstos tienden a presentar distribuciones con pocos iones y de cargas pequeñas (+7, +8 y +9), mientras que los conformeros desplegados (**Figura 5.16 B-iii**) producen distribuciones con un mayor número de iones y centradas en cargas mucho mayores (+13, +14 y +15). Estas diferencias estarían relacionadas con la accesibilidad de los grupos ionizables básicos de la secuencia aminoacídica, los cuales tenderían a cargarse más fácilmente en los conformeros desplegados [239].

Así mismo, un análisis exhaustivo del clúster de iones multicargados obtenidos en el espectro de masas del conformero desplegado, demostró también la presencia de una pequeña cantidad de DI. Tal y como se puede observar en el espectro de masas, los iones moleculares con m/z nominales de 1984, 2314, y 2777 corresponderían a la suma de la contribución del MO (cargas +7, +6 y +5, respectivamente) y del DI (cargas +14, +12 y +10, respectivamente), mientras que los iones moleculares con m/z nominales de 2136 y 2525 se detectarían exclusivamente para el DI (cargas +13 y +11, respectivamente) (**Tabla 5.1**, sección 5.1.1 de esta tesis doctoral).

A) CE-MS



B) Fab'-IA-SPE-CE-MS

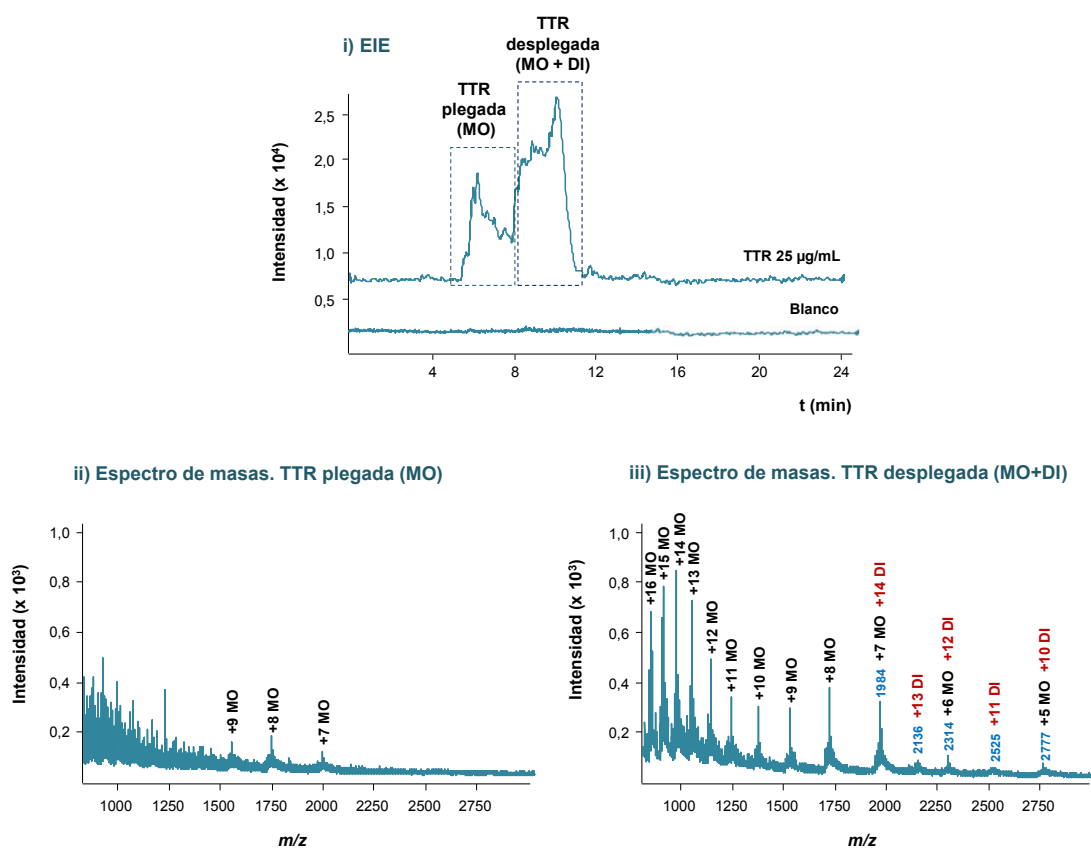


Figura 5.16. EIEs (i) y espectros de masas (ii e iii) obtenidos para un patrón de TTR por (A) CE-MS (50 µg/mL) y (B) Fab'-IA-SPE-CE-MS (25 µg/mL) con un BGE de NH_4Ac 10 mM pH=7,00.

Muchos autores han demostrado la capacidad de la CE para separar confórmers proteicos [240–243]. En este trabajo, el hecho de que los dos confórmers se observen por Fab'-IA-SPE-CE-MS y no por CE-MS empleando el mismo BGE (**Figura 5.16**), podría ser debido a que los

dos confórmers de la TTR interaccionan de manera diferente con el sorbente de IA con fragmentos Fab' de anticuerpo. De esta manera, el primer pico correspondiente al conformero plegado de la TTR, presentaría una menor afinidad por la fase estacionaria y, por consiguiente, se eluiría antes. Por contra, el segundo pico correspondiente al conformero desplegado de la TTR, interaccionaría más con el sorbente de IA y, por tanto, se eluiría más tarde. Es importante destacar que en el anterior trabajo de nuestro grupo de investigación para el análisis de TTR por IA-SPE-CE-MS con un sorbente de IA con anticuerpo intacto [211], la TTR eluía como un único pico correspondiente, por su espectro de masas, con un conformero desplegado (**Figura 5.17**). Esto podría estar relacionado con el uso de diferentes eluyentes, NH_3 100 mM pH=11,30 [211] vs. glicina 100 mM/ NH_3 100 mM pH=9,50 (**artículo 2.4**). Sin embargo, tal y como se ha comentado anteriormente, el eluyente consistente en NH_3 100 mM pH=11,30 no pudo ser utilizado en este trabajo debido a la poca estabilidad del sorbente de IA con los fragmentos Fab' de anticuerpo a pHs tan básicos.

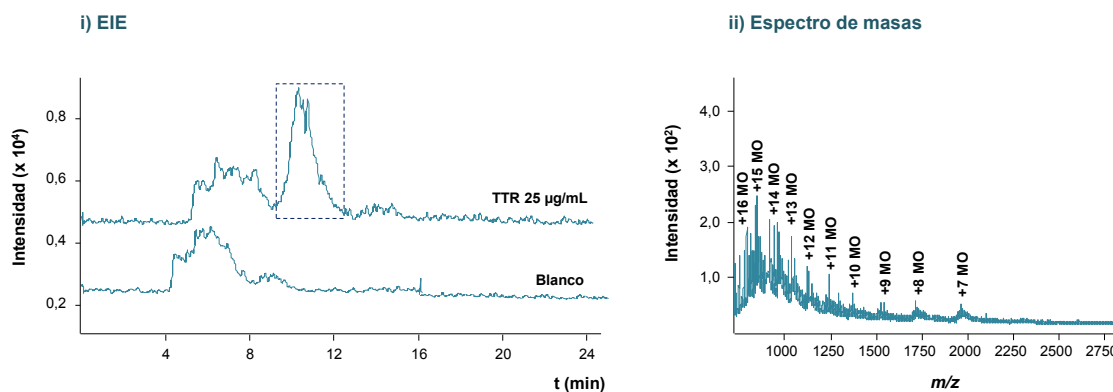


Figura 5.17. EIE (i) y espectro de masas (ii) obtenidos para un patrón de TTR de 25 µg/mL por IA-SPE-CE-MS con un sorbente de IA con anticuerpo intacto (partículas magnéticas UAAF [211]).

La **Tabla 5.5** muestra las M_r teóricas y experimentales, los E_r (en ppm) y las %A para las proteoformas monoméricas de la TTR detectadas con CE-MS (50 µg/mL), IA-SPE-CE-MS con anticuerpo intacto (25 µg/mL) [211] y Fab'-IA-SPE-CE-MS (25 µg/mL). Tal y como se observa, los dos métodos de IA-SPE-CE-MS permitieron detectar tres proteoformas adicionales

que no pudieron ser detectadas por CE-MS para un patrón de TTR de 50 $\mu\text{g/mL}$ (estas proteoformas sí fueron detectadas a 1000 $\mu\text{g/mL}$ empleando un BGE de NH_4Ac 10 mM pH=7,00, **artículo 2.1**), lo que pone de manifiesto la capacidad de preconcentración de ambas técnicas. Las proteoformas detectadas correspondieron a la TTR-Cys, la TTR-Libre, la TTR-Fosforilada o TTR Sulfonada, la TTR- Dihidroxilada o TTR-Sulfinico, la isoforma (10) C-G y las PTMs originadas cuando la cisteína de la posición 10 forma un puente disulfuro con el péptido cisteinil glicina (TTR-CysGly), el péptido glutatión (TTR-Glutatión) y el péptido cisteinil ácido glutámico (TTR-CysGlu). Además de no existir diferencias significativas entre los resultados obtenidos con ambos sorbentes de IA, las %A calculadas para todas las proteoformas monoméricas de la TTR fueron, en general, mayores con IA-SPE-CE-MS (anticuerpo intacto y Fab'), probablemente porque la recuperación de la TTR-Cys, tomada como referencia para calcular las %A, fue más baja que para el resto de proteoformas.

Por lo que hace referencia a los parámetros de calidad, el método de Fab'-IA-SPE-CE-MS proporcionó una repetitividad (%RSD, $n=6$) y reproducibilidad (%RSD, $n=9$) adecuadas para las A_p (entre 8-13% y 9-12%, respectivamente) y los t_m (entre 7-13% y 6-12%, respectivamente). Es importante destacar que, por lo que hace referencia a las A_p , se llevó a cabo una única integración teniendo en cuenta la contribución de los dos conformeros de la TTR (plegado y desplegado). Por lo que respecta a los LODs, se consiguieron unos valores del orden de los 0.5 $\mu\text{g/mL}$, 50 y 2 veces inferiores a los obtenidos con CE-MS ($\sim 25 \mu\text{g/mL}$) y con IA-SPE-CE-MS utilizando un sorbente de IA con anticuerpo intacto ($\sim 1 \mu\text{g/mL}$) [211], respectivamente. El tiempo de vida de los preconcentradores, en todos los casos, fue superior a 20 análisis sin problemas de extracción o inestabilidad en la corriente. Además, a diferencia de lo que ocurría con los microcartuchos preparados con sorbentes de IA con anticuerpo intacto [211], los cuales tenían que ser reemplazados cada día, los preconcentradores con fragmentos Fab' de anticuerpo pudieron ser utilizados durante días consecutivos sin disminuir sus prestaciones (guardados en agua o en BGE durante la noche).

Tabla 5.5. M_r teóricas y experimentales, E_r (en ppm) y %A para las proteoformas del monómero (MO) de la TTR detectadas por (A) CE-MS (50 $\mu\text{g/mL}$), IA-SPE-CE-MS con anticuerpo intacto (25 $\mu\text{g/mL}$) [211] y Fab'-IA-SPE-CE-MS (25 $\mu\text{g/mL}$). Las %A se calcularon normalizando respecto a la proteoforma más abundante (TTR-Cys). En el caso de Fab'-IA-SPE-CE-MS, se llevó a cabo una única integración teniendo en cuenta los dos conformeros de la TTR (plegado y desplegado).

Patrón de TTR			(A) CE-MS (50 $\mu\text{g/mL}$)			(B) IA-SPE-CE-MS (25 $\mu\text{g/mL}$) Anticuerpo intacto			(C) Fab'-IA-SPE-CE-MS (25 $\mu\text{g/mL}$)		
N	Proteoformas MO	M_r teórica	M_r experimental	E_r (ppm)	%A	M_r experimental	E_r (ppm)	%A	M_r experimental	E_r (ppm)	%A
1	TTR-Cys	13880,4022	13880,7000	21	100	13881,0508	47	100	13880,7900	28	100
2	TTR-Libre	13761,2640	13761,1500	8	40	13761,7772	37	62	13761,7500	35	44
3	TTR-Fosforilada	13841,2439	13841,7100	34	39	13841,0479	14	60	13841,6200	27	63
	TTR-Sulfonada	13841,3283		28			20			21	
4	TTR-Dihidroxilada TTR-Sulfinico	13793,2628	13793,7000	32	39	13793,6805	30	53	13793,7500	35	45
5	(10) C-G	13715,1713	13715,3000	9	33	13715,5892	30	62	13715,6000	31	41
6	TTR-CysGly	13937,4590	-	-	-	13937,7082	18	35	13937,9000	32	59
7	TTR-Glutatión	14066,5732	-	-	-	14067,2015	45	48	14066,9500	27	42
8	TTR-CysGlu	14009,5218	-	-	-	14009,3011	16	27	14010,0300	36	36

5.1.4.3. Análisis de TTR en muestras de suero humano por IA-SPE-CE-MS

Una vez optimizada y validada la metodología analítica para el análisis de patrones de TTR mediante Fab'-IA-SPE-CE-MS, se procedió a analizar la TTR en muestras de suero procedentes de controles sanos y de un paciente sintomático de FAP-I. Tal y como ya había pasado en el trabajo de Perú-Gascón et al. con el sorbente de IA con anticuerpo intacto [211], el análisis directo de muestras de suero por Fab'-IA-SPE-CE-MS no fue posible, principalmente debido a la retención no específica en el sorbente y a la precipitación de algunas proteínas mayoritarias del suero (por ejemplo, albúmina) en el capilar de separación. Por este motivo, y con el objetivo de prevenir la saturación de los microcartuchos, se llevó a cabo un sencillo pretratamiento *off-line* consistente en una doble precipitación con un 5% (v/v) de fenol (siguiendo el método optimizado previamente para IA-SPE-CE-MS con anticuerpo intacto [211]). Además, también fue necesario incrementar la fuerza iónica del eluyente pasándose a emplear glicina 200 mM/NH₃ 200 mM (pH=9,50) para conseguir eluir la TTR de una manera reproducible.

Las **Figuras 5.18 A y B** muestran los EIEs (i) y los espectros de masas (ii e iii) obtenidos por Fab'-IA-SPE-CE-MS para muestras de suero procedentes de (A) un paciente sintomático de FAP-I y (B) un control sano. A diferencia de lo ocurrido con el sorbente de IA con anticuerpo intacto [211] (**Figura 5.19**), en ambos casos la TTR eluyó como varios picos con formas diferentes, aunque los espectros de masas solo mostraron diferencias significativas entre el primero y los restantes picos. Tal y como se muestra en los espectros de masas de la TTR para las dos muestras de suero, la presencia de distribuciones de iones amplias centradas en cargas elevadas confirmó únicamente la presencia de conformeros de la TTR con estructuras desplegadas (a diferencia de lo que se observó para muestras patrón de TTR, mirar **Figura 5.16 B**), cada una de ellos interaccionando de manera diferente con el sorbente de IA con fragmentos Fab' de anticuerpo. Es importante tener en cuenta que, una vez desplegada la estructura, la proteína puede agregarse de manera reversible llevando, por consiguiente, a nuevos estados conformacionales desplegados e, incluso, a la oligomerización [240–242]. Este hecho es

especialmente relevante para la TTR, una proteína con una gran tendencia a oligomerizar y a agregarse en ciertos tejidos.

A) Paciente sintomático de FAP-I

B) Control sano

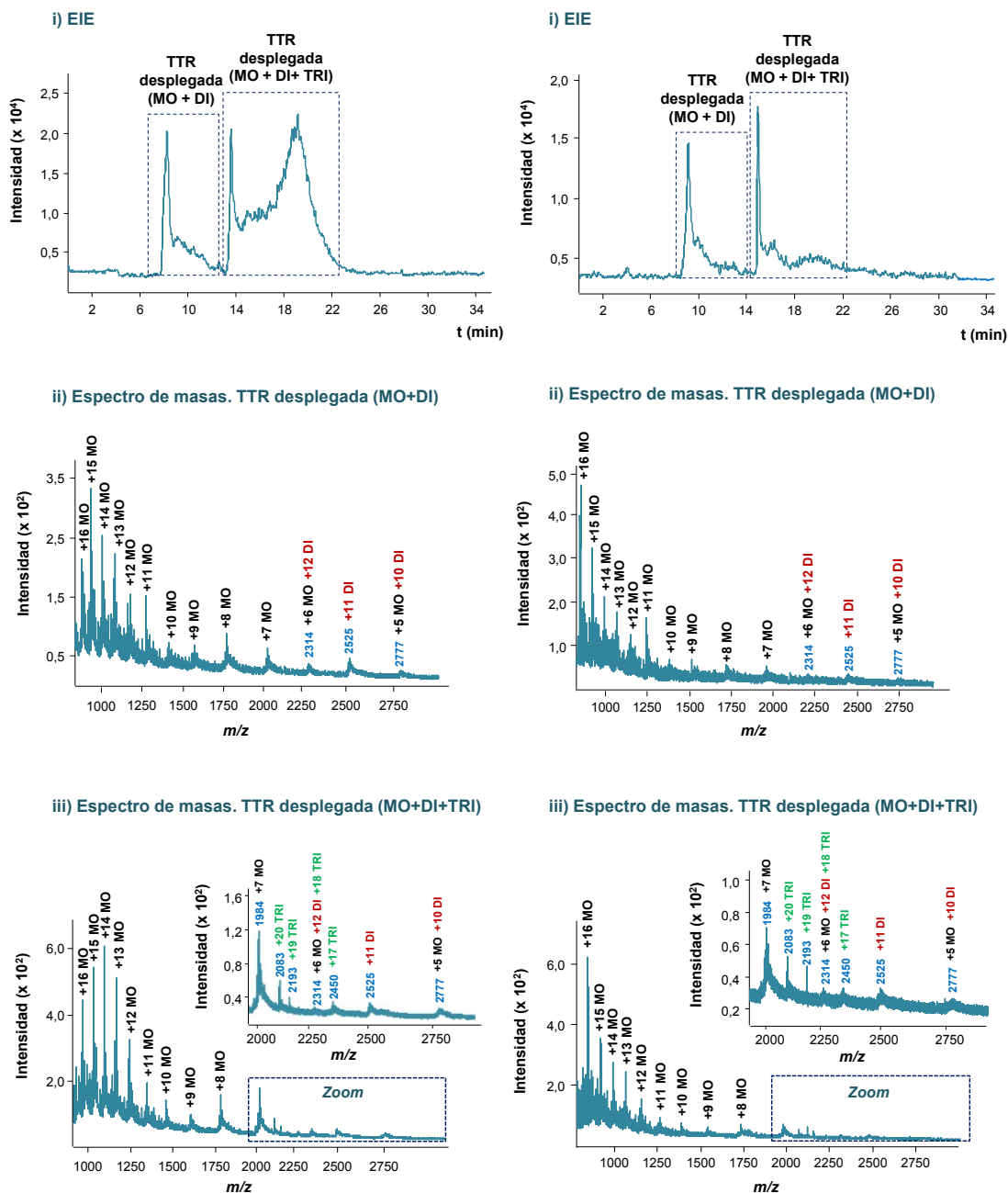


Figura 5.18. EIEs (i) y espectros de masas (ii e iii) obtenidos por Fab'-IA-SPE-CE-MS para (A) una muestra de suero de un paciente sintomático de FAP-I y (B) una muestra de suero control.

Igual que en caso de los patrones de TTR, se llevó a cabo un análisis exhaustivo del clúster de iones multicargados obtenidos en el espectro de masas de los diferentes confórmers para cada una de las muestras, lo que demostró la presencia de MO y DI en el primer confórmer, y la presencia de MO, DI y TRI en los restantes. En el caso del primer confórmer, los iones moleculares con m/z nominales de 2314 y 2777 corresponderían a la suma de la contribución del MO (cargas +6 y +5, respectivamente) y del DI (cargas +12 y +10, respectivamente), mientras que el ión molecular con m/z nominal de 2525 se detectaría exclusivamente para el DI (carga +11) (**Tabla 5.1**, sección 5.1.1 de esta tesis doctoral). En el caso del otro confórmer, el ión molecular con m/z nominal de 2314 correspondería a la suma de la contribución del MO (carga +6), DI (carga +12) y TRI (carga +18), mientras que los iones moleculares con m/z nominales de 2083, 2193 y 2450 se detectarían exclusivamente para el TRI (cargas +20, +19 y +17, respectivamente) (**Tabla 5.1**, sección 5.1.1 de esta tesis doctoral). En las **Figuras 5.18 A y B-i** también se puede observar que la abundancia de uno de los picos del segundo confórmer era mucho menor para la muestra de suero procedente del control sano respecto a la del paciente de FAP-I, sugiriendo, de esta manera, que los perfiles electroforéticos podrían estar relacionados con las diferencias en las proteoformas de la TTR.

En condiciones óptimas para el análisis de muestras de suero, la repetitividad (%RSD, $n=6$) fue adecuada para las A_p (entre 15% y 12%) y los t_m (entre 12 y 14%) para todas las proteoformas monoméricas de la TTR detectadas en muestras de suero control y de pacientes de FAP-I. De nuevo, por lo que hace referencia a las A_p , se llevó a cabo una única integración teniendo en cuenta la contribución de todos los confórmers desplegados de la TTR. El tiempo de vida de los preconcentradores, en todos los casos, fue inferior a los 20 análisis conseguidos con los patrones de TTR, resultados similares a los obtenidos por IA-SPE-CE-MS con anticuerpo intacto [211].

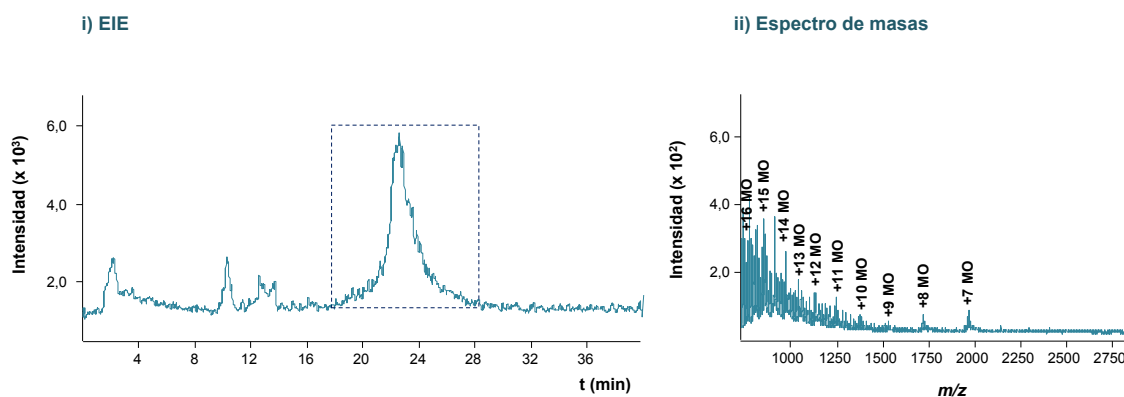


Figura 5.19. EIE (i) y espectros de masas (ii) obtenidos por IA-SPE-CE-MS con un sorbente de IA con anticuerpo intacto (partículas magnéticas UAAF [211]) para una muestra de suero control.

La **Tabla 5.6** muestra las M_r teóricas y experimentales, los E_r (en ppm) y las %A para las proteoformas monoméricas de la TTR detectadas por (A) IA-SPE-CE-MS con anticuerpo intacto y (B) Fab'-IA-SPE-CE-MS para muestras de suero procedentes de un control sano y de un paciente sintomático de FAP-I. A diferencia de lo observado en el anterior trabajo de nuestro grupo de investigación con el sorbente de IA con anticuerpo intacto [211], en el cual sólo se pudieron detectar cinco de las proteoformas monoméricas más abundantes de la TTR normal (TTR-Cys, TTR-Libre, TTR-Fosforilada o TTR-Sulfonada, TTR-Dihidroxilada o TTR-Sulfínico y (10) C-G), el método de Fab'-IA-SPE-CE-MS desarrollado en el cuarto trabajo de esta tesis doctoral (**artículo 2.4**), permitió detectar tres proteoformas adicionales, correspondientes a la TTR-CysGly, la TTR-Glutatión y la TTR-CysGlu. En este sentido, es importante resaltar que de todos los métodos de purificación de TTR descritos hasta ahora (IP *off-line* convencional en solución (**artículo 2.1**), IP *off-line* con partículas magnéticas (**artículo 2.2**), IA-SPE-CE-MS con un sorbente de IA con anticuerpo intacto [211] y Fab'-IA-SPCE-CE-MS (**artículo 2.4**)), el método de Fab'-IA-SPE-CE-MS es el único que permitió detectar, en suero humano, las mismas proteoformas monoméricas de la TTR que en un patrón de 1000 $\mu\text{g/mL}$, poniendo de manifiesto su gran capacidad de purificación y preconcentración. En esta tabla también se puede observar que ambos métodos de IA-SPE-CE-MS permitieron detectar

las proteoformas asociadas a la TTR mutante (TTR(Met30)-Cys y TTR(Met30)-Libre) en los pacientes sintomáticos de FAP-I. Sin embargo, tal y como se apuntó para el método *off-line* de IP con partículas magnéticas (**artículo 2.2** de esta tesis doctoral), el E_r no permitió diferenciar entre la TTR(Met30)-Libre y la TTR-Dihidroxilada o TTR-Sulfinico (M_r similares). Por este motivo, sólo la detección de la proteoforma TTR(Met30)-Cys permitiría confirmar de manera inequívoca la presencia de la TTR(Met30) en pacientes de FAP-I. La exactitud y resolución del espectrómetro de masas empleado tampoco fueron suficientes para identificar inequívocamente las proteoformas relacionadas con los diferentes conformeros. Así, para poder interpretar las diferencias asociadas a los perfiles electroforéticos obtenidos por Fab'-IA-SPE-CE-MS para los controles y los pacientes de FAP-I, sería necesario analizar un mayor número de muestras y con un instrumento de mejores prestaciones, que permitiera incluso realizar experimentos de masas en tándem (MS/MS) para caracterizar en profundidad e inequívocamente las diferentes proteoformas.

Tabla 5.6. M_r teóricas y experimentales, E_r (en ppm) y %A para las proteoformas del monómero (MO) de la TTR detectadas por (A) IA-SPE-CE-MS con anticuerpo intacto [211] y (B) Fab'-IA-SPE-CE-MS para una muestra de suero control y de un paciente sintomático de FAP-I. Las %A se calcularon normalizando respecto a la proteoforma más abundante (TTR-Cys). En el caso de Fab'-IA-SPE-CE-MS, se llevó a cabo una única integración teniendo en cuenta los diferentes conformeros desplegados de la TTR.

Muestras de suero			(A) IA-SPE-CE-MS Anticuerpo intacto						(B) Fab'-IA-SPE-CE-MS					
			Control sano			Paciente sintomático FAP-I			Control sano			Paciente sintomático FAP-I		
N	Proteoformas MO	M_r teórica	M_r experimental	E_r (ppm)	%A	M_r experimental	E_r (ppm)	%A	M_r experimental	E_r (ppm)	%A	M_r experimental	E_r (ppm)	%A
1	TTR-Cys	13880,4022	13880,9511	40	100	13881,1016	50	100	13880,6600	19	100	13880,8800	34	100
1'	TTR(Met30)-Cys	13912,4683	-	-	-	13912,9765	37	44	-	-	-	13912,6500	13	74
2	TTR-Libre	13761,2640	13761,9247	48	64	13761,7566	36	61	13760,8300	32	36	13760,9300	24	70
3	TTR-Fosforilada	13841,2439	13841,7207	34	69	13841,0863	11	67	13841,8000	40	42	13841,3500	8	74
	TTR-Sulfonada	13841,3283		28			17			34			2	
4	TTR-Dihidroxilada	13793,2628	13793,9606	51	51	13793,6805	30	64	13794,0100	54	32	13793,7500	35	48
2'	TTR(Met30)-Libre	13793,3301	-	-	-		25		-	-	-		30	
5	(10) C-G	13715,1713	13715,4042	17	53	13715,4228	18	47	13714,7800	29	37	13715,2000	2	50
6	TTR-CysGly	13937,4590	-	-	-	-	-	-	13936,9500	37	52	13937,6000	10	65
7	TTR-Glutación	14066,5732	-	-	-	-	-	-	14067,1000	37	40	14067,0500	34	56
8	TTR-CysGlu	14009,5218	-	-	-	-	-	-	14010,0500	38	35	14010,0100	35	49

5.2. Análisis dirigido de biomarcadores proteicos. Análisis *Bottom-up* del antígeno carcinoembrionario humano

El análisis de proteínas intactas (*top-down*) por CE-MS, tal y como se muestra en el capítulo anterior para la TTR, es una buena alternativa para la detección de un gran número de proteínas, así como para la caracterización de sus diferentes isoformas y PTMs. Sin embargo, esta estrategia presenta algunas limitaciones debidas, principalmente, a la dificultad en la ionización de las proteínas de elevada masa molecular, tales como glicoproteínas con un alto grado de glicosilación, y a la complejidad estructural intrínseca que genera una mezcla tan compleja de proteoformas. En este sentido, el análisis *bottom-up* permite simplificar el problema obteniendo información estructural a partir de los péptidos resultantes de digerir enzimáticamente la proteína intacta. Paralelamente, otra alternativa para mejorar la sensibilidad de los análisis por CE-MS consiste en utilizar una interfase sin líquido auxiliar (*sheathless*), en contraposición a la interfase con líquido auxiliar (*sheath flow*) utilizada en el capítulo anterior. Esta interfase, tal y como se describe en la sección 1.4.3.1 de la introducción, permite mejorar la sensibilidad y los LODs por CE-MS, ya que evita la dilución de los compuestos de interés a la salida del capilar de separación, antes de la ionización por ESI.

Este capítulo se centra en la identificación y caracterización de los *N*-glicopéptidos del antígeno carcinoembrionario humano (CEA) mediante electroforesis capilar acoplada a la espectrometría de masas en tándem con interfase *sheathless* (*sheathless* CE-MS/MS). CEA es una glicoproteína que se produce durante el desarrollo fetal y generalmente no es detectable en la sangre de personas sanas adultas. Sin embargo, la elevación de sus niveles en sangre se ha relacionado con procesos tumorales, especialmente con el cáncer de colon y la metástasis de hígado (procedente de cáncer de colon). Cabe resaltar que, más allá de su abundancia en sangre, hoy en día, no se obtiene otra información estructural de esta glicoproteína que permita relacionarla con los

diferentes estados de evolución del cáncer, sus tipos, su pronóstico o la eficacia de los tratamientos.

Con el objetivo de profundizar en el estudio de esta proteína como biomarcadora de cáncer, en el quinto trabajo de esta tesis doctoral (**artículo 3.1**), se ha desarrollado una metodología analítica mediante *sheathless* CE-MS/MS para la identificación de sus diferentes *N*-glicopéptidos, obtenidos tras digestión enzimática. Cabe destacar que se han descrito 28 puntos de *N*-glicosilación en esta glicoproteína, por lo que en este estudio se ha pretendido la identificación del mayor número posible de éstos, así como la determinación de su grado de ocupación. Para ello, ha sido necesario optimizar la digestión enzimática empleando diferentes tipos de enzimas, tanto específicas (tripsina y endoproteinasa Glu-C) como no específicas (pronasa). Finalmente, se ha aplicado la metodología óptima al análisis de tres muestras comerciales de CEA, dos de ellas procedentes de pacientes con cáncer de colon y una con metástasis de hígado. El posterior tratamiento de los datos ha permitido la identificación de 219 glicoformas de *N*-glicopéptidos (que a partir de ahora se referirán simplemente como *N*-glicopéptidos) de CEA correspondientes a 20 puntos de *N*-glicosilación diferentes. Además, se ha podido observar que los perfiles de *N*-glicosilación de esta proteína son diferentes en el cáncer de colon y la metástasis de hígado (procedente de cáncer de colon), por lo que la metodología desarrollada podría ser de utilidad para explicar el proceso de metástasis en pacientes con cáncer de colon.

5.2.1. Análisis de los *N*-glicopéptidos de CEA mediante *sheathless* CE-MS/MS

Existen diferentes tipos de enzimas para la hidrólisis *in vitro* de los enlaces peptídicos de las proteínas con el objetivo de generar péptidos de menor tamaño, y la combinación de enzimas específicas y no específicas puede ayudar a maximizar la cobertura (*coverage*) de los puntos de *N*-glicosilación y mejorar la caracterización de las diferentes proteoformas [112,113,244]. La tripsina y la endoproteinasa Glu-C son las enzimas más utilizadas. La tripsina corta

específicamente en el extremo carboxilo terminal de los residuos de lisina (K) y arginina (R), mientras que la Glu-C corta en el extremo terminal de los residuos de ácido glutámico (E) y ácido aspártico (D). Aunque estas enzimas proporcionan información suficiente para caracterizar un gran número de carbohidratos, glicanos y puntos de glicosilación en diferentes glicoproteínas, en algunos casos, su eficacia a la hora de cubrir toda la secuencia aminoacídica es limitada, especialmente debido a la generación de péptidos con un gran número de aminoácidos, lo que puede resultar en la supresión de la ionización de los *N*-glicopéptidos de interés [244]. En este sentido, la pronasa es una mezcla de enzimas proteolíticas no específicas que permite superar algunos de los inconvenientes descritos anteriormente para la tripsina y la Glu-C. La pronasa permite obtener péptidos con un número pequeño de aminoácidos (generalmente entre 1 y 8), evitando de esta manera la supresión de la ionización de los *N*-glicopéptidos en estudio. Sin embargo, la dificultad en la caracterización de los digestos aumenta a medida que se reduce el número de aminoácidos en los péptidos resultantes, a la que se añade la ambigüedad de los cortes realizados [245,246].

El **artículo 3.1** muestra como resulta conveniente combinar la información obtenida tras digerir las diferentes muestras comerciales de CEA con tripsina, Glu-C y pronasa para conseguir la máxima cobertura de la secuencia de la glicoproteína. De esta manera es posible identificar el mayor número posible de puntos de *N*-glicosilación, así como determinar su grado de ocupación.

5.2.1.1. Interpretación de los electroferogramas y los espectros de MS/MS

Tal y como se ha comentado anteriormente, las muestras de CEA digeridas se analizaron mediante *sheathless* CE-MS/MS. Es importante destacar que la utilización de la espectrometría de MS/MS, a diferencia de lo ocurrido en el capítulo anterior para el análisis de TTR, permitió obtener información estructural a partir de los fragmentos característicos de los diferentes glicopéptidos generados tras la digestión enzimática. Los espectros de MS/MS se adquirieron en

modo *data dependent*, donde se fragmentaron todos los iones con intensidades por encima de un determinado umbral (*threshold*). Con el objetivo de encontrar la región de los electroferogramas de MS/MS en la que migraban los *N*-glicopéptidos, se inspeccionó cuidadosamente la presencia de éstos mediante la obtención del EIE (MS/MS) de *m/z* 366,14, correspondiente a la liberación del fragmento HN de los diferentes *N*-glicopéptidos (H, manosa y N, N-acetilglucosamina, **Tabla 1.3** de la introducción) (mirar **Figura 5.20 A, B y C-ii**).

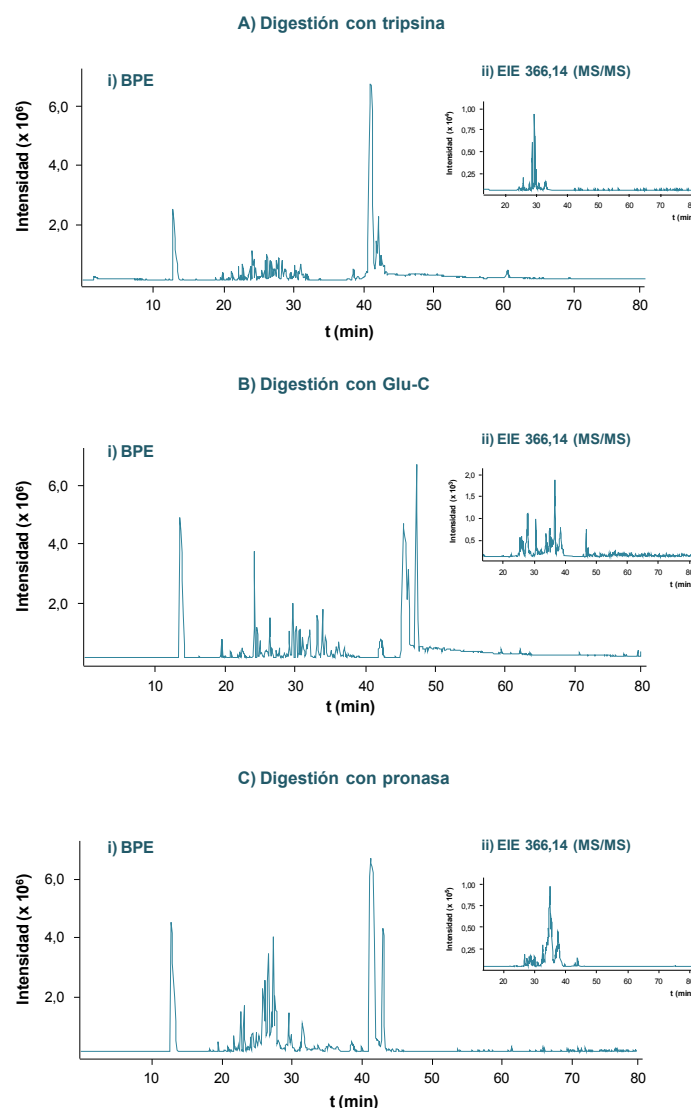


Figura 5.20. Electroferogramas de pico base (BPEs) (i) y EIEs (MS/MS) de *m/z* nominal 366 (ii) obtenidos por *sheathless* CE-MS/MS después de digerir una muestra de CEA de cáncer de colon (CEA2) con (A) tripsina, (B) Glu-C y (C) pronasa.

Tal y como se puede observar en la **Figura 5.20 A, B y C-ii**, el perfil de los electroferogramas de MS/MS indicó la presencia de *N*-glicopéptidos entre 25-32 min (tripsina), 25-49 min (Glu-C) y 28-45 min (pronasa). Una vez obtenido este perfil, se inspeccionaron los espectros de MS/MS con el objetivo de encontrar pérdidas neutras correspondientes a una H (manosa, $\Delta M_r=162,0528$), una N (N-acetilglucosamina, $\Delta M_r=203,0794$), una F (fucosa, $\Delta M_r=146,0579$), un S (ácido siálico, $\Delta M_r=291,0954$) o combinaciones de estas estructuras. La estructura de los diferentes *N*-glicopéptidos se determinó teniendo en cuenta las adiciones sucesivas de diferentes unidades de H, N, F o S a la m/z experimental correspondiente a los péptidos generados tras la digestión enzimática (**Figura 5.21**). Es conveniente destacar que, en todos los casos, se consideró que la m/z experimental correspondía a la del péptido en cuestión si la exactitud de masas (en términos de error relativo, E_r) era igual o inferior a 10 ppm. Como ejemplo, la **Figura 5.21** muestra la fragmentación obtenida por MS/MS para el N_{580} -glicopéptido H3N4F1 después de digerir una muestra de CEA de cáncer de colon (CEA1) con Glu-C.

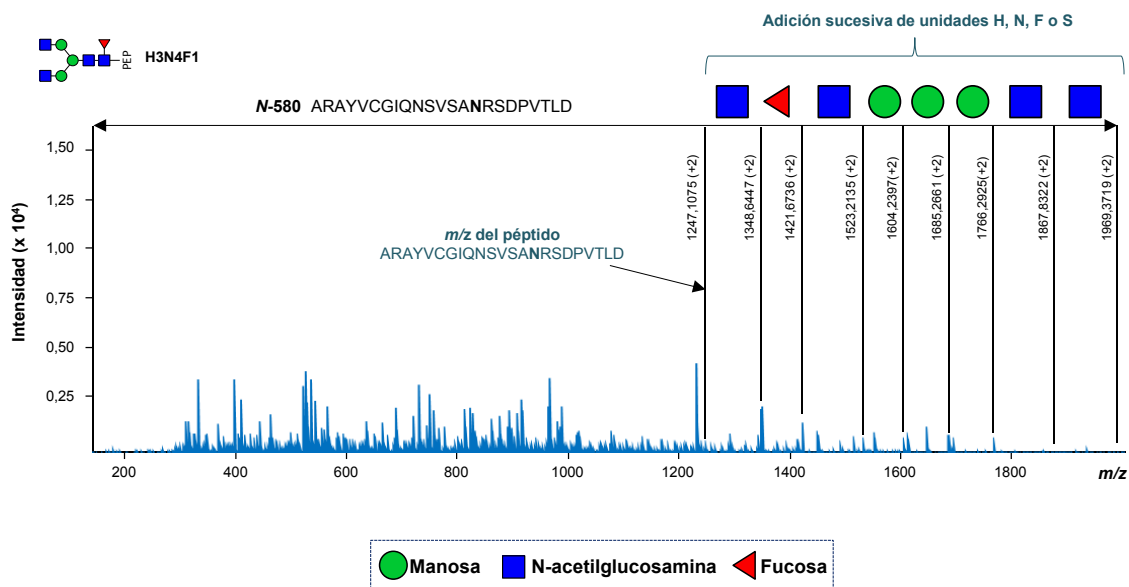


Figura 5.21. Fragmentación obtenida por MS/MS para el N_{580} -glicopéptido H3N4F1 después de digerir con Glu-C una muestra de CEA de cáncer de colon (CEA1). La estructura de este *N*-glicopéptido se determinó por adiciones sucesivas de unidades de H, N, F o S a la m/z experimental correspondiente al péptido generado (ARAYVCGIQNSVSA NRSDPVTLD, m/z 1247,1075 (carga +2)).

En el caso de la pronasa y, debido a la ambigüedad en la generación de digestos con un número pequeño de aminoácidos, fue necesario utilizar el programa *Findpept tool* (www.expasy.org/tools/findpept.html) con el objetivo de encontrar las m/z teóricas (de los péptidos generados) correspondientes a las m/z experimentales deducidas a partir de los espectros de MS/MS.

5.2.1.2. Identificación de los *N*-glicopéptidos de CEA

La **Tabla 5.7** muestra los puntos de *N*-glicosilación (y sus correspondientes *N*-glicopéptidos) detectados por *sheathless* CE-MS/MS después de digerir diferentes muestras comerciales de CEA, dos de ellas procedentes de pacientes con cáncer de colon (CEA1 y CEA2) y una de pacientes con metástasis de hígado (CEA3), con (A) tripsina, (B) Glu-C y (C) pronasa.

Tripsina

Tal y como se puede observar en la **Tabla 5.7 A** para la digestión con tripsina, esta enzima permitió detectar un total de 128 *N*-glicopéptidos correspondientes a 9 puntos de *N*-glicosilación (*N*-197/*N*-553, *N*-204/*N*-560, *N*-375, *N*-580, *N*-612, *N*-650 y *N*-665) para los tres tipos de muestras de CEA analizadas. Es importante destacar que no se pudo diferenciar entre los *N*-glicopéptidos pertenecientes a los puntos de glicosilación *N*-197 y *N*-553, y *N*-240 y *N*-560, ya que se obtuvieron las mismas secuencias peptídicas tras digerir las muestras con tripsina (LQLSNGN_{197/553}R y TLTLFN_{204/560}VTR, respectivamente). En general, se observaron similitudes en la composición de los glicanos para las dos muestras de cáncer de colon (CEA1 y CEA2), mientras que se hicieron evidentes algunas diferencias para la muestra de metástasis de hígado (CEA3). El mayor número de *N*-glicopéptidos fue detectado en el punto de glicosilación *N*-197/*N*-553 (52 glicopéptidos, si tenemos en cuenta que este punto de *N*-glicosilación puede estar deamidado), seguido del *N*-204/*N*-560 (40 glicopéptidos), y el *N*-665 (18 glicopéptidos). Para el resto de puntos de glicosilación, *N*-650, *N*-375, *N*-612 y *N*-580, se detectaron 11, 4, 2 y 1

glicopéptidos, respectivamente. Se caracterizaron los diferentes tipos de glicanos para cada punto de *N*-glicosilación, incluyendo glicanos ricos en manosa, híbridos y, especialmente, complejos (con 2, 3 y 4 antenas). Los glicanos altamente fucosilados (entre 3 y 7 fucosas) y sialilados (entre 1 y 2 ácidos siálicos) se encontraron predominantemente en los puntos de glicosilación *N*-204/*N*-560, *N*-580, *N*-612, *N*-650 y *N*-665, especialmente para la muestra de CEA de metástasis de hígado (CEA3). Por lo que hace referencia a los puntos de *N*-glicosilación no detectados, este hecho puede ser debido a la generación de péptidos con un gran número de aminoácidos, por lo general superior a 20 (mirar la longitud de las secuencias peptídicas generadas en los diferentes *N*-glicopéptidos tras digerir las muestras con tripsina en la **Tabla 5.7**), lo que podría dificultar la ionización de los *N*-glicopéptidos en estudio.

Endoproteinasa Glu-C

Tal y como se puede observar en la **Tabla 5.7 B**, la digestión con Glu-C permitió detectar un total de 50 *N*-glicopéptidos correspondientes a 4 puntos de *N*-glicosilación (*N*-152/*N*-508, *N*-466 y *N*-580) para los tres tipos de muestras de CEA analizadas. Ninguno de estos puntos de *N*-glicosilación fue identificado tras digerir las muestras de CEA con tripsina, con excepción del punto de glicosilación *N*-580 (9 vs. 1 *N*-glicopéptidos identificados con Glu-C y tripsina, respectivamente). En este caso, no se pudo diferenciar entre los *N*-glicopéptidos pertenecientes a los puntos de glicosilación *N*-152 y *N*-508, ya que se obtuvo la misma secuencia peptídica tras digerir las muestras de CEA con Glu-C (LPKPSISSN_{152/508}NSKPVE). De nuevo, en general, se observaron similitudes en la composición de los glicanos para las dos muestras de CEA de cáncer de colon (CEA1 y CEA2), mientras que la muestra de metástasis de hígado (CEA3) presentó algunas diferencias notables por lo que hace referencia a la estructura de los *N*-glicopéptidos identificados. El mayor número de *N*-glicopéptidos fue detectado en el punto de glicosilación *N*-466 (21 glicopéptidos), seguido del *N*-152/*N*-508 (20 glicopéptidos), y el *N*-580 (9 glicopéptidos). Se caracterizaron diferentes tipos de glicanos para cada punto de *N*-glicosilación, incluyendo glicanos ricos en manosa en el punto de glicosilación *N*-152/*N*-508, y

glicanos complejos (con 2, 3 y 4 antenas) en los puntos de glicosilación *N*-466 y *N*-580. Tal y como se puede observar en esta tabla, la Glu-C permitió identificar un menor número de puntos de *N*-glicosilación y *N*-glicopéptidos que la tripsina, aunque la información obtenida con ambas enzimas es complementaria. Este hecho es debido a su diferente especificidad y a la generación, con Glu-C, de digestos con un mayor número de aminoácidos (comparar la longitud de las secuencias peptídicas generadas tras digerir las muestras con tripsina y Glu-C en la **Tabla 5.7 A** y **B**). En este sentido, a diferencia de lo que ocurría con tripsina, conviene resaltar que la digestión enzimática con Glu-C generó secuencias peptídicas correspondientes a la ausencia de algunos de los cortes esperados (*missed cleavages*, MCs), generando, de esta manera, secuencias peptídicas más largas. Todos los péptidos correspondientes a 0, 1, 2 y 3 MCs se consideraron, tanto para la caracterización estructural como para la interpretación y comparación de los datos obtenidos (**Tabla 5.7 B**).

Pronasa

Tal y como se puede observar en la **Tabla 5.7 C** para la digestión con pronasa, esta enzima permitió detectar un total de 41 *N*-glicopéptidos correspondientes a 11 puntos de *N*-glicosilación (*N*-104, *N*-208, *N*-256, *N*-274, *N*-330, *N*-351, *N*-375, *N*-432, *N*-29, *N*-612 y *N*-665) para los tres tipos de muestras de CEA analizadas. Ninguno de estos puntos de *N*-glicosilación fue identificado tras digerir las muestras de CEA con tripsina y Glu-C, con excepción de los puntos de glicosilación *N*-375 (1 vs. 4 *N*-glicopéptidos identificados con pronasa y tripsina, respectivamente), *N*-612 (6 (pronasa) vs. 2 *N*-glicopéptidos (tripsina)) y *N*-665 (3 (pronasa) vs. 18 *N*-glicopéptidos (tripsina)). Como en los casos anteriores, en general, se observaron similitudes en la composición de los glicanos para las dos muestras de cáncer de colon (CEA1 y CEA2), mientras que se hicieron evidentes algunas diferencias para la muestra de metástasis de hígado (CEA3). El mayor número de *N*-glicopéptidos fue detectado en el punto de glicosilación *N*-529 (10 glicopéptidos), seguido del *N*-274 (7 glicopéptidos) y el *N*-612 (6 glicopéptidos). Para el resto de puntos de *N*-glicosilación, se detectaron entre 1 y 5 glicopéptidos. Es importante

destacar que la digestión con pronasa permitió identificar un mayor número de puntos de *N*-glicosilación, probablemente debido a la generación de digestos con un menor número de aminoácidos (entre 1 y 8), evitando, de esta manera, la supresión de la ionización de los *N*-glicopéptidos en estudio. Sin embargo, tal y como se ha comentado anteriormente, la dificultad en la caracterización de los digestos aumenta a medida que se reduce el tamaño de los péptidos generados y la especificidad de la enzima. Se caracterizaron diferentes tipos de glicanos para cada punto de *N*-glicosilación, incluyendo glicanos ricos en manosa, híbridos y, especialmente, complejos (con 2, 3 y 4 antenas). Los glicanos altamente fucosilados (entre 3 y 7 fucosas) y sialilados (entre 1 y 2 ácidos siálicos) se identificaron predominantemente en los puntos de glicosilación *N*-274 y *N*-665, especialmente para la muestra de CEA de metástasis de hígado (CEA3).

Tabla 5.7. Puntos de *N*-glicosilación (y sus correspondientes *N*-glicopéptidos) identificados por *sheathless* CE-MS/MS (*n*=6) tras digerir las muestras de CEA de pacientes con cáncer de colon (CEA1 y CEA2) y metástasis de hígado (CEA3) con (A) tripsina, (B) endoproteinasa Glu-C y (C) pronasa. BGE empleado: HAc 10% (v/v) pH=2,30. (aa=aminoácido).

			Número total de <i>N</i> -glicopéptidos=128			
A) Tripsina	Puntos de <i>N</i> -glicosilación	Número de aa	CEA1 Cáncer de colon	CEA2 Cáncer de colon	CEA3 Metástasis de hígado	Total
	<i>N</i> -104+ <i>N</i> -115 ^a	28	-	-	-	-
	<i>N</i> -152	21	-	-	-	-
	<i>N</i> -182	30	-	-	-	-
	<i>N</i> -197/ <i>N</i> -553 ^b	8	25	25	23	26
	<i>N</i> -197/ <i>N</i> -553-De ^c	8	26	26	21	26
	<i>N</i> -204/ <i>N</i> -560 ^b	9	35	34	16	40
	<i>N</i> -208	7	-	-	-	-
	<i>N</i> -246	25	-	-	-	-
	<i>N</i> -256+ <i>N</i> -274+ <i>N</i> -288+ <i>N</i> -292+ <i>N</i> -309 ^a	60	-	-	-	-
	<i>N</i> -330+ <i>N</i> -351+ <i>N</i> -360 ^a	58	-	-	-	-
	<i>N</i> -375	8	4	4	2	4
	<i>N</i> -432+ <i>N</i> -466 ^a	72	-	-	-	-
	<i>N</i> -480	18	-	-	-	-
	<i>N</i> -508	24	-	-	-	-
	<i>N</i> -529	30	-	-	-	-
	<i>N</i> -580	14	-	-	1	1
	<i>N</i> -612	47	-	-	2	2
	<i>N</i> -650	20	-	-	11	11
	<i>N</i> -665	6	11	4	18	18

			Número total de <i>N</i> -glicopéptidos=50			
B) Glu-C ^d	Puntos de <i>N</i> -glicosilación	Número de aa	CEA1 Cáncer de colon	CEA2 Cáncer de colon	CEA3 Metástasis de hígado	Total
	<i>N</i> -104+ <i>N</i> -115 ^a	17	-	-	-	-
	<i>N</i> -152/ <i>N</i> -508 ^b	15	20	19	18	20
	<i>N</i> -182+ <i>N</i> -197+ <i>N</i> -204+ <i>N</i> -208 ^a	36	-	-	-	-
	<i>N</i> -246	15	-	-	-	-
	<i>N</i> -256+ <i>N</i> -274 ^a	30	-	-	-	-
	<i>N</i> -288+ <i>N</i> -292 ^a	22	-	-	-	-
	<i>N</i> -309+ <i>N</i> -330 ^a	31	-	-	-	-
	<i>N</i> -351+ <i>N</i> -360 ^a	26	-	-	-	-
	<i>N</i> -375	13	-	-	-	-
	<i>N</i> -432	36	-	-	-	-
	<i>N</i> -466	8	21	21	9	21
	<i>N</i> -480	30	-	-	-	-
	<i>N</i> -529+ <i>N</i> -553+ <i>N</i> -560 ^a	39	-	-	-	-
	<i>N</i> -580	23	9	9	-	9
	<i>N</i> -612/ <i>N</i> -650/ <i>N</i> -665 ^a	83	-	-	-	-
			Número total de <i>N</i> -glicopéptidos=41			
C) Pronasa ^e	Puntos de <i>N</i> -glicosilación	Número de aa	CEA1 Cáncer de colon	CEA2 Cáncer de colon	CEA3 Metástasis de hígado	Total
	<i>N</i> -104	4	1	1	-	1
	<i>N</i> -115	-	-	-	-	-
	<i>N</i> -152	-	-	-	-	-
	<i>N</i> -182	-	-	-	-	-
	<i>N</i> -197	-	-	-	-	-
	<i>N</i> -204	-	-	-	-	-
	<i>N</i> -208	4	-	-	1	1
	<i>N</i> -246	-	-	-	-	-

			Número total de <i>N</i> -glicopéptidos=41			
C) Pronasa ^c	Puntos de <i>N</i> -glicosilación	Número de aa	CEA1 Cáncer de colon	CEA2 Cáncer de colon	CEA3 Metástasis de hígado	Total
	<i>N</i> -256	7	3	3	4	5
	<i>N</i> -274	2-4 ^f	5	5	7	7
	<i>N</i> -288	-	-	-	-	-
	<i>N</i> -292	-	-	-	-	-
	<i>N</i> -309	-	-	-	-	-
	<i>N</i> -330	5	1	1	1	1
	<i>N</i> -351	7	5	5	5	5
	<i>N</i> -360	-	-	-	-	-
	<i>N</i> -375	4	1	1	1	1
	<i>N</i> -432	5	1	1	-	1
	<i>N</i> -466	-	-	-	-	-
	<i>N</i> -480	-	-	-	-	-
	<i>N</i> -508	-	-	-	-	-
	<i>N</i> -529	4	7	7	8	10
	<i>N</i> -553	-	-	-	-	-
	<i>N</i> -560	-	-	-	-	-
	<i>N</i> -580	-	-	-	-	-
	<i>N</i> -612	6	3	3	4	6
	<i>N</i> -650	-	-	-	-	-
	<i>N</i> -665	3	2	2	3	3

^aExisten diferentes puntos de *N*-glicosilación en la misma secuencia peptídica. ^bNo se pueden diferenciar debido a que presentan la misma secuencia peptídica. ^cEl punto de glicosilación *N*-197/*N*-553 también se encuentra deamidado. ^dSe utilizan todos los *missed cleavages* (de 0 a 3) para la interpretación estructural. ^eComo la pronasa es una enzima no específica, sólo se puede indicar el número de aa para los puntos de *N*-glicosilación identificados. ^fSe identifican dos secuencias peptídicas para este punto de *N*-glicosilación (NG y NGTF).

5.2.1.3. Modelos semiempíricos de predicción de la migración electroforética

El uso de modelos semiempíricos de predicción de la migración electroforética es una buena herramienta complementaria para confirmar la interpretación estructural llevada a cabo a partir de los espectros de MS/MS. Estos modelos han sido investigados por diversos autores para predecir la migración electroforética de diferentes péptidos y glicopéptidos teniendo en cuenta su M_r , su carga (q) y su movilidad electroforética (μ_e). En particular, en un trabajo previo de nuestro grupo de investigación [247], se demostró que el modelo clásico de los polímeros ($\mu_e \propto q/M_r^{1/2}$) es el más adecuado para predecir la migración electroforética de péptidos y *N*-glicopéptidos. Es importante destacar que, mientras que el número y tipo de glicanos unidos a la cadena peptídica afecta al tamaño de los glicopéptidos y, por consiguiente, a su M_r , la presencia de ácidos siálicos, los cuales están cargados negativamente al pH de trabajo (BGE de HAc 10% (v/v), pH=2,30), influye significativamente en su carga.

La **Figura 5.22** muestra los EIEs, la M_r , la q y $q/M_r^{1/2}$ para los 11 *N*-glicopéptidos identificados en el punto de glicosilación *N*-650 tras digerir una muestra de CEA de metástasis de hígado (CEA3) con tripsina. Tal y como se puede observar en esta figura y en la tabla que le acompaña, los *N*-glicopéptidos sin ácidos siálicos migraron a un tiempo inferior (33-35 min). Sin embargo, los *N*-glicopéptidos con ácidos siálicos aparecieron a tiempos de migración superiores ya que éstos presentan carga negativa a pH=2,30 y, por consiguiente, la carga molecular global positiva disminuye. En este sentido, los *N*-glicopéptidos con ácidos siálicos migraron en dos ventanas de tiempo correspondientes a los *N*-glicopéptidos con uno y dos ácidos siálicos (37-39 min y 42-44 min, respectivamente). La correlación lineal entre los tiempos de migración experimentales y su $q/M_r^{1/2}$ teórica fue excelente ($R^2 > 0,97$ en este caso), lo que permitió obtener una confirmación adicional de la interpretación estructural llevada a cabo a partir de los espectros de MS/MS.

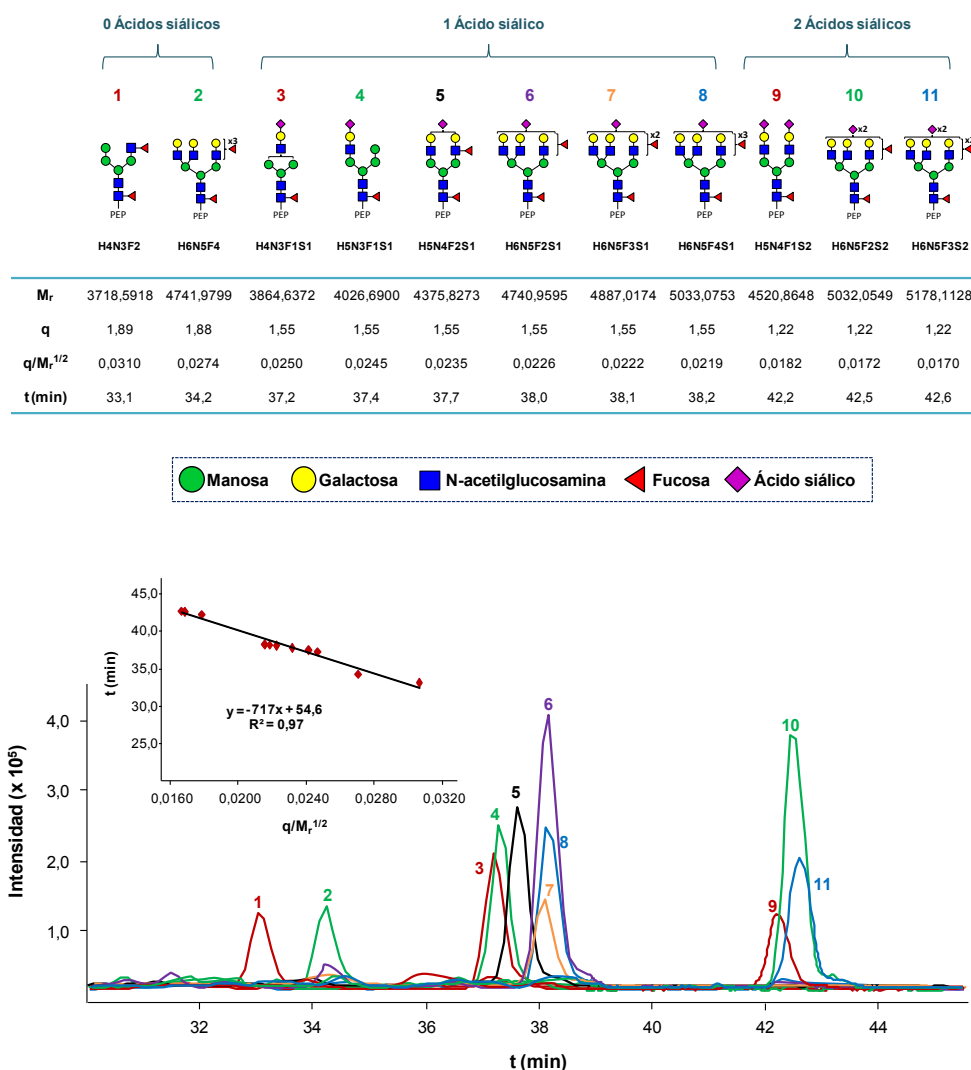


Figura 5.22. EIEs, M_r , q y $q/M_r^{1/2}$ para los 11 *N*-glicopéptidos identificados por *sheathless* CE-MS/MS en el punto de glicosilación *N*-650 tras digerir una muestra de CEA de metástasis de hígado (CEA3) con tripsina. (1: H4N3F2, 2: H6N5F4, 3: H4N3F1S1, 4: H5N3F1S1, 5: H5N4F2S1, 6: H6N5F2S1, 7: H6N5F3S1, 8: H6N5F4S1, 9: H5N4F1S2, 10: H6N5F2S2, 11: H6N5F3S2).

5.2.2. Análisis multivariante

Con el objetivo de comparar de manera rigurosa los resultados obtenidos con las muestras de CEA de cáncer de colon (CEA1 y CEA2) y de metástasis de hígado (CEA3) analizadas (seis replicados de cada una, S1-6 para CEA1, S7-12 para CEA2, y S13-18 para CEA3), e identificar qué glicopéptidos permitían diferenciar entre ellas, se estudiaron los mapas glicopeptídicos

obtenidos con herramientas quimiométricas tales como el análisis por componentes principales (PCA) y el análisis discriminante por mínimos cuadrados parciales (PLS-DA). Para evitar diferencias debidas a la variabilidad en las condiciones experimentales, fue necesario llevar a cabo un pretratamiento de los datos que consistió en normalizar las A_p de los glicopéptidos respecto a un péptido de referencia que se comportase de manera similar a los *N*-glicopéptidos identificados por lo que respecta a los t_m y a las A_p ($S_{127}DLVNEEATGQFR_{139}$, $S_{40}TPFNVAE_{47}$ y $A_{89}TPGP_{93}/T_{90}PGPA_{94}$ para la digestión con tripsina, Glu-C y pronasa, respectivamente). Tras la normalización, la repetitividad (%RSD, $n=6$) fue adecuada para todos los *N*-glicopéptidos identificados con las tres enzimas. Por ejemplo, en el caso de la tripsina, se obtuvieron unos valores de repetitividad de entre 0,5% y 7,9% para los t_m y de entre 2,6% y 10,0% para las áreas normalizadas (A_{norm}).

5.2.2.1. PCA

Una vez construida la matriz con las A_{norm} para los diferentes glicopéptidos, se construyeron tres modelos de PCA (para las muestras de CEA digeridas con tripsina, Glu-C y pronasa) con el objetivo de explorar las clases presentes en las muestras analizadas, así como para detectar la presencia de aberrantes (*outliers*). En todos los casos, se escogieron 2 componentes principales (PCs) ya que permitieron explicar un 90%, 87% y 82% de la varianza para las muestras de CEA digeridas con tripsina, Glu-C y pronasa, respectivamente. Es importante destacar que en el caso de la digestión enzimática con pronasa, la muestra S9 (de cáncer de colon, CEA2) fue detectada como un *outlier* (**Figura 5.23**), probablemente debido a las bajas relaciones S/N en esta muestra en concreto. Los tres modelos de PCA permitieron discriminar entre muestras de cáncer de colon (CEA1 y CEA2) y de metástasis de hígado (CEA3), siendo el PC1 el responsable de dicha discriminación. Una vez confirmada la separación de los dos grupos o clases de muestras en los datos analizados, se llevó a cabo un análisis por PLS-DA con el objetivo de identificar qué glicopéptidos permitían diferenciar entre estos dos grupos.

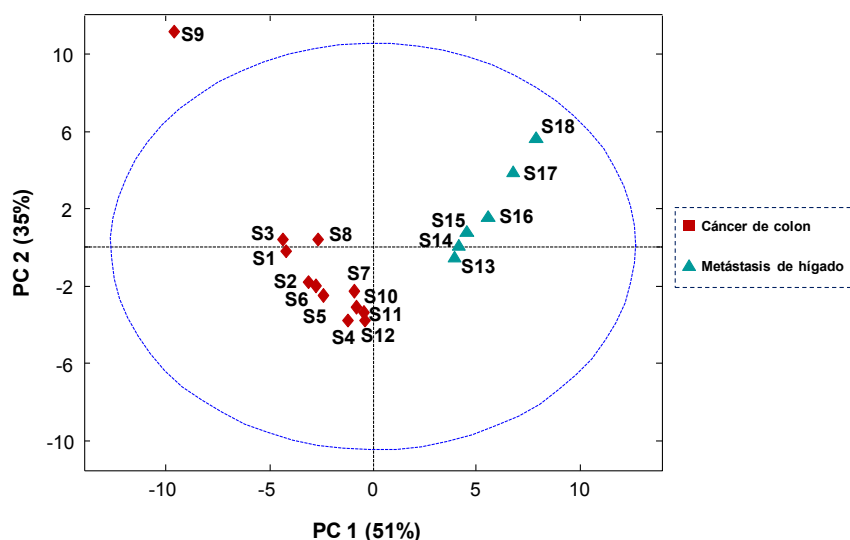


Figura 5.23. Gráfico de *scores* del PCA para las muestras de CEA analizadas por *sheathless* CE-MS/MS después de llevar a cabo la digestión con pronasa. La muestra S9 (de cáncer de colon, CEA2) fue detectada como un *outlier* y descartada para el resto de estudios.

5.2.2.2. PLS-DA

Se construyeron tres modelos de PLS-DA (considerando los dos grupos de muestras observados por PCA) para las muestras de CEA digeridas con tripsina, Glu-C y pronasa. En todos los casos se escogieron 2 variables latentes (LVs), ya que permitieron explicar un 90%, 84% y 68% de la varianza en X, y un 99%, 89% y 90% de la varianza en Y para las muestras de CEA digeridas con tripsina, Glu-C y pronasa, respectivamente. Los tres modelos de PLS-DA permitieron discriminar entre las muestras de CEA de cáncer de colon (CEA1 y CEA2) y de metástasis de hígado (CEA3), siendo la LV1 la responsable de dicha discriminación. La **Figura 5.24** muestra los gráficos de *scores* para las muestras de CEA digeridas con (A) tripsina, (B) Glu-C y (C) pronasa. Una vez contruidos los tres modelos de PLS-DA, se estimó la contribución de cada variable (*N*-glicopéptido) en las LVs mediante los gráficos de *loadings* correspondientes. Sin embargo, estos gráficos sólo nos permiten obtener información cualitativa sobre los *N*-glicopéptidos que influyen. Con el objetivo de cuantificar la influencia de las diferentes variables en la separación de las muestras de CEA, se seleccionaron todas las variables

importantes en la proyección (VIPs) con un valor superior a 1. Cabe resaltar que el método de los VIPs es uno de los más utilizados para cuantificar la importancia de las diferentes variables en los modelos de PLS-DA, principalmente debido a su capacidad para trabajar con datos multicolineales [229,248]. En este caso, un total de 121 *N*-glicopéptidos fueron seleccionados por su relevancia para diferenciar entre los dos grupos de muestras (71 de 128 para la tripsina, 26 de 50 para la Glu-C y 24 de 41 para la pronasa).

Con el objetivo de encontrar alguna tendencia que permitiese explicar las diferencias observadas entre las muestras de CEA de cáncer de colon y de metástasis de hígado, los gráficos de barras de la **Figura 5.24** muestran las A_{norm} de los *N*-glicopéptidos con VIPs superiores a 1 para las muestras de CEA digeridas con (A) tripsina, (B) Glu-C y (C) pronasa. Tal y como se observa en la **Figura 5.24 A** para la digestión con tripsina, las A_{norm} para los *N*-glicopéptidos fucosilados (entre 1 y 7 fucosas) y sialilados (entre 1 y 2 ácidos siálicos) aumentaron en la muestra de CEA de metástasis de hígado (CEA3). Esta tendencia se observó para todos los puntos de *N*-glicosilación, excepto para el punto *N*-204/*N*-560, en el que algunos *N*-glicopéptidos fucosilados y sialilados presentaron una tendencia claramente opuesta. Esta tendencia opuesta también se observó en el caso de la digestión con Glu-C (**Figura 5.24 B**), para la que las A_{norm} de los *N*-glicopéptidos fucosilados (entre 1 y 3 fucosas) aumentaron en las muestras de CEA de cáncer de colon (CEA y CEA2). En el caso de la digestión con pronasa (**Figura 5.24 C**), se observaron tendencias diferentes en los *N*-glicopéptidos complejos y ricos en manosa para las muestras de cáncer de colon (CEA1 y CEA2) y de metástasis de hígado (CEA3). Por ejemplo, el A_{norm} para el *N*₂₅₆-glicopéptido H6N2 aumentó en muestras de CEA de cáncer de colon, mientras que se observó una tendencia opuesta para los *N*₂₅₆-glicopéptidos H8N2 y H9N2. Por el contrario, tal y como se observó tras digerir las muestras de CEA con tripsina, las A_{norm} para los *N*-glicopéptidos fucosilados (entre 2 y 4 fucosas) y sialilados (entre 1 y 2 ácidos siálicos) aumentaron en la muestra de CEA procedente de metástasis de hígado (CEA3).

A) Digestión con tripsina

i) Gráfico de scores del PLS-DA

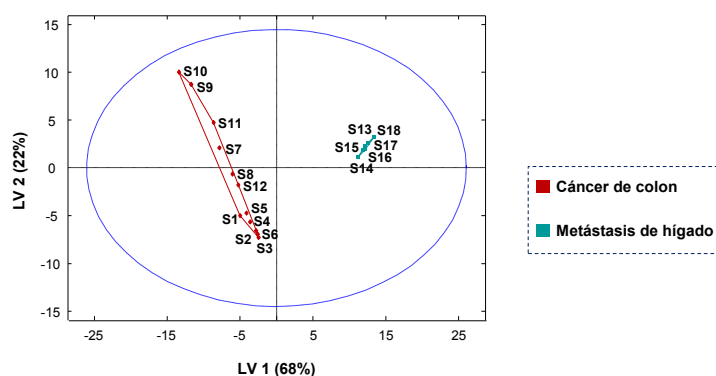
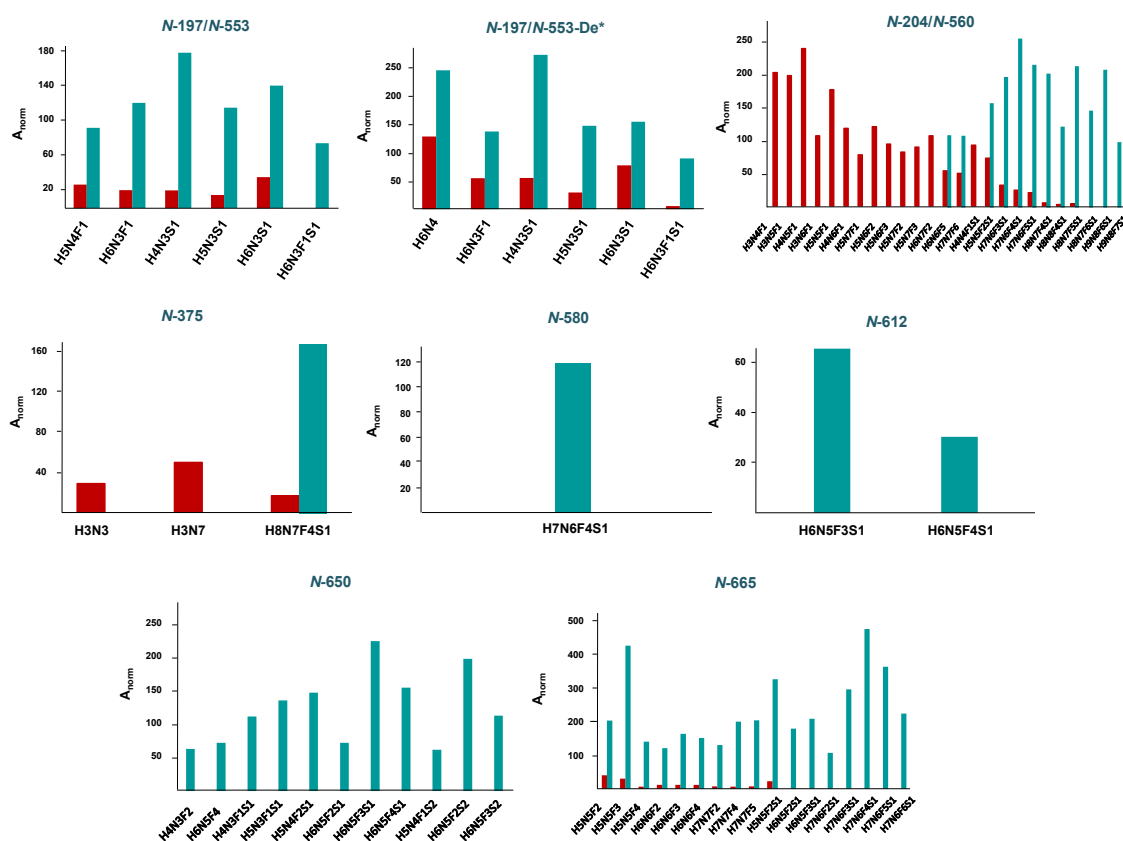

 ii) A_{norm} para los 71 *N*-glicopéptidos con $VIP > 1$


Figura 5.24 A. Gráfico de *scores* (i) y áreas normalizadas (A_{norm}) (ii) del PLS-DA para los *N*-glicopéptidos relevantes ($VIP > 1$) para diferenciar entre muestras de cáncer de colon (CEA1 y CEA2) y de metástasis de hígado (CEA3) después de la digestión con tripsina. *El punto de glicosilación *N*-197/*N*-553 puede estar deamidado.

B) Digestión con Glu-C

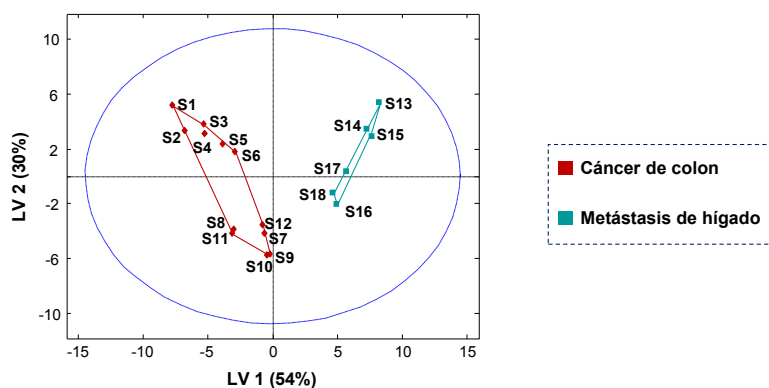
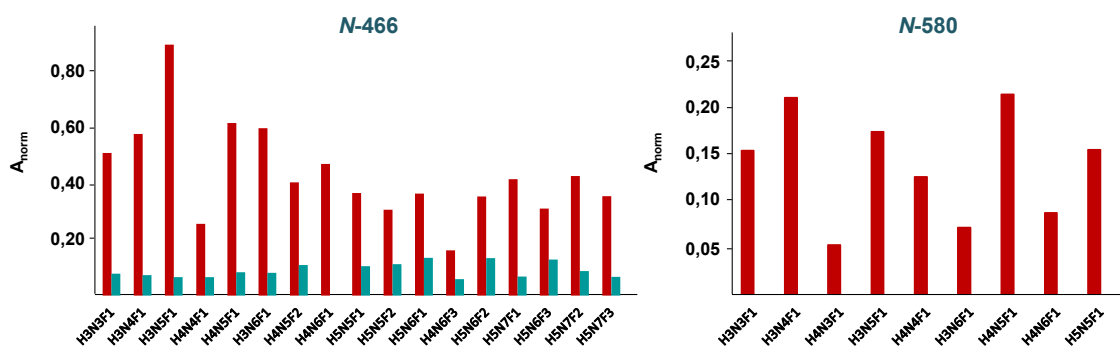
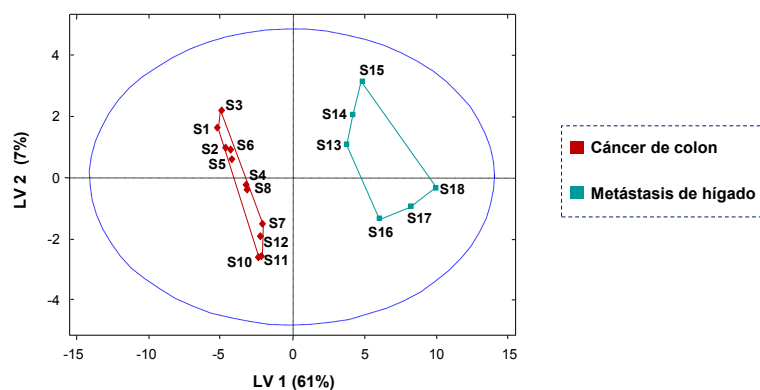
i) Gráfico de *scores* del PLS-DAii) A_{norm} para los 26 *N*-glicopéptidos con $\text{VIP}>1$ 

Figura 5.24 B (continúa). Gráfico de *scores* (i) y áreas normalizadas (A_{norm}) (ii) del PLS-DA para los *N*-glicopéptidos relevantes ($\text{VIP}>1$) para diferenciar entre muestras de cáncer de colon (CEA1 y CEA2) y de metástasis de hígado (CEA3) después de llevar a cabo la digestión con Glu-C.

C) Digestión con pronasa

i) Gráfico de scores del PLS-DA



ii) A_{norm} para los 24 *N*-glicopéptidos con $VIP > 1$

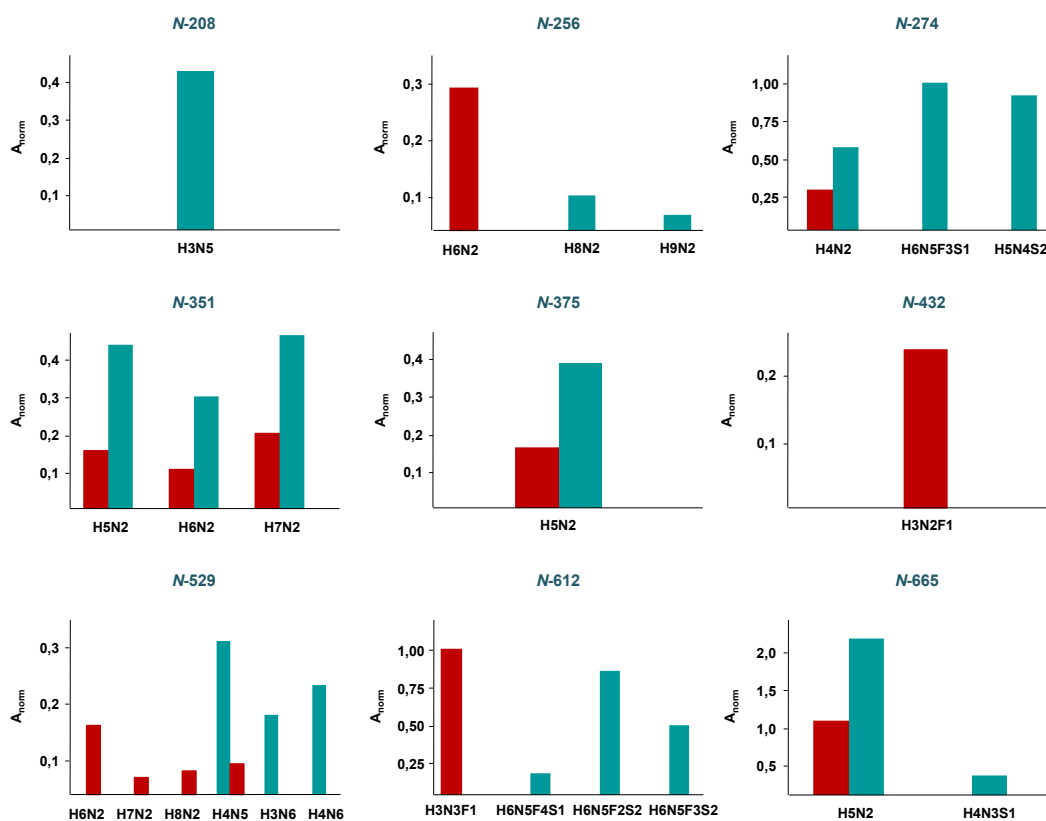


Figura 5.24 C (continúa). Gráfico de *scores* (i) y áreas normalizadas (A_{norm}) (ii) del PLS-DA para los *N*-glicopéptidos relevantes ($VIP > 1$) para diferenciar entre muestras de cáncer de colon (CEA1 y CEA2) y de metástasis de hígado (CEA3) después de llevar a cabo la digestión con pronasa.

En este sentido, algunos autores han descrito la importancia de la fucosilación y, en especial, de la sialilación, en una multitud de procesos biológicos relacionados con el reconocimiento y señalización celulares [249,250]. En concreto, se han descrito valores elevados de fucosilación y sialilación en procesos tumorales, observándose un aumento en el número de fucosas y ácidos siálicos a medida que aumenta la gravedad y la invasión del tumor en cuestión [251–254]. En general, este hecho concuerda con los resultados obtenidos en este trabajo (**artículo 3.1**), en el que se observa un mayor grado de fucosilación (por lo general, entre 3 y 7 fucosas) y sialilación (entre 1 y 2 ácidos siálicos) en la muestra de CEA procedente de metástasis de hígado (CEA3). Sin embargo, un análisis multivariante de datos sistemático y detallado mediante herramientas quimiométricas avanzadas, permitió poner de manifiesto diferencias entre algunos puntos de glicosilación en concreto, los cuales podrían ser de utilidad para explicar el proceso de metástasis en pacientes con cáncer de colon. Así mismo, se pudo observar que para los puntos de glicosilación *N*-204/*N*-560 (identificados con tripsina) y los puntos de glicosilación *N*-466 y *N*-580 (identificados con Glu-C), la abundancia de los *N*-glicopéptidos poco fucosilados (en general, entre 1 y 3 fucosas) era superior en las muestras de CEA de cáncer de colon, tendencia claramente opuesta a la observada para los *N*-glicopéptidos altamente fucosilados (entre 3 y 7 fucosas).

A pesar de los buenos resultados obtenidos con la metodología desarrollada, deberían analizarse un mayor número de muestras de CEA para poder confirmar la eficacia del método propuesto y, por consiguiente, profundizar en el estudio de los mecanismos de fucosilación y sialilación que tienen lugar en diferentes procesos tumorales. Para estos estudios, sería necesario desarrollar métodos de purificación adecuados similares a los descritos en los **artículos 2.1, 2.2, 2.3 y 2.4** de esta tesis doctoral para aislar la CEA de los fluidos y tejidos biológicos donde se encuentra.

5.3. Análisis no dirigido para la identificación de biomarcadores metabólicos (*untargeted analysis*). Aplicación a la enfermedad del Huntington

A veces no es suficiente con estudiar una proteína biomarcadora relevante para una determinada enfermedad, sino que es necesario integrar con la información proteómica la conseguida con otras ciencias ómicas: genómica, transcriptómica y metabolómica. En este capítulo de tesis se ha hecho especial énfasis en el estudio de la metabolómica ya que permite obtener, de forma directa, información relacionada con la actividad celular y, por consiguiente, es más fácil de correlacionar con el fenotipo.

En este sentido, la detección e identificación de compuestos de baja masa molecular en fluidos biológicos, tales como el plasma, tiene un papel fundamental en la investigación metabolómica para la determinación de nuevos biomarcadores. En esta tesis doctoral se ha establecido una metodología adecuada para el análisis de compuestos de baja masa molecular en plasma mediante extracción en fase sólida en línea a la electroforesis capilar acoplada a la espectrometría de masas (SPE-CE-MS). Para ello, se ha utilizado un sorbente C₁₈, ya que este tipo de fases estacionarias hidrofóbicas tienen una selectividad adecuada para un gran número de compuestos de baja masa molecular. Así mismo, nuestro grupo de investigación ya ha demostrado previamente que estos sorbentes son los más apropiados para el análisis de neuropéptidos de baja masa molecular en muestras de plasma humano por C₁₈-SPE-CE-MS [201].

Una vez optimizada la metodología de C₁₈-SPE-CE-MS con neuropéptidos modelo, ésta se ha aplicado al análisis de muestras de plasma de ratones con el fin de identificar biomarcadores de baja masa molecular que puedan resultar de utilidad en el diagnóstico precoz y el seguimiento de la enfermedad del Huntington (HD), una enfermedad neurodegenerativa hereditaria. Para ello, se han analizado muestras de plasma de ratones sanos (*wild-type*, wt) y de ratones

modificados genéticamente para desarrollar esta enfermedad mediante C_{18} -SPE-CE-MS y se han comparado los resultados obtenidos con el fin de encontrar diferencias significativas entre ambos grupos. El tratamiento de los datos obtenidos se ha abordado siguiendo una estrategia de análisis no dirigido (*untargeted analysis*), la cual consiste en evaluar los resultados desde un punto de vista global (*global profiling*), a fin de identificar compuestos que puedan discriminar entre plasma de ratones wt y HD y dar lugar a nuevos biomarcadores de esta enfermedad. En este caso, ha sido necesario emplear herramientas quimiométricas tales como la resolución multivariante de curvas por mínimos cuadrados alternados (MCR-ALS) y el análisis discriminante por mínimos cuadrados parciales (PLS-DA). El método empleado ha permitido la identificación de posibles biomarcadores metabólicos del HD, así como la interpretación de las diferentes rutas metabólicas que podrían verse alteradas durante la progresión de la enfermedad.

5.3.1. Optimización de un tratamiento de muestra de plasma humano previo al análisis por C_{18} -SPE-CE-MS

El análisis de muestras complejas por SPE-CE-MS, especialmente si se utilizan fases estacionarias de selectividad limitada, como el C_{18} , requiere de un tratamiento previo de las muestras con el objetivo de prevenir la saturación del preconcentrador, el cual contiene una cantidad limitada de fase estacionaria.

En un estudio anterior de nuestro grupo de investigación, se propuso un tratamiento de muestra previo al análisis de neuropéptidos en muestras de plasma humano por CE-MS y C_{18} -SPE-CE-MS [201]. Éste consistía en llevar a cabo una precipitación con ACN (proporción 6:1 (v/v), ACN:plasma) seguida de una etapa de ultrafiltración con filtros Microcon Amicon, con membrana de acetato de celulosa y de corte molecular de M_r 10000 (también llamados YM-10). Sin embargo, estos filtros YM-10 con la membrana en posición horizontal para la filtración directa fueron sustituidos por el fabricante por otros con la membrana en posición vertical para

la filtración tangencial (Amicon Ultra-0.5). Estos nuevos filtros proporcionaron peores resultados utilizando las condiciones previas óptimas, por lo que en esta tesis doctoral (**artículo 4.1**) se revisó el tratamiento propuesto para poder utilizar los nuevos filtros y obtener buenas recuperaciones. Para ello, como en el trabajo anterior de nuestro grupo de investigación [201], se emplearon los tres neuropéptidos modelo que se muestran en la **Tabla 5.8**.

Tabla 5.8. Secuencia, M_r y m/z para los neuropéptidos dinorfina A (fragmento 1-7) (Dyn A), endomorfina 1 (End 1) y met-enkefalina (Met) estudiados por CE-MS y C_{18} -SPE-CE-MS.

Péptido	Dyn A (1-7)	End 1	Met
Secuencia	Tyr-Gly-Gly-Phe-Leu-Arg-Arg	Tyr-Pro-Trp-Phe-NH ₂	Tyr-Gly-Gly-Phe-Met
M_r	867,4715	610,2904	573,2257
$[M+zH]^{z+} m/z$	868,4788 ($z=1$) 434,7430 ($z=2$)	611,2976 ($z=1$)	574,2330 ($z=1$)

En la **Figura 5.25** se muestra la separación obtenida por CE-MS para una mezcla patrón de Dyn A, End 1 y Met de 100 $\mu\text{g/mL}$ de concentración.

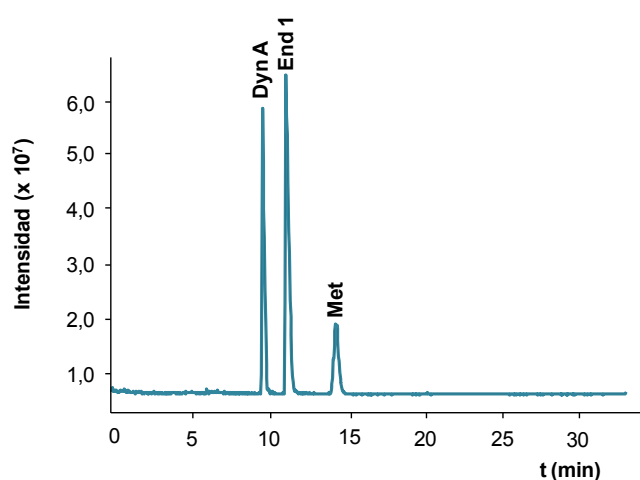


Figura 5.25. Suma de EIEs obtenidos por CE-MS para una mezcla patrón de Dyn A, End 1 y Met de 100 $\mu\text{g/mL}$ de concentración. BGE: 50 mM HAc:50 mM HFor, ajustado a pH=3,50 con NH₃.

La **Figura 5.26** muestra el tratamiento que se propuso en nuestro trabajo previo para el análisis de estos neuropéptidos en muestras de plasma humano [201] y que se siguió en el **artículo 4.2** de esta tesis doctoral, haciéndose especial énfasis en la optimización de la etapa de ultrafiltración.

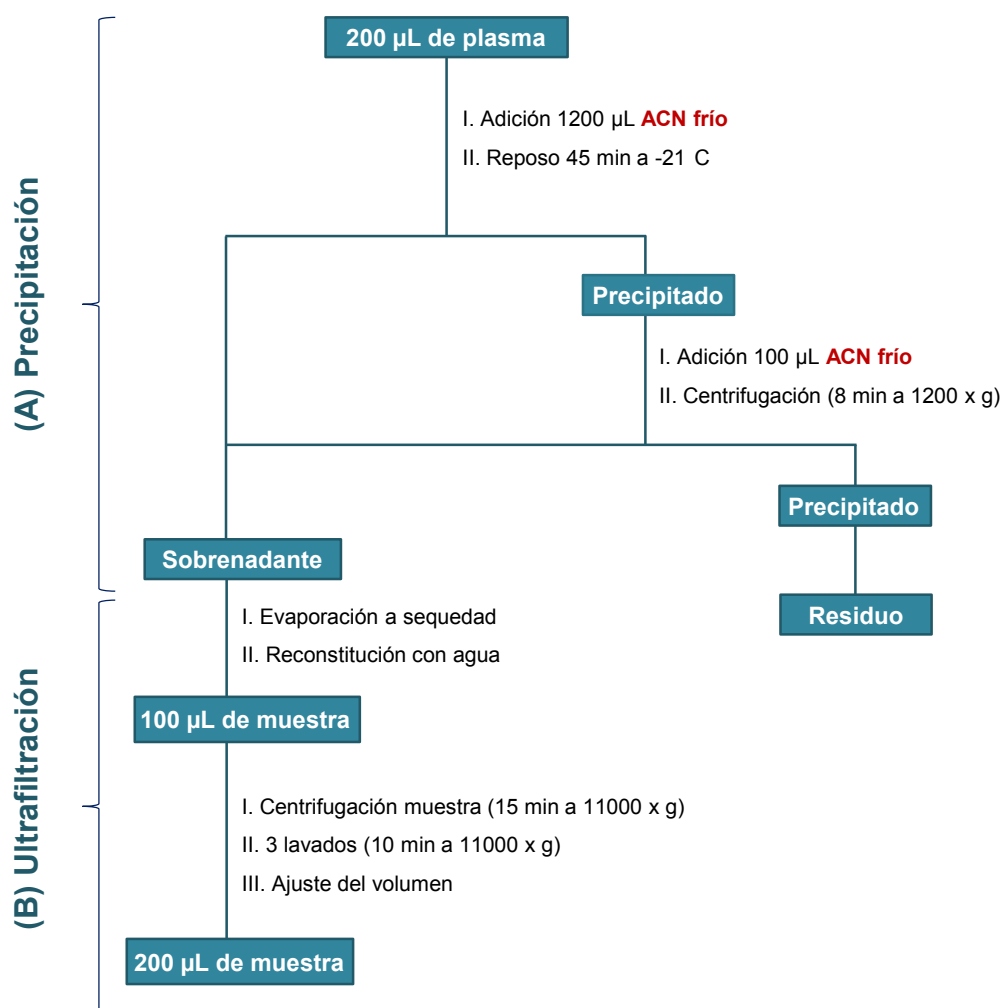


Figura 5.26. Tratamiento de muestras de plasma humano previo al análisis por CE-MS y C₁₈-SPE-CE-MS: (A) Precipitación y (B) ultrafiltración.

En este trabajo previo [201], se determinó la recuperación de cada péptido en la etapa de precipitación con ACN (**Figura 5.26 A**) mediante la comparación de las A_p obtenidas por CE-MS para muestras de plasma humano fortificadas con 10 µg/mL de cada péptido antes y después de la precipitación (49%, 91% y 102% para Dyn A, End 1 y Met, respectivamente). Las

recuperaciones más bajas correspondieron a la Dyn A, probablemente porque, de los tres neuropéptidos estudiados, es el que presenta una M_r superior.

En esta tesis se estudiaron, en primer lugar por CE-MS, las recuperaciones para la Dyn A, la End 1 y la Met después de aplicar un tratamiento de precipitación seguido de una etapa de ultrafiltración a muestras de plasma humano fortificadas con 10 $\mu\text{g/mL}$ de cada péptido usando los siguientes filtros: filtros Amicon Ultra-0.5 (*Millipore*) de corte molecular de M_r 3000, 10000 y 30000 con membrana de acetato de celulosa (3-CA, 10-CA y 30-CA, respectivamente) y filtros Vivaspin (*GE Healthcare*) de corte molecular de M_r 10000 con membrana de polietersulfona (10-PES).

Como los neuropéptidos de interés se encontraban en el filtrado, en primer lugar, se evaluó la influencia del tamaño de poro de los filtros 3-CA, 10-CA, 30-CA y 10-PES para poder descartar la retención debida a mecanismos de exclusión por tamaño. Es conveniente recordar que los filtros de corte molecular están pensados fundamentalmente para concentrar sustancias de M_r superior a la indicada y no para aislar las de M_r inferior. Así mismo, dependiendo de la M_r de los neuropéptidos, éstos quedarán más o menos retenidos en el filtro durante la filtración. Por eso, en todos los experimentos, el menor porcentaje de recuperación correspondió a la Dyn A, que además de tener una recuperación limitada en la primera etapa de precipitación, presenta una M_r superior. Tal y como se muestra en la **Figura 5.27**, en ningún caso se consiguieron recuperar los tres péptidos con filtros 3-CA, 10-CA y 10-PES. Por ejemplo, utilizando filtros 10-CA y 10-PES sólo se pudo detectar la Met con porcentajes de recuperación del 43% y 41%, respectivamente. Las mejores recuperaciones se obtuvieron con filtros 30-CA, pero en contrapartida, los filtrados contenían mayor cantidad de interferencias de la matriz de plasma debido a su mayor tamaño de poro, lo que es poco recomendable para obtener buenas reproducibilidades y alargar la vida de los microcartuchos en SPE-CE-MS.

A continuación, se estudió la posible adsorción de los analitos en la superficie de plástico del recipiente de los filtros pasivándolos con un 1% (m/v) de BSA en PBS y con una solución acuosa al 5% (v/v) de PEG. El uso de estas disoluciones tiene como objetivo principal bloquear los principales sitios activos de los filtros, evitando de esta manera, la adsorción de los compuestos de interés en sus paredes. Tal y como se muestra en la **Figura 5.27**, en todos los casos las recuperaciones mejoraron después de la pasivación de los filtros 3-CA y 10-CA con BSA y con PEG. Sin embargo, ninguna de las dos soluciones empleadas para pasivar pudo ser propuesta como estándar de pasivación, ya que los resultados obtenidos para los diferentes filtros mostraron tendencias contrapuestas. Las mejores recuperaciones para los 3-CA se obtuvieron pasivando los filtros con BSA, mientras que para los 10-CA, las mejores recuperaciones se consiguieron con PEG. En ambos casos se obtuvo una reproducibilidad similar (%RSD media del 10% y 9%, respectivamente), pero las recuperaciones fueron ligeramente mejores con los filtros 10-CA pasivados con PEG.

Todavía se intentó evaluar el efecto de la temperatura antes de la ultrafiltración, ya que un aumento de ésta puede ayudar a romper las interacciones hidrofóbicas e iónicas que mantienen los péptidos en la superficie de un gran número de proteínas del plasma, lo que podría provocar una menor concentración en los filtrados [255]. El tratamiento consistió en incubar las muestras a 85°C durante 15 min antes de la etapa de ultrafiltración con filtros 10-CA pasivados con PEG. Tal y como se observa en la **Figura 5.27**, se obtuvieron recuperaciones significativamente más bajas empleando este tratamiento. Finalmente, fue posible concluir que los resultados para los filtros 10-CA pasivados con PEG fueron similares a los obtenidos con los filtros YM-10 en trabajos anteriores de nuestro grupo de investigación [201] y que ya no se encontraban comercialmente disponibles.

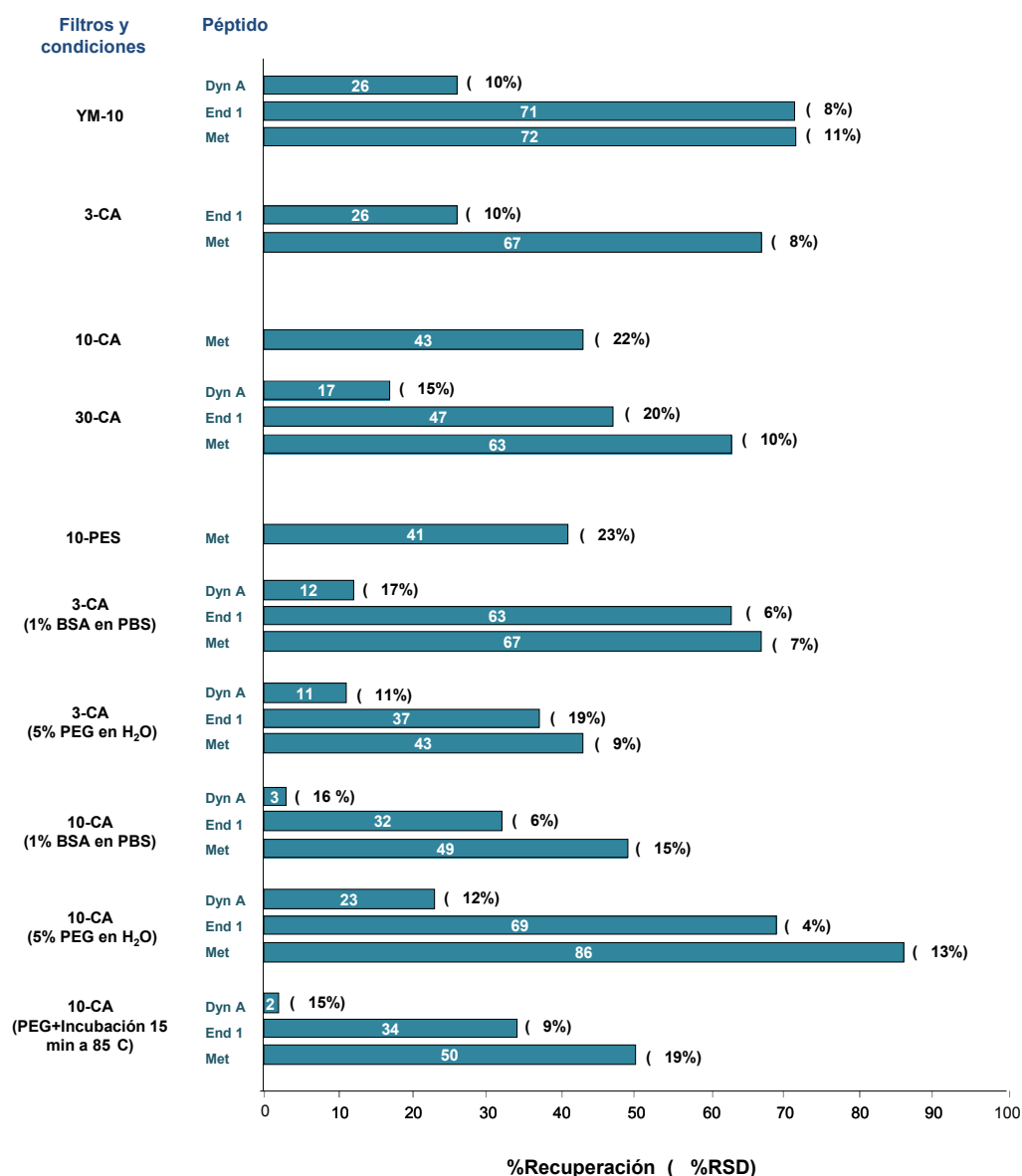


Figura 5.27. Evaluación de la recuperación de los péptidos por CE-MS utilizando, tras la precipitación con ACN: filtros 3-CA, 10-CA, 30-CA y 10-PES en diferentes condiciones (la temperatura es 25°C si no se indica lo contrario). Las muestras de plasma humano se fortificaron antes y después del tratamiento de muestra completo con 10 µg/mL de cada péptido. YM-10 hace referencia a los filtros utilizados en el trabajo previo de nuestro grupo de investigación [201].

La **Figura 5.28** muestra los EIEs obtenidos por CE-MS para blancos y muestras de plasma humano fortificadas con 10 µg/mL de cada péptido después de llevar a cabo el tratamiento de precipitación seguido de la ultrafiltración con filtros 10-CA (A) sin pasivar y (B) pasivados con

PEG. Tal y como se puede observar, utilizando los filtros 10-CA sin pasivar sólo se pudo detectar la Met, mientras que pasivando los filtros con PEG, se consiguieron detectar los tres péptidos.

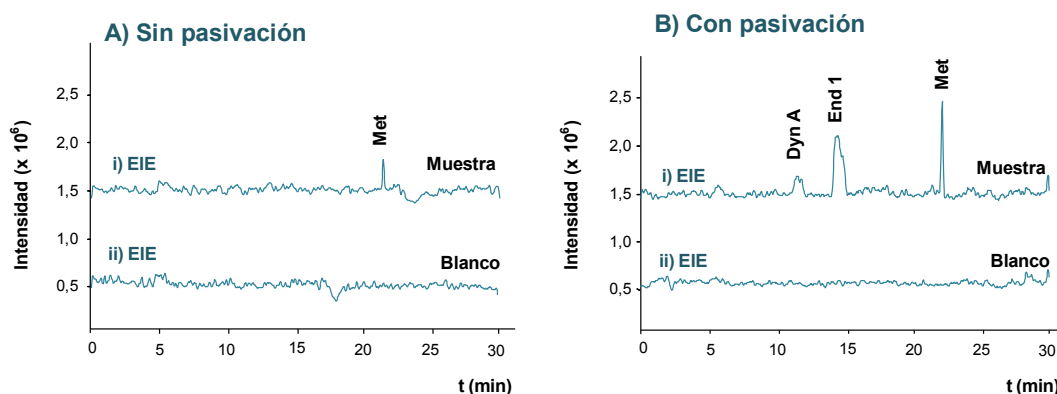


Figura 5.28. Suma de EIEs (Dyn A, End 1 y Met) obtenidos por CE-MS para blancos y muestras fortificadas con 10 µg/ml de cada péptido utilizando filtros 10-CA (A) sin pasivar y (B) pasivados con PEG tras la precipitación con ACN.

Una vez establecido el tratamiento de plasma óptimo por CE-MS, se estudió la validez de este tratamiento para el análisis de los neuropéptidos modelo en muestras de plasma humano por C₁₈-SPE-CE-MS. Para llevar a cabo los análisis, se fortificaron muestras de plasma humano con 100 ng/mL de cada uno de los péptidos y se sometieron al tratamiento de precipitación con ACN seguido de la ultrafiltración utilizando filtros 10-CA pasivados con PEG y sin pasivar. Tal y como se observa en la **Figura 5.29**, utilizando los filtros 10-CA sin pasivar, no fue posible detectar la Dyn A (péptido con recuperaciones más bajas, **Figura 5.27**) y la altura del pico correspondiente a la End 1 decreció considerablemente respecto a la altura del pico obtenido después de pasivar los filtros. No se observó ninguna diferencia significativa para los picos de la Met obtenidos pasivando y sin pasivar los filtros. Utilizando los filtros 10-CA pasivados, fue posible detectar los neuropéptidos en muestras de plasma a 10 ng/ml para la Dyn A y 1 ng/ml para la End 1 y la Met, tal y como sucedía en los estudios anteriores de nuestro grupo de

investigación utilizando los filtros YM-10 sin pasivar. Es por ello que, este tratamiento de plasma utilizando los filtros 10-CA pasivados con PEG, se aplicó a la detección e identificación de compuestos de baja masa molecular en plasma de ratones para identificar biomarcadores metabólicos del HD mediante C_{18} -SPE-CE-MS.

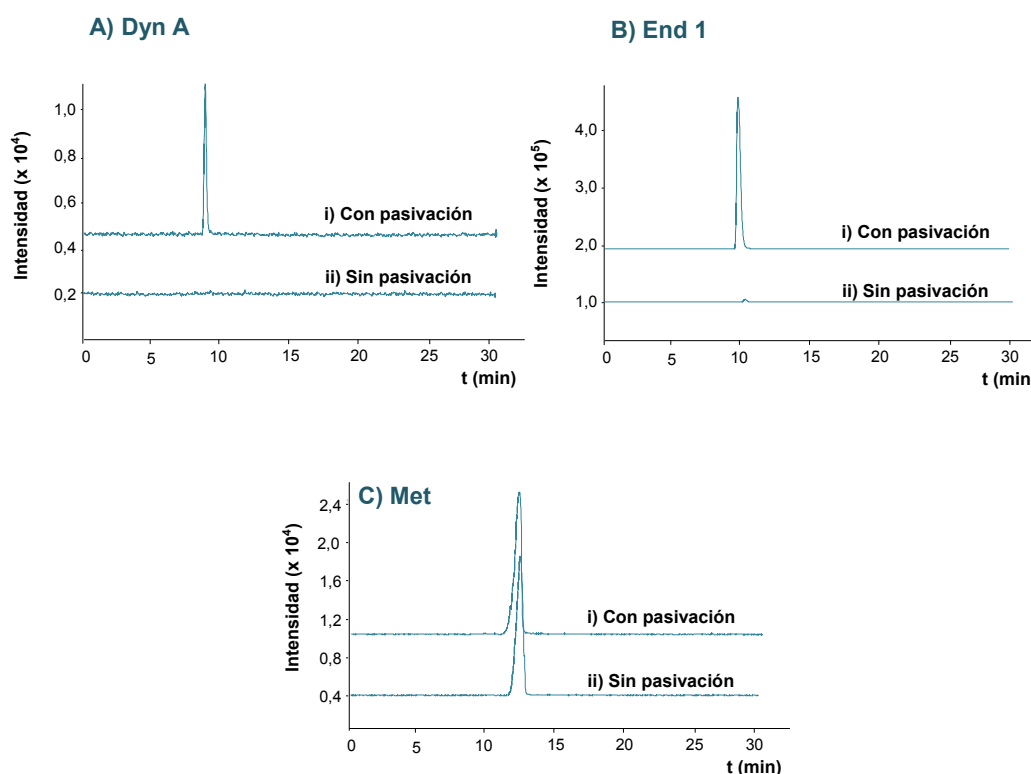


Figura 5.29. EIEs obtenidos por C_{18} -SPE-CE-MS para muestras de plasma fortificadas con 100 ng/ml de (A) Dyn A, (B) End 1 y (C) Met utilizando filtros 10-CA (i) pasivados con PEG y (ii) sin pasivar.

5.3.2. Análisis de compuestos de baja masa molecular en plasma de ratones wt y HD por C_{18} -SPE-CE-MS

La enfermedad del Huntington (HD) afecta a una proteína expresada en numerosos tejidos, llamada huntingtina. El defecto se debe a la expansión, en la secuencia del gen que codifica esta proteína, de una cantidad variable de tripletes CAG (citosina, adenina y guanina, respectivamente). En la secuencia original hay menos de 28 repeticiones, y en la enfermedad,

más de 39. En la actualidad, el análisis de la huntingtina no es suficientemente eficaz para poder diagnosticar y predecir un debut temprano de esta enfermedad. Por ello, el análisis de esta proteína se complementa con otros estudios, entre los cuales la metabolómica juega un papel muy importante. Uno de los objetivos de esta tesis doctoral es la identificación de compuestos de baja masa molecular en muestras de plasma, que puedan resultar de utilidad en el diagnóstico y el seguimiento del HD. Por este motivo, se analizaron muestras de plasma de ratones sanos (wt) y de ratones modificados genéticamente para desarrollar esta enfermedad (R6/1 o HD) mediante el método de C₁₈-SPE-CE-MS optimizado previamente (**artículo 4.1**) y se compararon los resultados obtenidos con el fin de encontrar las diferencias más significativas entre ambos grupos. Además, para poder extraer conclusiones sobre el inicio y la progresión de la enfermedad, se analizaron muestras de ambos grupos a 8, 12 y 30 semanas de edad. En los ratones HD estas muestras se corresponderían con individuos asintomáticos, presintomáticos y sintomáticos terminales, respectivamente (aunque esta clasificación se basa únicamente en los déficits de coordinación motora). Es conveniente destacar que, dada la limitada cantidad de sangre que se puede extraer de un único ratón (entre 1 y 2 mL en ratones wt, y este volumen es menor en ratones HD de edades avanzadas), cada grupo de muestras corresponde a plasma obtenido de la combinación de la sangre de varios ratones a cada una de las edades (entre 4 y 5 por genotipo y edad).

El procedimiento experimental consistió en preparar los seis tipos de muestras de plasma (por triplicado, a excepción de 12wt, 12HD y 30 HD, para los cuales se analizaron dos replicados debido a la limitada cantidad de muestra) con el tratamiento de precipitación con ACN (6:1 (v/v), ACN:plasma) seguido de la ultrafiltración utilizando filtros 10-CA pasivados con PEG, con los que se obtuvieron mejores recuperaciones para los neuropéptidos (**artículo 4.1**), asumiendo que esto era extrapolable en general a otros compuestos de baja masa molecular. Una vez preparadas, se analizaron todas las muestras en orden aleatorio mediante C₁₈-SPE-CE-MS. Para el análisis de cada seis muestras (8, 12 y 30 semanas wt y 8, 12 y 30 semanas HD) se empleó un preconcentrador diferente, ya que con muestras de plasma, los preconcentradores

tienen una vida media de 10 análisis. Al principio y al final de cada secuencia de análisis, y con el objetivo de comprobar el buen estado y funcionamiento de los preconcentradores, se analizó una mezcla patrón de Dyn A, End 1 y Met de 10 ng/mL.

En la **Figura 5.30** se muestran los TIEs obtenidos por C₁₈-SPE-CE-MS para cada uno de los seis tipos de muestras de plasma.

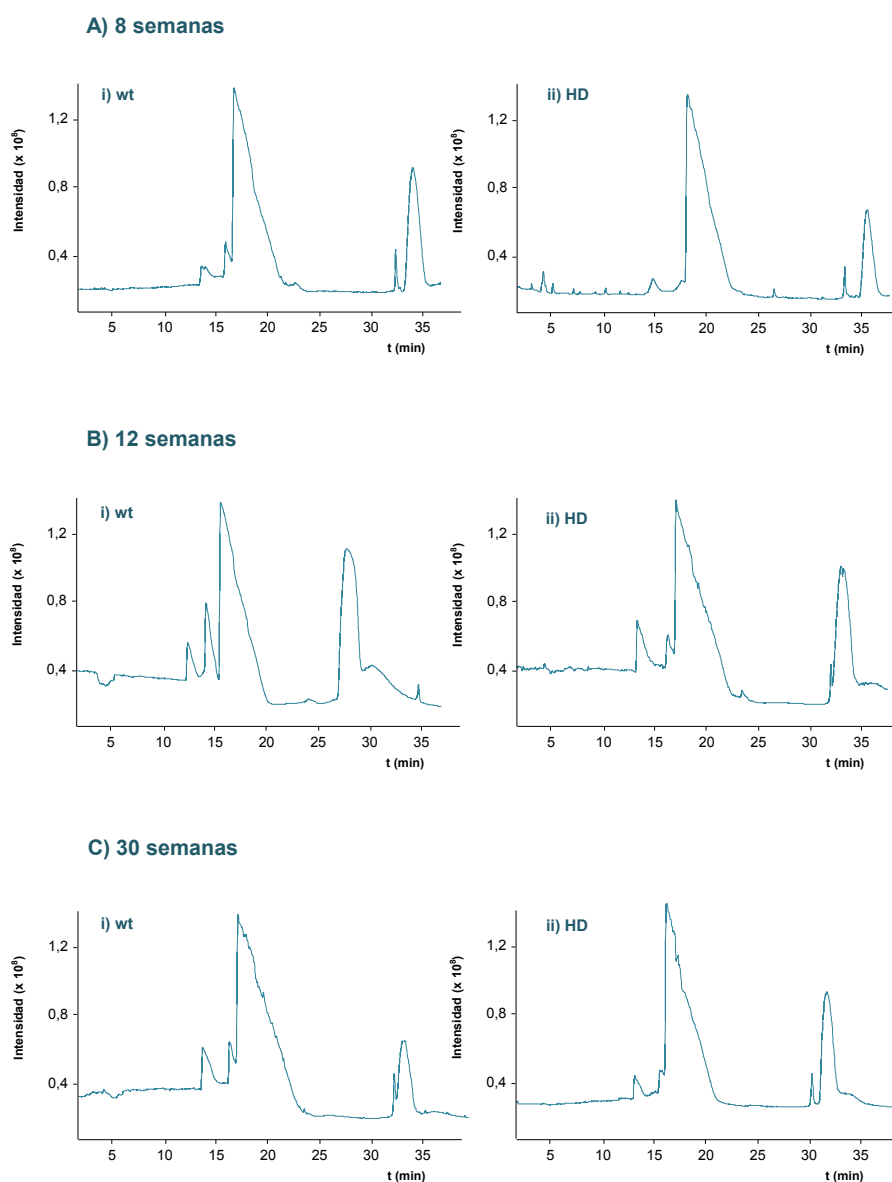


Figura 5.30. TIEs entre 40 y 3200 m/z obtenidos por C₁₈-SPE-CE-MS para muestras de plasma a (A) 8 semanas, (B) 12 semanas y (C) 30 semanas de edad para ratones (i) wt y (ii) HD.

Tal y como se puede observar, el perfil de los electroferogramas a simple vista es bastante similar. Por este motivo, y con el objetivo de evaluar si existían diferencias significativas entre ambos grupos y en función de la edad, el tratamiento de los datos obtenidos se abordó siguiendo una estrategia de análisis no dirigido (*untargeted analysis*), la cual consistió en evaluar los resultados desde un punto de vista global (*global profiling*) a fin de detectar los compuestos que discriminaban entre los diferentes grupos e identificar aquellos metabolitos que podían dar lugar a nuevos biomarcadores del HD. Para extraer automática y sistemáticamente la información sobre los compuestos detectados, fue necesario utilizar una herramienta quimiométrica denominada resolución multivariante de curvas por mínimos cuadrados alternados (MCR-ALS), la cual permitió obtener los perfiles electroforéticos de los diferentes componentes de las muestras analizadas y sus correspondientes espectros de masas. Esta información fue posteriormente analizada empleando los métodos de análisis multivariante de datos descritos en la sección anterior (**artículo 3.1**): PCA y PLS-DA.

5.3.3. Herramientas quimiométricas para la detección e identificación de biomarcadores metabólicos del Huntington

5.3.3.1. MCR-ALS

En la **Figura 5.31** se muestra la estrategia utilizada para el pretratamiento de los datos y la aplicación de MCR-ALS a las matrices obtenidas previamente por C₁₈-SPE-CE-MS (8wt, 12wt, 30wt, 8HD, 12HD y 30HD). Este pretratamiento de los datos consistió en una reducción del modo de resolución del MS (de 0,0001 Da/e a 0,01 Da/e), una reducción del modo de tiempo seleccionando únicamente las regiones del electroferograma con picos más intensos (2 ventanas de tiempo, aproximadamente entre 10-25 min y entre 30-40 min, **Figura 5.30**) y, finalmente, una reducción del intervalo de m/z en intervalos más pequeños (30 intervalos de m/z). En total,

resultaron 60 matrices aumentadas (2 ventanas de tiempo x 30 intervalos de m/z) que se analizaron por separado con MCR-ALS.

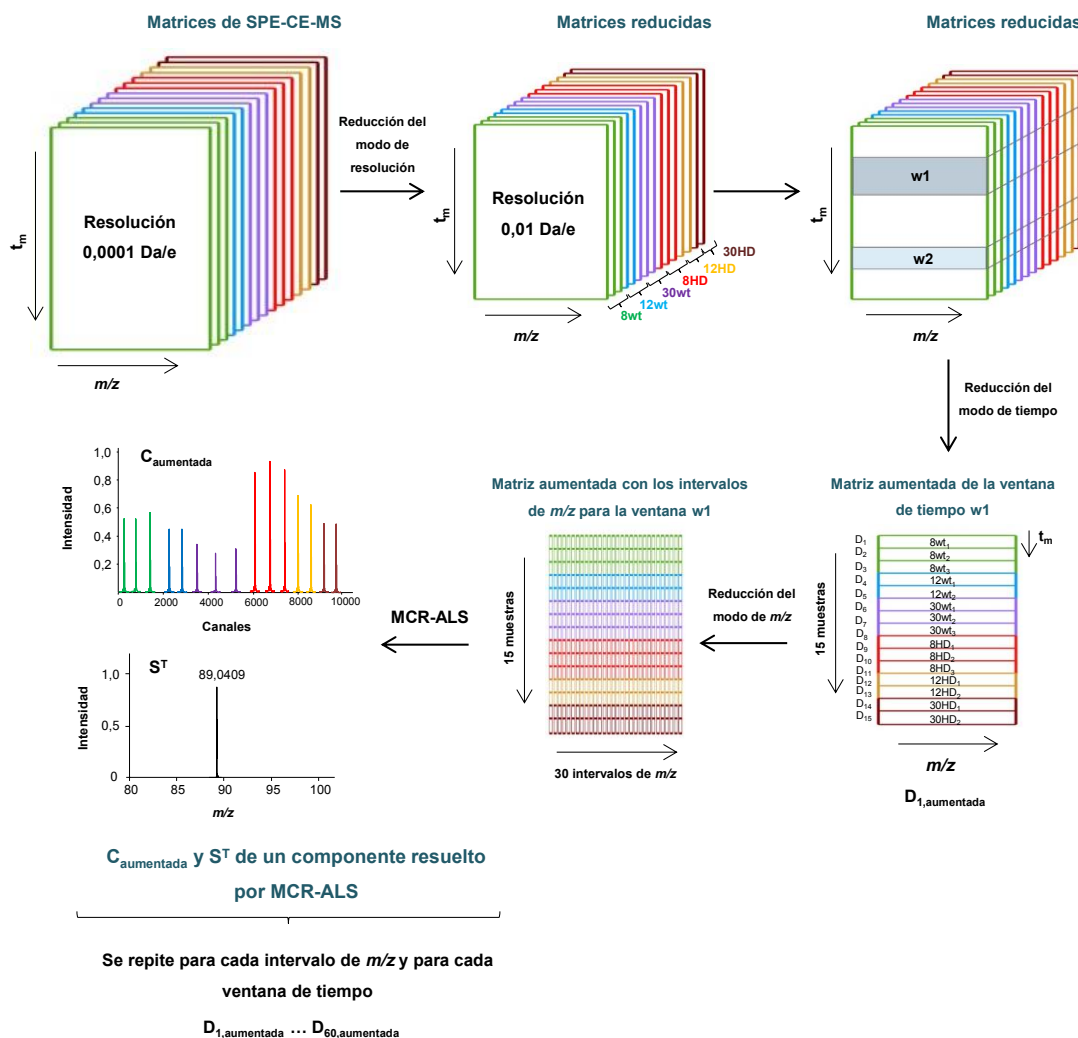


Figura 5.31. Estrategia empleada para el pretratamiento de los datos y el análisis por MCR-ALS.

El método de MCR-ALS se aplicó a cada una de las matrices obtenidas con el objetivo de resolver los perfiles electroforéticos (C_{aug}) y sus espectros de masas (S^T) correspondientes (mirar sección 1.6.1 de la introducción). El número de componentes se escogió visualmente de acuerdo al número de picos electroforéticos resueltos, a pesar de que algunos de los componentes fueron descartados por ser debidos a contribuciones tales como el ruido instrumental. Los espectros de masas (S^T) de los componentes seleccionados se usaron para

identificar las m/z de los iones moleculares de los compuestos mayoritarios detectados en los diferentes componentes. Un total de 74 compuestos (*features*) fueron detectados después de resolver y analizar las 60 matrices aumentadas. Finalmente, y con el objetivo de confirmar la presencia de los 74 *features* en los datos originales, se generaron los EIEs a partir de los datos originales y las m/z candidatas y se llevó a cabo una revisión cuidadosa de los electroferogramas obtenidos.

5.3.3.2. Análisis multivariante. PCA y PLS-DA

Una vez integrados los EIEs obtenidos a partir de los datos originales para los 74 *features* y normalizadas las A_p respecto a un compuesto presente en todas las muestras y que no era importante para discriminar entre plasma de ratones wt y HD (m/z 72,9858), se llevó a cabo un análisis exploratorio por PCA con el objetivo de determinar las clases presentes en las muestras, así como la presencia de *outliers*. Una vez observados por PCA los diferentes grupos, se construyeron diferentes modelos de PLS-DA (con resultados similares a los obtenidos por PCA) con el objetivo de diferenciar entre muestras wt y HD (primer modelo), muestras wt a diferentes semanas de edad (segundo modelo) y muestras HD a diferentes semanas de edad (tercer modelo). Es importante destacar que en el caso del segundo y del tercer modelo, a pesar de disponer de muestras de plasma de ratones wt y HD a 8, 12 y 30 semanas de edad, sólo dos clases (8 y 12-30 semanas) fueron seleccionadas para construir los modelos de PLS-DA, principalmente debido a la cantidad limitada de muestras analizadas (sólo 2 muestras wt a 12 semanas de edad y 2 muestras HD a 12 y 30 semanas de edad). La **Figura 5.32** muestra los gráficos de *scores* para las diferentes muestras de plasma de ratones teniendo en cuenta los tres modelos comentados anteriormente, para los que se seleccionaron dos LVs en cada caso (34% y 89%, 47% y 99%, 52% y 88% de varianzas explicadas en X e Y para el primer, segundo y tercer modelo, respectivamente). Tal y como se puede observar, los tres modelos de PLS-DA permitieron discriminar entre los diferentes grupos de muestras, lo que permitió posteriormente determinar las diferencias debidas a la enfermedad (**Figura 5.32 A**), al envejecimiento en

ratones wt (**Figura 5.32 B**) y a la progresión del HD a diferentes semanas de edad (**Figura 5.32 C**).

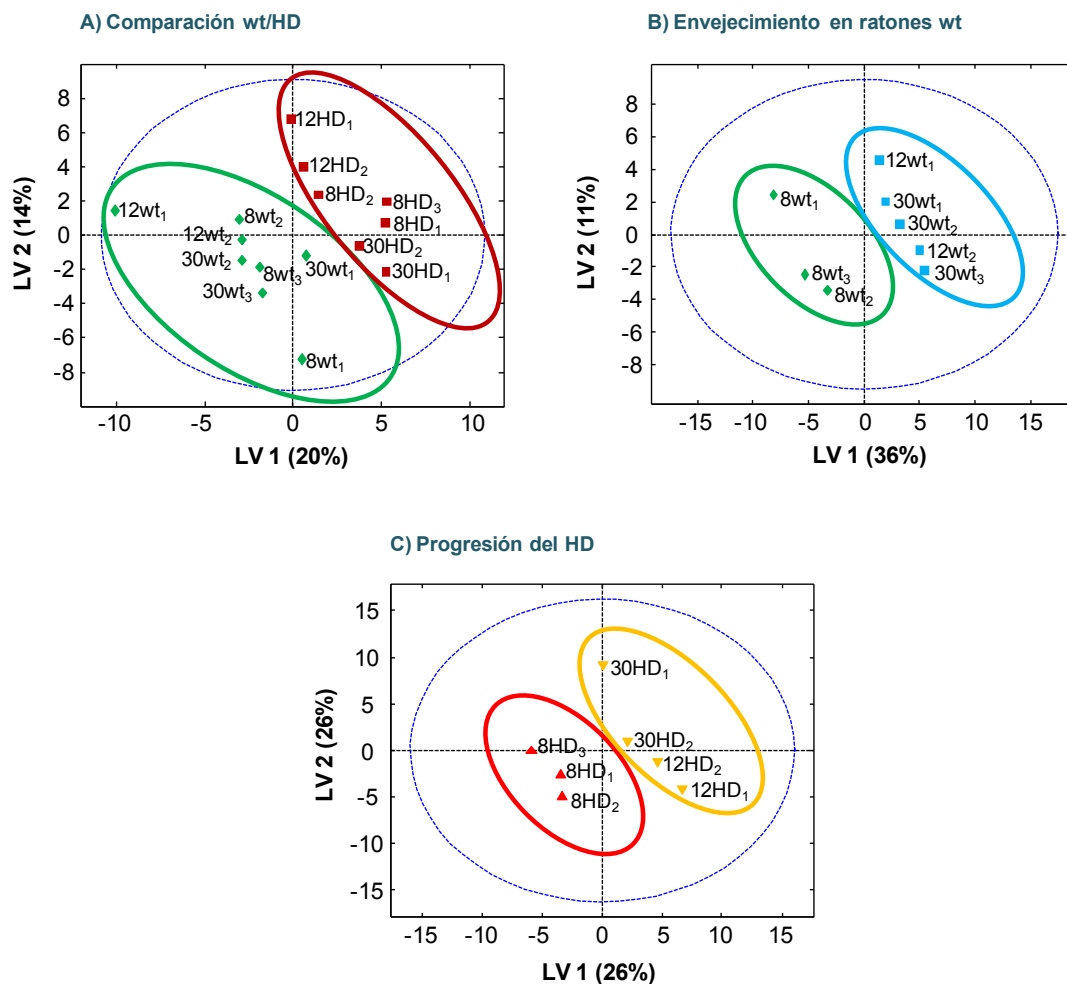


Figura 5.32. Gráficos de *scores* del PLS-DA para las diferentes muestras de plasma de ratones analizadas por C₁₈-SPE-CE-MS con el fin de diferenciar entre (A) muestras wt y HD, (B) muestras wt a diferentes semanas de edad y (C) muestras HD a diferentes semanas de edad.

Una vez contruidos los tres modelos de PLS-DA, las variables o compuestos (*features*) más significativos se seleccionaron teniendo en cuenta las variables importantes en la proyección (VIPs) superiores a 1 para cada uno de los modelos comentados anteriormente [229,248]. Un total de 33 *features* fueron seleccionados por su relevancia para discriminar entre los diferentes

grupos de muestras y para determinar las diferencias debidas a la enfermedad (comparación wt/HD), al envejecimiento en ratones wt y a la progresión del HD a diferentes semanas de edad.

5.3.4. Identificación de biomarcadores metabólicos y rutas metabólicas asociadas al Huntington

Una vez seleccionados los 33 *features* capaces de distinguir entre los diferentes grupos de muestras ($VIP > 1$), se procedió a su identificación empleando bases de datos en línea para la investigación metabolómica, tales como la *METLIN Metabolite Database* [256] y la *Human Metabolome Database* [47]. Esta búsqueda permitió proponer una identidad para 29 de los 33 *features* anteriores. Es conveniente destacar que se consideró que la m/z experimental correspondía a la del metabolito en cuestión si la exactitud de masas (en términos de error relativo, E_r) era igual o inferior a 20 ppm. Tal y como se puede observar en la **Tabla 5.9**, se produjeron ciertas ambigüedades por lo que respecta a la identificación de los compuestos. Por ejemplo, en algunos casos, diferentes metabolitos fueron propuestos para una misma fórmula molecular y M_r teórica (mirar metabolitos con ID 2, 3, 4, 6, 7, 9, 13, 15, 16, 17, 18, 19 y 29). En otros casos, no fue posible diferenciar entre metabolitos con M_r teóricas muy similares (mirar metabolitos con ID 6 ($E_r=7$ y 14 ppm), ID 8 ($E_r=16$ y 18 ppm), ID 9 ($E_r=2$ y 15 ppm), ID 10 ($E_r=13$ y 17 ppm) e ID 25 ($E_r=7$ y 16 ppm)). Por ello, en un futuro, tal y como se mostró para la glicoproteína CEA en el **artículo 3.1** de esta tesis doctoral, se podría confirmar la identidad de dichos metabolitos mediante el uso de patrones adecuados, experimentos de masas en tándem (MS/MS con un analizador híbrido cuadrupolo-tiempo de vuelo (Q-TOF), o de resonancia ciclotrónica de iones con transformada de Fourier (FTICR)) o utilizando modelos semiempíricos de predicción de la migración electroforética.

Tabla 5.9. Ventana de tiempo (min), fórmula molecular, M_r teórica, ión molecular, m/z teórica y experimental y E_r (en ppm) para los metabolitos identificados mediante la combinación de C_{18} -SPE-CE-MS, MCR-ALS, PLS-DA (VIP>1) y búsqueda en bases de datos ($E_r \leq 20$ ppm).

ID	Metabolito	Ventana de tiempo (min)	Fórmula molecular	M_r teórica	Ión molecular	m/z experimental	m/z teórica	E_r (ppm)
1	<i>3-(Methylthio)-1-propene</i>	10-25	C_4H_8S	88,0347	$[M+H]^+$	89,0401	89,0419	20
2	<i>m-Cresol</i>	30-40	C_7H_8O	108,0575	$[M+H]^+$	109,0653	109,0648	5
	<i>p-Cresol</i>		C_7H_8O	108,0575	$[M+H]^+$	109,0653	109,0648	5
3	<i>3-Methyl-1,2-cyclopentanedione</i>	30-40	$C_6H_8O_2$	112,0524	$[M+H]^+$	113,0602	113,0597	4
	<i>2,5-Dimethyl-3(2H)-furanone</i>		$C_6H_8O_2$	112,0524	$[M+H]^+$	113,0602	113,0597	4
	<i>(2E,4E)-2,4-Hexadienoic acid</i>		$C_6H_8O_2$	112,0524	$[M+H]^+$	113,0602	113,0597	4
	<i>(E)-4-Oxo-2-hexen-1-al</i>		$C_6H_8O_2$	112,0524	$[M+H]^+$	113,0602	113,0597	4
	<i>1,2-Cyclohexanedione</i>		$C_6H_8O_2$	112,0524	$[M+H]^+$	113,0602	113,0597	4
	<i>trans-1,2-Dihydrobenzene-1,2-diol</i>		$C_6H_8O_2$	112,0524	$[M+H]^+$	113,0602	113,0597	4
4	<i>2-Hepten-4-one</i>	30-40	$C_7H_{12}O$	112,0888	$[M+H]^+$	113,0968	113,0961	6
	<i>(Z)-4-Heptenal</i>		$C_7H_{12}O$	112,0888	$[M+H]^+$	113,0968	113,0961	6
	<i>3-Hepten-2-one</i>		$C_7H_{12}O$	112,0888	$[M+H]^+$	113,0968	113,0961	6
	<i>2-Methylcyclohexanone</i>		$C_7H_{12}O$	112,0888	$[M+H]^+$	113,0968	113,0961	6
	<i>trans-2-trans-4-Heptadien-1-ol</i>		$C_7H_{12}O$	112,0888	$[M+H]^+$	113,0968	113,0961	6
	<i>5-Methyl-5-hexen-2-one</i>		$C_7H_{12}O$	112,0888	$[M+H]^+$	113,0968	113,0961	6
	<i>4-Methylcyclohexanone</i>		$C_7H_{12}O$	112,0888	$[M+H]^+$	113,0968	113,0961	6
	<i>5-Methyl-3-hexen-2-one</i>		$C_7H_{12}O$	112,0888	$[M+H]^+$	113,0968	113,0961	6
5	<i>Dimethylbenzimidazole</i>	30-40	$C_9H_{10}N_2$	146,0844	$[M+H]^+$	147,0919	147,0917	2
6	<i>Hydantoin-5-propionic acid</i>	30-40	$C_6H_8N_2O_4$	172,0484	$[M+Na]^+$	195,0389	195,0376	7
	<i>Menadione</i>		$C_{11}H_8O_2$	172,0524	$[M+Na]^+$	195,0389	195,0416	14
	<i>Methyl (Z)-2-decene-4,6,8-triynoate</i>		$C_{11}H_8O_2$	172,0524	$[M+Na]^+$	195,0389	195,0416	14

ID	Metabolito	Ventana de tiempo (min)	Fórmula molecular	M _r teórica	Ión molecular	m/z experimental	m/z teórica	E _r (ppm)
7	<i>(-)-Epinephrine</i> <i>Normetanephine</i>	10-25	C ₉ H ₁₃ NO ₃	183,0895	[M+Na] ⁺	206,0747	206,0788	20
			C ₉ H ₁₃ NO ₃	183,0895	[M+Na] ⁺	206,0747	206,0788	20
8	<i>Vanylglycol</i> <i>Phosphorylcholine</i>	10-25	C ₉ H ₁₂ O ₄	184,0736	[M+Na] ⁺	207,0594	207,0628	16
			C ₅ H ₁₅ NO ₄ P	184,0739	[M+Na] ⁺	207,0594	207,0631	18
9	<i>3-Indolebutyric acid</i> <i>Glycyl-glutamine</i> <i>Glycyl-gamma-glutamate</i> <i>Asparaginyl-alanine</i> <i>Glutaminy-glycine</i> <i>Alanyl-asparagine</i> <i>Gamma-glutamyl-glycine</i>	10-25	C ₁₂ H ₁₃ NO ₂	203,0946	[M+Na] ⁺	226,0833	226,0838	2
			C ₇ H ₁₃ N ₃ O ₄	203,0906	[M+Na] ⁺	226,0833	226,0798	15
			C ₇ H ₁₃ N ₃ O ₄	203,0906	[M+Na] ⁺	226,0833	226,0798	15
			C ₇ H ₁₃ N ₃ O ₄	203,0906	[M+Na] ⁺	226,0833	226,0798	15
			C ₇ H ₁₃ N ₃ O ₄	203,0906	[M+Na] ⁺	226,0833	226,0798	15
			C ₇ H ₁₃ N ₃ O ₄	203,0906	[M+Na] ⁺	226,0833	226,0798	15
			C ₇ H ₁₃ N ₃ O ₄	203,0906	[M+Na] ⁺	226,0833	226,0798	15
10	<i>Propyl propane thiosulfonate</i> <i>5-Methoxycanthin-6-one</i>	30-40	C ₁₁ H ₁₆ O ₅	228,0998	[M+Na] ⁺	251,0857	251,0890	13
			C ₁₅ H ₁₀ N ₂ O ₂	250,0742	[M+H] ⁺	251,0857	251,0815	17
11	<i>L-Hexanoylcarnitine</i>	30-40	C ₁₃ H ₂₅ NO ₄	259,1784	[M+H] ⁺	260,1858	260,1858	0
12	<i>Histidinyl-histidine</i>	30-40	C ₁₂ H ₁₆ N ₆ O ₃	292,1284	[M+H] ⁺	293,1327	293,1357	10
13	<i>Phenylalanyl-arginine</i> <i>Arginyl-phenylalanine</i>	10-25	C ₁₅ H ₂₃ N ₅ O ₃	321,1801	[M+H] ⁺	322,1920	322,1873	14
			C ₁₅ H ₂₃ N ₅ O ₃	321,1801	[M+H] ⁺	322,1920	322,1873	14
14	<i>2-[Octahydro-4,7-dimethyl-1-oxocyclopenta[c]pyran-3-yl]nepetalactam</i>	30-40	C ₂₀ H ₂₉ NO ₃	331,2147	[M+H] ⁺	332,2186	332,2220	10
15	<i>Prostaglandin D2</i> <i>Prostaglandin E2</i> <i>Prostaglandin H2</i> <i>Prostaglandin I2</i> <i>Thromboxane A2</i> <i>Lipoxin A4</i>	10-25	C ₂₀ H ₃₂ O ₅	352,2250	[M+Na] ⁺	375,2097	375,2142	12
			C ₂₀ H ₃₂ O ₅	352,2250	[M+Na] ⁺	375,2097	375,2142	12
			C ₂₀ H ₃₂ O ₅	352,2250	[M+Na] ⁺	375,2097	375,2142	12
			C ₂₀ H ₃₂ O ₅	352,2250	[M+Na] ⁺	375,2097	375,2142	12
			C ₂₀ H ₃₂ O ₅	352,2250	[M+Na] ⁺	375,2097	375,2142	12
			C ₂₂ H ₂₈ N ₂ O ₂	352,2250	[M+Na] ⁺	375,2097	375,2142	12

ID	Metabolito	Ventana de tiempo (min)	Fórmula molecular	M _r teórica	Ión molecular	m/z experimental	m/z teórica	E _r (ppm)
15	<i>Lipoxin B4</i> <i>1520-Hydroxy-leukotriene B4</i> <i>15-Keto-prostaglandin F2a</i>	10-25	C ₂₀ H ₃₂ O ₅	352,2250	[M+Na] ⁺	375,2097	375,2142	12
			C ₂₀ H ₃₂ O ₅	352,2250	[M+Na] ⁺	375,2097	375,2142	12
			C ₂₀ H ₃₂ O ₅	352,2250	[M+Na] ⁺	375,2097	375,2142	12
16	<i>18-Hydroxycorticosterone</i> <i>Cortisol</i>	10-25	C ₂₁ H ₃₀ O ₅	362,2093	[M+H] ⁺	363,2100	363,2166	18
			C ₂₁ H ₃₀ O ₅	362,2093	[M+H] ⁺	363,2100	363,2166	18
17	<i>alpha-Peroxyachifolide</i> <i>Diosbulbin F</i> <i>Diosbulbin A</i> <i>Carissanol</i>	10-25	C ₂₀ H ₂₄ O ₇	376,1522	[M+H] ⁺	377,1531	377,1595	17
			C ₂₀ H ₂₄ O ₇	376,1522	[M+H] ⁺	377,1531	377,1595	17
			C ₂₀ H ₂₄ O ₇	376,1522	[M+H] ⁺	377,1531	377,1595	17
			C ₂₀ H ₂₄ O ₇	376,1522	[M+H] ⁺	377,1531	377,1595	17
18	<i>PGD2-Dihydroxypropylamine</i> <i>PGE2-Dihydroxypropylamine</i>	10-25	C ₂₃ H ₃₉ NO ₆	425,2778	[M+Na] ⁺	448,2600	448,2670	16
			C ₂₃ H ₃₉ NO ₆	425,2778	[M+Na] ⁺	448,2600	448,2670	16
19	<i>Artonol E</i> <i>Cycloartomunoxanthone</i>	10-25	C ₂₆ H ₂₄ O ₇	448,1522	[M+H] ⁺	449,1605	449,1595	2
			C ₂₆ H ₂₄ O ₇	448,1522	[M+H] ⁺	449,1605	449,1595	2
20	<i>6-Deoxohomodolichosterone</i>	10-25	C ₂₉ H ₅₀ O ₄	462,3709	[M+Na] ⁺	485,3584	485,3601	4
21	<i>Limocitrin 3-glucoside</i>	10-25	C ₂₃ H ₂₄ O ₁₃	508,1217	[M+Na] ⁺	531,1156	531,1109	9
22	<i>Biochanin A 7-(6-malonylglucoside)</i>	10-25	C ₂₅ H ₂₄ O ₁₃	532,1217	[M+H] ⁺	533,1183	533,1290	20
23	<i>Maclurin 3-C-(6''-p-hydroxybenzoyl-glucoside)</i>	10-25	C ₂₆ H ₂₄ O ₁₃	544,1217	[M+H] ⁺	545,1295	545,1289	1
24	<i>L-Urobilinogen</i>	10-25	C ₃₃ H ₄₈ N ₄ O ₆	596,3574	[M+H] ⁺	597,3724	597,3647	13
25	<i>Tsugarioside B</i> <i>Oryzanol C</i>	30-40	C ₃₇ H ₆₀ O ₇	616,4339	[M+2Na] ²⁺	331,2084	331,2062	7
			C ₄₁ H ₆₀ O ₄	616,4492	[M+2Na] ²⁺	331,2084	331,2138	16
26	<i>PC(14:1(9Z)/14:1(9Z))</i>	10-25	C ₃₆ H ₆₈ NO ₈ P	673,4683	[M+2Na] ²⁺	359,7223	359,7233	3
27	<i>Nostoxanthin sulfate</i>	10-25	C ₄₀ H ₅₅ NaO ₇ S	702,3566	[M+H] ⁺	703,3700	703,3639	10
28	<i>PI(13:0/12:0)</i>	10-25	C ₃₄ H ₆₅ O ₁₃ P	712,4164	[M+2Na] ²⁺	379,1987	379,1974	3
29	<i>Ganglioside GD1b (d18:1/12:0)</i> <i>Ganglioside GD1a (d18:1/12:0)</i>	30-40	C ₇₉ H ₁₃₇ N ₃ O ₃₉	1751,8829	[M+2Na] ²⁺	898,9184	898,9307	14
			C ₇₉ H ₁₃₇ N ₃ O ₃₉	1751,8829	[M+2Na] ²⁺	898,9184	898,9307	14

Una vez identificados los metabolitos responsables de la diferenciación entre los diferentes grupos de muestras (wt/HD, envejecimiento wt y progresión HD), se procedió a investigar las posibles rutas metabólicas asociadas a dichos metabolitos. Para ello, se llevó a cabo una búsqueda exhaustiva en la *Kyoto Encyclopedia of Genes and Genomes (KEGG Database)* [48], pudiéndose encontrar rutas metabólicas asociadas a 13 de los 29 metabolitos identificados (**Tabla 5.10**). Es importante destacar que la mayoría de rutas metabólicas del HD encontradas en este trabajo están asociadas a mecanismos de señalización celular, lo que pone de manifiesto la implicación de estos mecanismos en los diferentes procesos patológicos que tienen lugar, tales como disfunciones neuronales, pérdida de diferentes tipos de receptores y alteraciones en la expresión de neurotransmisores y proteínas [78,79].

Tabla 5.10. Rutas metabólicas asociadas a 13 de los metabolitos identificados mediante la combinación de C₁₈-SPE-CE-MS, MCR-ALS, PLS-DA (VIP>1) y búsqueda en bases de datos. (**cAMP**, adenosín monofosfato cíclico; **TRP**, receptores de potencial transitorio; **VEGF**, factor de crecimiento endotelial vascular).

ID	Metabolito	Rutas metabólicas
2	<i>m-Cresol</i> <i>p-Cresol</i>	Degradación de compuestos aromáticos
		Degradación de compuestos aromáticos / Digestión y absorción de proteínas
5	<i>Dimethylbenzimidazole</i>	Metabolismo de la riboflavina / Metabolismo de la porfirina
7	<i>(-)-Epinephrine</i> <i>Normetanephine</i>	Señalización del cAMP* / Señalización adrenérgica de los cardiomiocitos* / Interacción ligando-receptor* / Metabolismo de la tirosina
		Metabolismo de la tirosina
8	<i>Vanylglycol</i> <i>Phosphorylcholine</i>	Metabolismo de la tirosina
		Metabolismo de los glicerosfolípidos
9	<i>3-Indolebutyric acid</i> <i>Glycyl-glutamine</i> <i>Glycyl-gamma-glutamate</i> <i>Asparaginyal-alanine</i> <i>Glutaminyl-glycine</i> <i>Alanyl-asparagine</i> <i>Gamma-glutamyl-glycine</i>	Metabolismo del triptófano
		Producto incompleto del catabolismo y de la digestión de las proteínas con efectos de señalización celular *
		Producto incompleto del catabolismo y de la digestión de las proteínas con efectos de señalización celular *
		Producto incompleto del catabolismo y de la digestión de las proteínas con efectos de señalización celular *
		Producto incompleto del catabolismo y de la digestión de las proteínas con efectos de señalización celular *
		Producto incompleto del catabolismo y de la digestión de las proteínas con efectos de señalización celular *
11	<i>L-Hexanoylcarnitine</i>	Producción de energía
12	<i>Histidinyl-histidine</i>	Producto incompleto del catabolismo y de la digestión de las proteínas con efectos de señalización celular *
13	<i>Phenylalanyl-arginine</i> <i>Arginyl-phenylalanine</i>	Producto incompleto del catabolismo y de la digestión de las proteínas con efectos de señalización celular *
		Producto incompleto del catabolismo y de la digestión de las proteínas con efectos de señalización celular *
15	<i>Prostaglandin D2</i> <i>Prostaglandin E2</i> <i>Prostaglandin H2</i> <i>Prostaglandin I2</i> <i>Thromboxane A2</i>	Metabolismo del ácido araquidónico* / Interacción ligando-receptor* / Sinapsis serotoninérgica* / Señalización del Fc epsilon*
		Metabolismo del ácido araquidónico* / Interacción ligando-receptor* / Sinapsis serotoninérgica* / Señalización del cAMP* / Señalización de la oxitocina* / Regulación de los canales TRP*
		Metabolismo del ácido araquidónico* / Sinapsis serotoninérgica* / Señalización retrógrada endocannabinoide* / Señalización de la oxitocina*
		Metabolismo del ácido araquidónico* / Interacción ligando-receptor* / Señalización del cAMP* / Señalización VEGF*
		Metabolismo del ácido araquidónico* / Interacción ligando-receptor* / Sinapsis serotoninérgica*

ID	Metabolito	Rutas metabólicas
15	<i>Lipoxin A4</i> <i>Lipoxin B4</i> <i>20-Hydroxy-leukotriene B4</i> <i>15-Keto-prostaglandin F2a</i>	Metabolismo del ácido araquidónico* / Interacción ligando-receptor*
		Metabolismo del ácido araquidónico*
		Metabolismo del ácido araquidónico*
		Metabolismo del ácido araquidónico*
16	<i>18-Hydroxycorticosterone</i> <i>Cortisol</i>	Biosíntesis de las hormonas esteroides
		Biosíntesis de las hormonas esteroides
24	<i>L-Urobilinogen</i>	Metabolismo de la porfirina
26	<i>PC(14:1(9Z)/14:1(9Z))</i>	Metabolismo del ácido araquidónico* / Señalización retrógrada endocannabinoide* / Metabolismo de los glicerofosfolípidos / Metabolismo del ácido linoleico
29	<i>Ganglioside GD1b (d18:1/12:0)</i> <i>Ganglioside GD1a (d18:1/12:0)</i>	Transducción celular*
		Transducción celular*

*Rutas metabólicas relacionadas con mecanismos de señalización celular.

La **Figura 5.33** muestra las relaciones entre las abundancias relativas de los diferentes metabolitos identificados (VIP>1) con el objetivo de interpretar las diferencias entre las muestras de ratones HD a diferentes semanas de edad (progresión HD), muestras de ratones wt a diferentes semanas de edad (envejecimiento wt) y, en general, entre muestras wt y HD sin tener en cuenta la progresión de la edad (wt/HD).

Progresión del HD (Figura 5.33 A)

El diagrama de Venn que aparece en la **Figura 5.33 A** muestra que siete (3+4) de los 29 metabolitos identificados en este trabajo con VIP>1 son específicos para explicar la progresión del HD (progresión HD, metabolitos con ID 1, 10, 28, 3, 13, 15 y 24). Tal y como se observa en la figura, la concentración de fenilalanil-arginina y arginil-fenilalanina (ID 13) aumentó en ratones HD después de 12 semanas de edad. Estos metabolitos, los cuales presentan efectos de señalización celular, son productos incompletos del catabolismo y de la digestión de las proteínas (**Tabla 5.10**) [257]. De manera similar, la concentración de prostaglandinas, tromboxanos, lipoxinas y leucotrienos (ID 15) aumentó en ratones HD a 12 semanas de edad. Estos metabolitos están relacionados con la regulación de procesos inflamatorios, la interacción ligando-receptor, la sinapsis serotoninérgica, la señalización del adenosín monofosfato cíclico (cAMP) y la señalización de la oxitocina (**Tabla 5.10**). Finalmente, los niveles de L-urobilinógeno (ID 24), metabolito relacionado con el metabolismo de la porfirina (**Tabla 5.10**), también aumentaron en ratones HD después de 12 semanas de edad. Todos estos cambios a 12 semanas de edad sugieren un punto de inflexión para la aparición de síntomas de disfunción neuronal, así como de alteración en la expresión de receptores y neurotransmisores.

Progresión HD + Envejecimiento wt (Figura 5.33 A-B)

Los diagramas de Venn que aparecen en la **Figura 5.33 A y B** muestran que 8 (1+7) de los 29 metabolitos identificados en este trabajo con VIP>1 sirven para explicar, tanto la progresión del

HD como el envejecimiento en muestras de ratones wt (metabolitos con ID 29, 4, 11, 26, 7, 8, 9 y 19). Cuatro de estos metabolitos presentaron tendencias de concentración diferentes entre muestras de ratones HD y wt a diferentes semanas de edad (metabolitos con ID 29, 4, 11 y 26), mientras que los otros cuatro mostraron tendencias similares (metabolitos con ID 7, 8, 9 y 19).

Tendencias de concentración diferentes

Tal y como se observa en la **Figura 5.33 A y B**, los gangliósidos (ID 29), componentes del plasma celular encargados de modular señales de transducción celular (**Tabla 5.10**), mostraron una tendencia claramente diferente entre muestras de ratones HD y wt a diferentes semanas de edad. Específicamente, la concentración de gangliósidos se redujo después de 12 semanas en ratones HD, mientras que sus niveles aumentaron en ratones wt. De manera similar, la concentración de L-hexanoilcarnitina (ID 11) se redujo en ratones wt, mientras que su concentración aumentó después de 12 semanas en ratones HD. Este hecho sugiere que HD conlleva alteraciones en la producción de energía, las cuales se caracterizan por la producción y excreción inusual de acilcarnitinas (**Tabla 5.10**) [258]. Finalmente, la concentración de PC(14:1(9Z)/14:1(9Z)) (ID 26), metabolito relacionado con el metabolismo de los glicerofosfolípidos y del ácido linoleico, así como con rutas metabólicas de señalización celular (**Tabla 5.10**), aumentó en ratones HD después de 12 semanas de edad, mientras que se redujo en ratones wt.

Tendencias de concentración similares

La concentración de epinefrina y normetanefrina (ID 7), metabolitos relacionados con el metabolismo de la tirosina y con rutas metabólicas de señalización celular (**Tabla 5.10**), se redujo a las 12 semanas de edad, tanto en ratones HD como en ratones wt. Se observó la misma tendencia para el vanilglicol y la fosforilcolina (ID 8), metabolitos relacionados con el metabolismo de la tirosina y de los glicerofosfolípidos (**Tabla 5.10**), y los metabolitos con ID 9.

En este caso, el ácido 3-indolbutírico está relacionado con el metabolismo del triptófano, mientras que los otros metabolitos son productos incompletos del catabolismo y de la digestión de proteínas relacionados con efectos de señalización celular (**Tabla 5.10**) [259].

Envejecimiento wt (Figura 5.33 B)

El diagrama de Venn que aparece en la **Figura 5.33 B** muestra que ocho (0+8) de los 29 metabolitos identificados en este trabajo con $VIP > 1$ son específicos para explicar el envejecimiento en ratones wt a diferentes semanas de edad (mirar metabolitos con ID 5, 6, 14, 16, 20, 21, 22, y 23). Tal y como se puede observar en la figura, la concentración de dimetilbenzimidazol (ID 5), metabolito relacionado con el metabolismo de la riboflavina y de la porfirina (**Tabla 5.10**), se redujo después de 12 semanas de edad. La misma tendencia fue observada para la 18-hidroxycorticosterona y el cortisol (ID 16), metabolitos relacionados con la biosíntesis de las hormonas esteroideas (**Tabla 5.10**).

wt/HD (Figura 5.33 C)

El diagrama de Venn que aparece en la **Figura 5.33 C** muestra que diez (6+3+1+0) de los 29 metabolitos identificados en este trabajo con $VIP > 1$ sirven para diferenciar entre ratones wt y HD, sin tener en cuenta la progresión de la edad (mirar metabolitos con ID 2, 12, 18, 25, 17, 27, 10, 15, 1 y 11). Tal y como se puede observar en esta figura, la concentración de m-cresol y p-cresol (ID 2), metabolitos relacionados con la absorción y digestión de las proteínas, así como con la degradación de compuestos aromáticos (**Tabla 5.10**), se redujo en muestras de ratones HD. La misma tendencia fue observada para la histidinil-histidina (ID 12), producto incompleto del catabolismo y de la digestión de las proteínas (**Tabla 5.10**) [260,261], y para las prostaglandinas, tromboxanos, lipoxinas y leucotrienos (ID 15). En el caso de la L-hexanoilcarnitina (ID 11), su concentración aumentó en muestras de ratones HD ya que, tal y

como se ha comentado anteriormente, el HD conlleva alteraciones en la producción de energía caracterizadas por la producción y excreción inusual de acilcarnitinas (**Tabla 5.10**) [258].

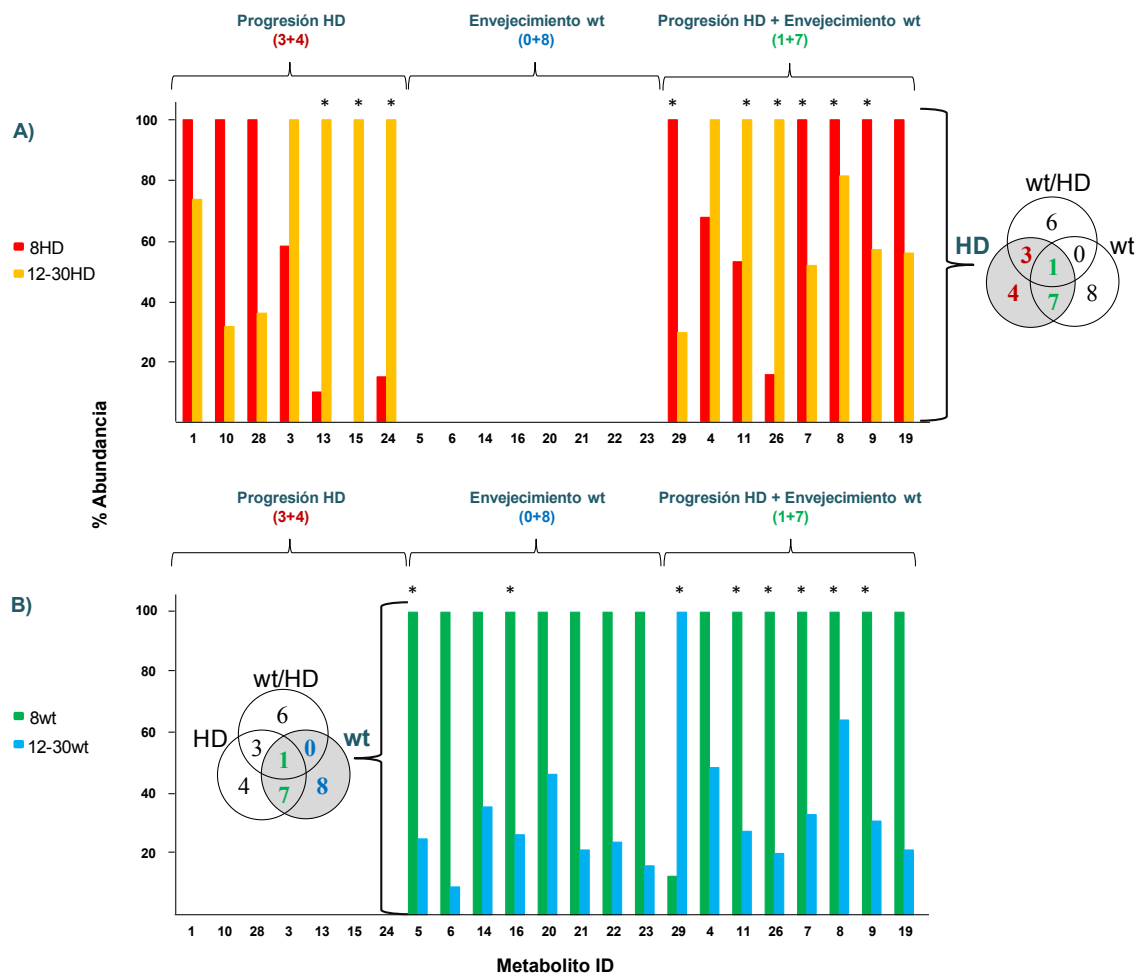


Figura 5.33 A y B. Relaciones entre los diferentes metabolitos identificados (VIP>1) para diferenciar entre (A) progresión del HD y (B) envejecimiento wt. La abundancia para cada metabolito se calculó normalizando respecto al mayor valor de área medio obtenido entre las muestras a 8 y 12-30 semanas de edad. *Metabolitos para los que se han identificado rutas metabólicas (**Tabla 5.10**).

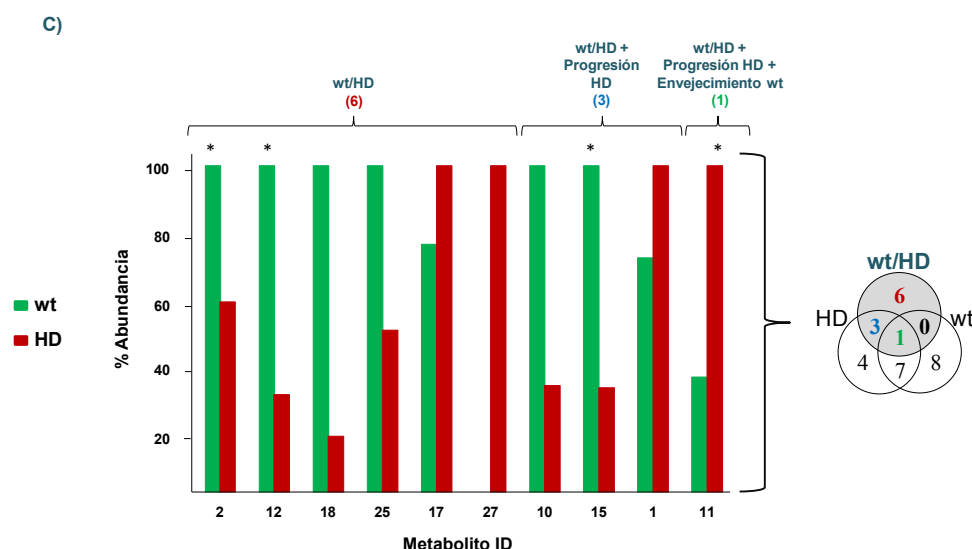


Figura 5.33 C (continúa). Relaciones entre los diferentes metabolitos identificados (VIP>1) para diferenciar entre (C) muestras wt y HD (wt/HD). La abundancia para cada metabolito se calculó normalizando respecto al mayor valor de área medio obtenido entre las muestras wt y HD. *Metabolitos para los que se han identificado rutas metabólicas (Tabla 5.10).

Es importante destacar que algunos metabolitos identificados en este trabajo (**artículo 4.2**) ya habían sido previamente descritos en el cerebro de algunos modelos de ratones, especialmente las prostaglandinas, tromboxanos, lipoxinas y leucotrienos (ID 15, **Tabla 5.9** y **5.10**) [262–265]. Sin embargo, la presencia de estos metabolitos relacionados con mecanismos de señalización celular no había sido demostrada en el plasma de ratones modificados genéticamente para desarrollar el HD. Este trabajo basado en una estrategia de análisis no dirigido (*untargeted analysis*) ha permitido proponer nuevos biomarcadores de HD directamente relacionados con productos de disfunción neuronal, los cuales podrían ser de gran utilidad para diagnosticar y predecir un debut temprano de esta enfermedad, así como para proponer nuevas dianas terapéuticas que detengan su progresión. La información obtenida con esta metodología podría integrarse con la proporcionada por estudios proteómicos basados en estrategias de análisis dirigido (*targeted analysis*), por ejemplo aquéllos enfocados al análisis de la huntingtina, con el objetivo de profundizar en el estudio de las relaciones entre las proteínas y el metabolismo para entender y detener la progresión del HD.

Conclusions

- A novel CE-MS methodology was developed to analyze TTR in standard solutions with several PTMs, and an extra minor isoform ((10) C-G) using acidic and neutral BGEs. With both BGEs, the native tetrameric structure of TTR was disrupted because of the acidic sheath liquid used with the sheath flow CE-MS interface and the high vacuum inside the mass spectrometer.

- A novel CapLC-MS methodology was also developed for the analysis of TTR in standard solutions. Under optimized conditions, CE-MS and CapLC-MS showed a similar performance. However, CE-MS was proposed as the best alternative, mainly because of the slightly lower LODs, the use of an organic solvent-free BGE and the lower cost of the analyses.

- Two different off-line IP procedures were developed for the purification of TTR from serum samples under denaturing conditions (acidic pHs) before the analysis by CE-MS and CapLC-MS: conventional IP in solution and IP with UAPA magnetic beads.
 - The conventional IP in solution allowed detecting the main normal (TTR-Cys and Free-TTR) and mutant (TTR(Met30)-Cys and Free-TTR(Met30)) monomeric TTR proteoforms.
 - The IP with UAPA magnetic beads enhanced the previous method based on conventional IP in solution and allowed the detection of normal monomeric TTR proteoforms found at lower concentrations (TTR-Phosphorylated or TTR-Sulfonated, TTR-Dehydroxylated or TTR-Sulfinic, (10) C-G and TTR-CysGly).
 - With both IP procedures, no mutant TTR proteoforms were detected in the control samples and in the liver transplanted patient, hence suggesting the treatment effectiveness.
 - The levels of TTR-Cys and especially, TTR(Met30)-Cys, in serum samples from FAP-I symptomatic patients were found elevated when compared to the normal and mutant Free-TTR proteoforms. This fact pointed out the importance of TTR oxidized proteoforms as biomarkers of the oxidative stress involved in amyloid deposit formation in FAP-I.

- A novel IM-MS methodology was developed for the analysis of TTR in standard solutions and serum samples under non-denaturing conditions.

- The proposed IM-MS methodology allowed detecting MO and TE ions in TTR standard solutions. However, the resolution of the mass spectrometer at the measured m/z was not enough to resolve the TTR proteoforms.
- For the analysis of TTR from serum samples, a sample pretreatment based on acid-free IP (pH=11.30) in solution was firstly developed. Under optimized conditions, the IM-MS methodology allowed the detection of DI, TRI and TE ions.
- A comparison between TE/DI and TE/TRI abundance ratios in serum samples from healthy controls and FAP-I patients demonstrated the presence of a higher abundance and stability of the TE in the FAP-I patients, probably due to the presence in patients of a higher amount of the oxidized TTR-Cys and TTR(Met30)-Cys proteoforms.
- The TE/DI and TE/TRI abundance ratios in the liver transplanted patient were similar to those found in healthy controls, hence confirming the treatment effectiveness.

- An IA-SPE-CE-MS methodology with an IA sorbent with Fab' antibody fragments was developed for the analysis of TTR in standard solutions and serum samples.

- Due to the poor recoveries at acidic pHs and the limited stability of the Fab' fragments and the succinimidyl silica particles at alkaline pHs, TTR elution was only possible at certain conditions (pH=9.50) where TTR was detected as different conformers.
- For TTR standard solutions, repeatability and reproducibility were acceptable, microcartridges lifetime was good (>20 analyses in consecutive days), LODs were 50 and 2-fold lower compared to the LODs obtained using CE-MS (~25 µg/mL)

and the intact antibody IA sorbent method developed in a previous work (~1 µg/mL) and different TTR conformers were detected (folded and unfolded).

- A simple off-line pretreatment based on protein precipitation with 5% (v/v) of phenol was necessary before the Fab'-IA-SPE-CE-MS analysis of TTR in serum samples to avoid microcartridge saturation.
- The proposed Fab'-IA-SPE-CE-MS methodology enhanced our previous results for the analysis of sera TTR using intact antibodies. It allowed the detection of MO proteoforms found at lower concentrations (TTR-CysGly, TTR-Glutathione and TTR-CysGlu) and the detection of different unfolded TTR conformations.
- A differential TTR conformer profile was observed between FAP-I patients and control sera, which could be related to differences in TTR proteoforms.

- A novel sheathless CE-MS/MS methodology was developed for the *N*-glycopeptide analysis of CEA samples retrieved from different sources: human colon carcinoma and human liver metastases.

- The information obtained using specific proteases (such as trypsin and endoproteinase Glu-C) and non-specific proteases (such as pronase) was combined in order to maximize the *N*-glycosylation site coverage.
- The proposed methodology allowed the characterization of 20 out of the 28 potential *N*-glycosylation sites. In total, 219 different *N*-glycopeptide glycoforms were identified.
- A targeted multivariate data analysis approach based on PCA and PLS-DA was applied to differentiate the characteristic glycopeptide maps of CEA samples from human colon carcinoma and human liver metastases. A hundred and twenty one *N*-glycopeptide glycoforms were found to be the most significant to discriminate between the two CEA types, being CEA samples from human liver metastases more fucosylated and sialylated.

- A sample pretreatment was optimized for the analysis of low molecular mass compounds in human plasma samples before C₁₈-SPE-CE-MS. The best recoveries for three model neuropeptides (Dyn A, End 1 and Met) were obtained after applying a precipitation step with ACN (6:1 (v/v), ACN:plasma) followed by ultrafiltration with Amicon Ultra-0.5 10,000 M_r cut-off cellulose acetate filters passivated with 5% (v/v) PEG in water. Under the optimized conditions, LODs for the studied peptides by C₁₈-SPE-CE-MS in plasma samples decreased down to 10 ng/mL for Dyn A and 1 ng/mL for End 1 and Met (100- and 1000-fold the LODs obtained using CE-MS).

- The optimized C₁₈-SPE-CE-MS methodology described in the previous study was combined with multivariate data analysis tools (MCR-ALS, PCA and PLS-DA) for the identification of biomarkers of Huntington's disease progression in mice plasma.

- The optimized sample pretreatment was applied to plasma samples from wt and HD mice of different ages (8, 12 and 30 weeks) prior analysis by C₁₈-SPE-CE-MS.
- The application of MCR-ALS to the data matrices generated by C₁₈-SPE-CE-MS allowed the resolution of electrophoretic peaks and mass spectra of a large number of low molecular mass compounds (74 features).
- A multivariate data analysis approach based on PCA and PLS-DA was applied to easily differentiate between wt and HD samples, as well as to follow-up the HD progression. Using the VIP method, 33 compounds were found to be the most significant to discriminate between the groups of samples.
- A list of potential metabolites useful to discriminate between control and HD plasma samples, as well as to follow up the HD progression, were tentatively identified using different on-line databases (such as METLIN and HMD). A total number of 29 metabolites were identified, and the most affected metabolic pathways were also searched against the KEGG Database, being the intracellular signaling the most affected metabolic pathway in HD mice after 12 weeks of birth.

References

- [1] R.P. Horgan, L.C. Kenny, SAC review “omic” technologies: proteomics and metabolomics, *Obstet. Gynaecol.* 13 (2011) 189–195.
- [2] C.E. Mason, S.G. Porter, T.M. Smith, Characterizing multi-omic data in systems biology, in: *Adv. Exp. Med. Biol.* (2013) 15–38.
- [3] D.L. Gerhold, R.V. Jensen, S.R. Gullans, Better therapeutics through microarrays, *Nat. Genet.* 32 (2002) 547–551.
- [4] W.E. Evans, M.V. Relling, Moving towards individualized medicine with pharmacogenomics, *Nature.* 429 (2004) 464–468.
- [5] D.B. Kell, Systems biology, metabolic modelling and metabolomics in drug discovery and development, *Drug Discov. Today.* 11 (2006) 1085–1092.
- [6] D.P. Bartel, MicroRNAs: genomics, biogenesis, mechanism, and function, *Cell.* 116 (2004) 281–297.
- [7] O. Morozova, M.A. Marra, Applications of next-generation sequencing technologies in functional genomics, *Genomics.* 92 (2008) 255–264.
- [8] Z. Wang, M. Gerstein, M. Snyder, RNA-seq: a revolutionary tool for transcriptomics, *Nat. Rev. Genet.* 10 (2009) 57–63.
- [9] P.A. McGettigan, Transcriptomics in the RNA-seq era, *Curr. Opin. Chem. Biol.* 17 (2013) 4–11.
- [10] Z. Zhang, S. Wu, D.L. Stenoien, L. Paša-Tolić, High-throughput proteomics, *Annu. Rev. Anal. Chem.* 7 (2014) 427–454.
- [11] T.K. Toby, L. Fornelli, N.L. Kelleher, Progress in top-down proteomics and the analysis of proteoforms, *Annu. Rev. Anal. Chem.* 9 (2016) 499–519.
- [12] L.M. Smith, N.L. Kelleher, Proteoform: a single term describing protein complexity, *Nat. Methods.* 10 (2013) 186–187.
- [13] M. Mann, O.N. Jensen, Proteomic analysis of post-translational modifications, *Nat. Biotechnol.* 21 (2003) 255–261.
- [14] C.T. Walsh, S. Garneau-Tsodikova, G.J. Gatto, Protein posttranslational modifications: the chemistry of proteome diversifications, *Angew. Chemie.* 44 (2005) 7342–7372.

- [15] E.S. Witze, W.M. Old, K.A. Resing, N.G. Ahn, Mapping protein post-translational modifications with mass spectrometry, *Nat. Methods*. 4 (2007) 798–806.
- [16] H. Johnson, C.E. Eyers, Analysis of post-translational modifications by LC-MS/MS, *Methods Mol. Biol.* 658 (2010) 93–108.
- [17] J.R. Yates, C.I. Ruse, A. Nakorchevsky, Proteomics by mass spectrometry: approaches, advances, and applications, *Annu. Rev. Biomed. Eng.* 11 (2009) 49–79.
- [18] A. Armirotti, Bottom-Up proteomics, *Curr. Anal. Chem.* 5 (2009) 116–130.
- [19] Y. Zhang, B.R. Fonslow, B. Shan, M.C. Baek, J.R. Yates, Protein analysis by shotgun/bottom-up proteomics, *Chem. Rev.* 113 (2013) 2343–2394.
- [20] J. Mayne, Z. Ning, X. Zhang, A.E. Starr, R. Chen, S. Deeke, C.K. Chiang, B. Xu, M. Wen, K. Cheng, D. Seebun, A. Star, J.I. Moore, D. Figeys, Bottom-up proteomics (2013–2015): keeping up in the era of systems biology, *Anal. Chem.* 88 (2016) 95–121.
- [21] A. Armirotti, G. Damonte, Achievements and perspectives of top-down proteomics, *Proteomics*. 10 (2010) 3566–3576.
- [22] D. Calligaris, C. Villard, D. Lafitte, Advances in top-down proteomics for disease biomarker discovery, *J. Proteomics*. 74 (2011) 920–934.
- [23] H. Zhou, Z. Ning, A.E. Starr, M. Abu-Farha, D. Figeys, Advancements in top-down proteomics, *Anal. Chem.* 84 (2012) 720–734.
- [24] A.D. Catherman, O.S. Skinner, N.L. Kelleher, Top down proteomics: facts and perspectives, *Biochem. Biophys. Res. Commun.* 445 (2014) 683–693.
- [25] M.S. Monteiro, M. Carvalho, M.L. Bastos, P. Guedes de Pinho, Metabolomics analysis for biomarker discovery: advances and challenges, *Curr. Med. Chem.* 20 (2013) 257–271.
- [26] S.J. Kim, S.H. Kim, J.H. Kim, S. Hwang, H.J. Yoo, Understanding metabolomics in biomedical research, *Endocrinol. Metab.* 31 (2016) 7–16.
- [27] A.R. Fernie, R.N. Trethewey, A.J. Krotzky, Metabolite profiling: from diagnostics to systems biology, *Mol. Cell Biol.* 5 (2004) 1–7.
- [28] S.A. Fancy, K. Rumpel, GC-MS-based metabolomics, *Biomark. Methods Drug Discov.*

- (2008) 317–340.
- [29] A. Garcia, C. Barbas, Gas chromatography-mass spectrometry (GC-MS)-based metabolomics, *Methods Mol. Biol.* 708 (2011) 191–204.
 - [30] X. Lu, X. Zhao, C. Bai, C. Zhao, G. Lu, G. Xu, LC-MS-based metabonomics analysis, *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* 866 (2008) 64–76.
 - [31] B. Zhou, J.F. Xiao, L. Tuli, H.W. Ransom, LC-MS-based metabolomics, *Mol. Biosyst.* 8 (2012) 470–481.
 - [32] M.R.N. Monton, T. Soga, Metabolome analysis by capillary electrophoresis-mass spectrometry, *J. Chromatogr. A.* 1168 (2007) 237–246.
 - [33] R. Ramautar, G.W. Somsen, G.J. de Jong, CE-MS for metabolomics: developments and applications in the period 2012-2014, *Electrophoresis.* 36 (2015) 212–224.
 - [34] A. Saghatelian, B.F. Cravatt, Global strategies to integrate the proteome and metabolome, *Curr. Opin. Chem. Biol.* 9 (2005) 62–68.
 - [35] E. Dudley, M. Yousef, Y. Wang, W.J. Griffiths, Targeted metabolomics and mass spectrometry, Elsevier Inc., 2010.
 - [36] M. Yuan, S.B. Breitkopf, X. Yang, J.M. Asara, A positive/negative ion-switching, targeted mass spectrometry-based metabolomics platform for bodily fluids, cells, and fresh and fixed tissue, *Nat. Protoc.* 7 (2012) 872–881.
 - [37] A. Maiolica, M.A. Jünger, I. Ezkurdia, R. Aebersold, Targeted proteome investigation via selected reaction monitoring mass spectrometry, *J. Proteomics.* 75 (2012) 3495–3513.
 - [38] R. Schiess, B. Wollscheid, R. Aebersold, Targeted proteomic strategy for clinical biomarker discovery, *Mol. Oncol.* 3 (2009) 33–44.
 - [39] X. Ye, J. Blonder, T.D. Veenstra, Targeted proteomics for validation of biomarkers in clinical samples, *Briefings Funct. Genomics Proteomics.* 8 (2009) 126–135.
 - [40] G.J. Patti, Separation strategies for untargeted metabolomics, *J. Sep. Sci.* 34 (2011) 3460–3469.
 - [41] G. Pinel, S. Weigel, J.P. Antignac, M.H. Mooney, C. Elliott, M.W.F. Nielen, B. Le

- Bizec, Targeted and untargeted profiling of biological fluids to screen for anabolic practices in cattle, *TrAC - Trends Anal. Chem.* 29 (2010) 1269–1280.
- [42] R. Aebersold, M. Mann, Mass spectrometry-based proteomics, *Nature*. 422 (2003) 198–207.
- [43] G. Theodoridis, H.G. Gika, I.D. Wilson, LC-MS-based methodology for global metabolite profiling in metabonomics/metabolomics, *TrAC - Trends Anal. Chem.* 27 (2008) 251–260.
- [44] R. Ramautar, G.W. Somsen, G.J. de Jong, CE-MS in metabolomics, *Electrophoresis*. 30 (2009) 276–291.
- [45] M. Dakna, Z. He, W.C. Yu, H. Mischak, W. Kolch, Technical, bioinformatical and statistical aspects of liquid chromatography-mass spectrometry (LC-MS) and capillary electrophoresis-mass spectrometry (CE-MS) based clinical proteomics: A critical assessment, *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* 877 (2009) 1250–1258.
- [46] T.S.K. Prasad, R. Goel, K. Kandasamy, S. Keerthikumar, S. Kumar, S. Mathivanan, D. Telikicherla, R. Raju, B. Shafreen, A. Venugopal, L. Balakrishnan, A. Marimuthu, S. Banerjee, D.S. Somanathan, A. Sebastian, S. Rani, S. Ray, C.J. Harrys Kishore, S. Kanth, M. Ahmed, M.K. Kashyap, R. Mohmood, Y.I. Ramachandra, V. Krishna, B.A. Rahiman, S. Mohan, P. Ranganathan, S. Ramabadran, R. Chaerkady, A. Pandey, Human protein reference database - 2009 update, *Nucleic Acids Res.* 37 (2009) D767–D772.
- [47] D.S. Wishart, T. Jewison, A.C. Guo, M. Wilson, C. Knox, Y. Liu, Y. Djoumbou, R. Mandal, F. Aziat, E. Dong, S. Bouatra, I. Sinelnikov, D. Arndt, J. Xia, P. Liu, F. Yallou, T. Bjorndahl, R. Perez-Pineiro, R. Eisner, F. Allen, V. Neveu, R. Greiner, A. Scalbert, HMDB 3.0-The human metabolome database in 2013, *Nucleic Acids Res.* 41 (2013) D801–D807.
- [48] M. Kanehise, S. Goto, KEGG: Kyoto encyclopedia of genes and genomes, *Nucleic Acids Res.* 28 (2000) 27–30.
- [49] C.J. Krieger, P. Zhang, L.A. Mueller, A. Wang, S. Paley, M. Arnaud, J. Pick, S.Y. Rhee, P.D. Karp, MetaCyc: a multiorganism database of metabolic pathways and enzymes,

- Nucleic Acids Res. 32 (2004) D438–D442.
- [50] R. Caspi, R. Billington, L. Ferrer, H. Foerster, C.A. Fulcher, I.M. Keseler, A. Kothari, M. Krummenacker, M. Latendresse, L.A. Mueller, Q. Ong, S. Paley, P. Subhraveti, D.S. Weaver, P.D. Karp, The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases, *Nucleic Acids Res.* 36 (2008) D623–D631.
- [51] V.G. De Gruttola, P. Clax, D.L. De Mets, G.J. Downing, S.S. Ellenberg, L. Friedman, M.H. Gail, R. Prentice, J. Wittes, S.L. Zeger, Considerations in the evaluation of surrogate endpoints in clinical trials, *Control. Clin. Trials.* 22 (2001) 485–502.
- [52] J. Atkinson, W.A. Colburn, V.G. De Gruttola, D.L. De Mets, G.J. Downing, D.F. Hoth, J.A. Oates, C.C. Peck, R.T. Schooley, B.A. Spilker, J. Woodcock, S.L. Zeger, Biomarkers and surrogate endpoints: preferred definitions and conceptual framework, *Clin. Pharmacol. Ther.* 69 (2001) 89–95.
- [53] P. Davidsson, M. Sjögren, The use of proteomics in biomarker discovery in neurodegenerative diseases, *Dis. Markers.* 21 (2005) 81–92.
- [54] X. Wei, L. Li, Mass spectrometry-based proteomics and peptidomics for systems biology and biomarker discovery in neurodegenerative diseases, *Int. J. Clin. Exp. Pathol.* 2 (2009) 132–148.
- [55] M. Shi, W.M. Caudle, J. Zhang, Biomarker discovery in neurodegenerative diseases: a proteomic approach, *Neurobiol. Dis.* 35 (2009) 157–164.
- [56] P.R. Srinivas, M. Verma, Y. Zhao, S. Srivastava, Proteomics for cancer biomarker discovery, *Clin. Chem.* 48 (2002) 1160–1169.
- [57] S.M. Hanash, S.J. Pitteri, V.M. Faca, Mining the plasma proteome for cancer biomarkers, *Nature.* 452 (2008) 571–579.
- [58] E.G. Armitage, C. Barbas, Metabolomics in cancer biomarker discovery: current trends and future perspectives, *J. Pharm. Biomed. Anal.* 87 (2014) 1–11.
- [59] C. Soto, L.D. Estrada, Protein misfolding and neurodegeneration, *Arch. Neurol.* 65 (2008) 184–189.

- [60] V. Kumar, N. Sami, T. Kashav, A. Islam, F. Ahmad, M.I. Hassan, Protein aggregation and neurodegenerative diseases: from theory to therapy, *Eur. J. Med. Chem.* 124 (2016) 1105–1120.
- [61] G. Merlini, V. Bellotti, Molecular mechanisms of amyloidosis, *N. Engl. J. Med.* 349 (2003) 583–596.
- [62] W. Zheng, Y.M. Lu, G.Y. Lu, Q. Zhao, O. Cheung, W.S. Blaner, Transthyretin, thyroxine, and retinol-binding protein in human cerebrospinal fluid: effect of lead exposure, *Toxicol. Sci.* 61 (2001) 107–114.
- [63] H.R. Bergen, S.R. Zeldenrust, S. Naylor, An on-line assay for clinical detection of amyloidogenic transthyretin variants directly from serum, *Amyloid J. Protein Fold. Disord.* 10 (2003) 190–197.
- [64] L.H. Connors, A. Lim, T. Prokaeva, V.A. Roskens, C.E. Costello, Tabulation of human transthyretin (TTR) variants, 2003, *Amyloid J. Protein Fold. Disord.* 10 (2003) 160–184.
- [65] K. Poulsen, J.M.C. Bahl, J.T. Tanassi, A.H. Simonsen, N.H.H. Heegaard, Characterization and stability of transthyretin isoforms in cerebrospinal fluid examined by immunoprecipitation and high-resolution mass spectrometry of intact protein, *Methods.* 56 (2012) 284–292.
- [66] O.B. Suhr, Y. Ando, P.I. Ohlsson, A. Olofsson, K. Andersson, E. Lundgren, M. Ando, G. Holmgren, Investigation into thiol conjugation of transthyretin in hereditary transthyretin amyloidosis, *Eur. J. Clin. Invest.* 28 (1998) 687–692.
- [67] O.B. Suhr, I.H. Svendsen, P.I. Ohlsson, J. Lendoire, P. Trigo, K. Tashima, P.J. Ranlov, Y. Ando, Impact of age and amyloidosis on thiol conjugation of transthyretin in hereditary transthyretin amyloidosis, *Amyloid Int. J. EXD. Clin. Invest.* 6 (1999) 187–191.
- [68] M. Saraiva, Hereditary transthyretin amyloidosis: molecular basis and therapeutical strategies, *Expert Rev. Mol. Med.* 4 (2002) 1–11.
- [69] P. Westermark, K. Sletten, B. Johansson, G.G. Cornwell, Fibril in senile systemic amyloidosis is derived from normal transthyretin, *Proc. Natl. Acad. Sci. U.S.A.* 87

- (1990) 2843–2845.
- [70] C. Andrade, A peculiar form of peripheral neuropathy, *Brain*. 75 (1952) 408–427.
 - [71] P. Coutinho, A. da Silva, J. Lima, A. Barbosa, Forty years of experience with type I amyloid neuropathy, *Amyloid Amyloidosis Excerpta Med.* (1980) 88–98.
 - [72] M. Munar-Ques, P. Costa, M. Saraiva, C. Viader-Farre, C. Munar-Bernat, Familial type I (Portuguese form) amyloidotic polyneuropathy in Majorca. Study using the TTR (Met30) genetic marker, *Med. Clin.* 91 (1988) 441–444.
 - [73] T. Nagasaka, Familial amyloidotic polyneuropathy and transthyretin, in: *Subcell. Biochem.* (2012) 565–607.
 - [74] Y. Ando, O. Suhr, T. Yamashita, P.I. Ohlsson, G. Holmgren, K. Obayashi, H. Terazaki, C. Mambule, M. Uchino, M. Ando, Detection of different forms of variant transthyretin (Met30) in cerebrospinal fluid, *Neurosci. Lett.* 238 (1997) 123–126.
 - [75] E.A. Pomfret, W.D. Lewis, R.L. Jenkins, P. Bergethon, S.W. Dubrey, J. Reisinger, R.H. Falk, M. Skinner, Effect of of orthotopic liver transplantation on the progression of familial amyloidotic polyneuropathy, *Transplantation*. 65 (1998) 918–925.
 - [76] S. Okamoto, J. Wixner, K. Obayashi, Y. Ando, M. Uchino, O.B. Suhr, Liver transplantation for familial amyloidotic polyneuropathy: impact on Swedish patients' survival, *Liver Transplant*. 15 (2009) 1229–1235.
 - [77] P. Hammarstrom, R.L. Wiseman, E.T. Powers, J.W. Kelly, Prevention of transthyretin amyloid disease by changing protein misfolding energetics, *Science*. 299 (2003) 713–716.
 - [78] S.D. Kim, V.S.C. Fung, An update on Huntington's disease: from the gene to the clinic, *Mov. Disord.* 27 (2014) 477–483.
 - [79] G.P. Bates, R. Dorsey, J.F. Gusella, M.R. Hayden, C. Kay, B.R. Leavitt, M. Nance, C.A. Ross, R.I. Scahill, R. Wetzel, E.J. Wild, S.J. Tabrizi, Huntington disease, *Nat. Rev. Dis. Prim.* 1 (2015) 1–21.
 - [80] H. Li, S.H. Li, H. Johnston, P.F. Shelbourne, X.J. Li, Amino-terminal fragments of mutant huntingtin show selective accumulation in striatal neurons and synaptic toxicity,

- Nat. Genet. 25 (2000) 385–389.
- [81] D.R. Langbehn, R.R. Brinkman, D. Falush, J.S. Paulsen, M.R. Hayden, A new model for prediction of the age of onset and penetrance for Huntington's disease based on CAG length, *Clin. Genet.* 65 (2004) 267–277.
- [82] S.E. Andrew, Y.P. Goldberg, B. Kremer, H. Telenius, J. Theilmann, S. Adam, E. Starr, F. Squitieri, B. Lin, M.A. Kalchman, R.K. Graham, M.R. Hayden, The relationship between tricluneotide (CAG) repeat length and clinical features of Huntington's disease, *Nat. Genet.* 4 (1993) 398–403.
- [83] C.A. Ross, E.H. Aylward, E.J. Wild, D.R. Langbehn, J.D. Long, J.H. Warner, R.I. Scahill, B.R. Leavitt, J.C. Stout, J.S. Paulsen, R. Reilmann, P.G. Unschuld, A. Wexler, R.L. Margolis, S.J. Tabrizi, Huntington disease: natural history, biomarkers and prospects for therapeutics, *Nat. Rev. Neurol.* 10 (2014) 204–216.
- [84] D.W. Weir, A. Sturrock, B.R. Leavitt, Development of biomarkers for Huntington's disease, *Lancet. Neurol.* 10 (2011) 573–590.
- [85] E.M. Rees, R.I. Scahill, N.Z. Hobbs, Longitudinal neuroimaging biomarkers in Huntington's disease, *J. Huntingtons. Dis.* 2 (2013) 21–39.
- [86] F. Niccolini, M. Politis, Neuroimaging in Huntington's disease, *World J. Radiol.* 6 (2014) 301–312.
- [87] S.M. Hersch, S. Gevorkian, K. Marder, C. Moskowitz, A. Feigin, M. Cox, P. Como, C. Zimmerman, M. Lin, L. Zhang, A.M. Ulug, M.F. Beal, W. Matson, M. Bogdanov, E. Ebbel, A. Zaleta, Y. Kaneko, B. Jenkins, N. Hevelone, H. Zhang, H. Yu, D. Schoenfeld, R. Ferrante, H.D. Rosas, Creatine in Huntington disease is safe, tolerable, bioavailable in brain and reduces serum 8OH²'dG, *Neurology.* 66 (2006) 250–252.
- [88] A. Ciammola, J. Sassone, M. Cannella, S. Calza, B. Poletti, L. Frati, F. Squitieri, V. Silani, Low brain-derived neurotrophic factor (BDNF) levels in serum of Huntington's disease patients, *Am. J. Med. Genet. Part B Neuropsychiatr. Genet.* 144B (2007) 574–577.
- [89] F. Borovecki, L. Lovrecic, J. Zhou, H. Jeong, F. Then, H.D. Rosas, S.M. Hersch, P.

- Hogarth, B. Bouzou, R.V. Jensen, D. Krainc, Genome-wide expression profiling of human blood reveals biomarkers for Huntington's disease, *Proc. Natl. Acad. Sci. U.S.A.* 102 (2005) 11023–11028.
- [90] A. Dalrymple, E.J. Wild, R. Joubert, K. Sathasivam, M. Björkqvist, Å. Peterse, G.S. Jackson, J.D. Isaacs, M. Kristiansen, G.P. Bates, B.R. Leavitt, G. Keir, M. Ward, S.J. Tabrizi, Proteomic profiling of plasma in Huntington's disease reveals neuroinflammatory activation and biomarker candidates, *J. Proteome Res.* 6 (2007) 2833–2840.
- [91] L.A. Torre, F. Bray, R.L. Siegel, J. Ferlay, J. Lortet-tieulent, A. Jemal, Global cancer statistics, 2012, *CA a Cancer J. Clin.* 65 (2015) 87–108.
- [92] Z. Zhai, X. Yu, B. Yang, Y. Zhang, L. Zhang, X. Li, H. Sun, Colorectal cancer heterogeneity and targeted therapy: clinical implications, challenges and solutions for treatment resistance, *Semin. Cell Dev. Biol.* 64 (2017) 107–115.
- [93] G. Lech, R. Słotwiński, M.I. Słodkowski, I.W. Krasnodębski, Colorectal cancer tumour markers and biomarkers: recent therapeutic advances, *World J. Gastroenterol.* 22 (2016) 1745–1755.
- [94] D.H. Lee, J.M. Lee, B.Y. Hur, I. Joo, N.J. Yi, K.S. Suh, K.W. Kang, J.K. Han, Colorectal cancer liver metastases: diagnostic performance and prognostic value of PET/MR imaging, *Radiology.* 280 (2016) 782–792.
- [95] U.H. Weidle, F. Birzele, A. Krüger, Molecular targets and pathways involved in liver metastasis of colorectal cancer, *Clin. Exp. Metastasis.* 32 (2015) 623–635.
- [96] B. Adamczyk, T. Tharmalingam, P.M. Rudd, Glycans as cancer biomarkers, *Biochim. Biophys. Acta.* 1820 (2012) 1347–1353.
- [97] Y. Zhang, J. Jiao, P. Yang, H. Lu, Mass spectrometry-based N-glycoproteomics for cancer biomarker discovery, *Clin. Proteomics.* 11 (2014) 1–14.
- [98] M.J. Kailemia, D. Park, C.B. Lebrilla, Glycans and glycoproteins as specific biomarkers for cancer, *Anal. Bioanal. Chem.* 409 (2017) 395–410.
- [99] R.A. Dwek, Glycobiology: towards understanding the function of sugars, *Chem. Rev.* 96

- (1996) 683–720.
- [100] S. Defaus, P. Gupta, D. Andreu, R. Gutiérrez-Gallego, Mammalian protein glycosylation-structure versus function, *Analyst*. 139 (2014) 2944–2967.
- [101] H. Lis, N. Sharon, Protein glycosylation. Structural and functional aspects, *Eur. J. Biochem*. 218 (1993) 1–27.
- [102] Consortium for Functional Glycomics, Functional Glycomics Gateway, <http://www.functionalglycomics.org/>.
- [103] D.J. Harvey, Proteomic analysis of glycosylation: structural determination of N-and O-linked glycans by mass spectrometry, *Expert Rev. Proteomics*. 2 (2005) 87–101.
- [104] B. Nilsson, Analysis of protein glycosylation by mass spectrometry, *Mol. Biotechnol*. 2 (1994) 243–280.
- [105] A. Dell, H.R. Morris, Glycoprotein structure determination by mass spectrometry, *Science*. 291 (2001) 2351–2356.
- [106] H.J. Jeong, Y.G. Kim, Y.H. Yang, B.G. Kim, High-throughput quantitative analysis of total N-glycans by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry, *Anal. Chem*. 84 (2012) 3453–3460.
- [107] M. Wührer, Glycomics using mass spectrometry, *Glycoconj. J*. 30 (2013) 11–22.
- [108] S. Amon, A.D. Zamfir, A. Rizzi, Glycosylation analysis of glycoproteins and proteoglycans using capillary electrophoresis-mass spectrometry strategies, *Electrophoresis*. 29 (2008) 2485–2507.
- [109] H. Geyer, R. Geyer, Strategies for analysis of glycoprotein glycosylation, *Biochim. Biophys. Acta - Proteins Proteomics*. 1764 (2006) 1853–1869.
- [110] J.J. Conboy, J.D. Henion, The determination of glycopeptides by liquid chromatography/mass spectrometry with collision-induced dissociation, *Am. Soc. Mass Spectrom*. 3 (1992) 804–814.
- [111] M. Wührer, 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.

- [112] L. Tsiatsiani, A.J.R. Heck, Proteomics beyond trypsin, *FEBS J.* 282 (2015) 2612–2626.
- [113] P. Giansanti, L. Tsiatsiani, T.Y. Low, A.J.R. Heck, Six alternative proteases for mass spectrometry-based proteomics beyond trypsin, *Nat. Protoc.* 11 (2016) 993–1006.
- [114] S. Hammarström, The carcinoembryonic antigen (CEA) family: structures, suggested functions and expression in normal and malignant tissues, *Semin. Cancer Biol.* 9 (1999) 67–81.
- [115] N.E. Lopez, C.Y. Peterson, Advances in biomarkers: going beyond the carcinoembryonic antigen, *Clin. Colon Rectal Surg.* 29 (2016) 196–204.
- [116] A.Z. Becerra, C.P. Probst, M.A. Tejani, C.T. Aquina, B.J. Hensley, K. Noyes, J.R. Monson, F.J. Fleming, Evaluating the prognostic role of elevated preoperative carcinoembryonic antigen levels on colon cancer patients: results from the national cancer database, *Ann. Surg. Oncol.* 23 (2016) 1554–1561.
- [117] C. Huang, T. Zhan, Y. Liu, Q. Li, H. Wu, D. Ji, Y. Li, Glycomic profiling of carcinoembryonic antigen isolated from human tumor tissue, *Clin. Proteomics.* 12 (2015) 1–7.
- [118] E. V. Chandrasekaran, M. Davila, D.W. Nixon, M. Goldfarb, J. Mendicino, Isolation and structures of the oligosaccharide units of carcinoembryonic antigen, *J. Biol. Chem.* 258 (1983) 7213–7222.
- [119] N.L. Anderson, The clinical plasma proteome: a survey of clinical assays for proteins in plasma and serum, *Clin. Chem.* 56 (2010) 177–185.
- [120] E.S. Boja, T.E. Fehniger, M.S. Baker, G. Marko-Varga, H. Rodriguez, Analytical validation considerations of multiplex mass-spectrometry-based proteomic platforms for measuring protein biomarkers, *J. Proteome Res.* 13 (2014) 5325–5332.
- [121] N.E. Labrou, Protein purification: an overview, *Methods Mol. Biol.* 1129 (2014) 3–10.
- [122] D. Guan, Z. Chen, Challenges and recent advances in affinity purification of tag-free proteins, *Biotechnol. Lett.* 36 (2014) 1391–1406.
- [123] P. Lescuyer, D.F. Hochstrasser, J.C. Sanchez, Comprehensive proteome analysis by chromatographic protein prefractionation, *Electrophoresis.* 25 (2004) 1125–1135.

- [124] W.C. Lee, K.H. Lee, Applications of affinity chromatography in proteomics, *Anal. Biochem.* 324 (2004) 1–10.
- [125] K. Ahrer, A. Jungbauer, Chromatographic and electrophoretic characterization of protein variants, *J. Chromatogr. B.* 841 (2006) 110–122.
- [126] G.M. Edelman, Antibody structure and molecular immunology, *Science.* 180 (1967) 830–840.
- [127] W. Wang, S. Singh, D.L. Zeng, K. King, S. Nema, Antibody structure, instability, and formulation, *J. Pharm. Sci.* 96 (2007) 1–26.
- [128] J.K.H. Liu, The history of monoclonal antibody development - progress, remaining challenges and future innovations, *Ann. Med. Surg.* 3 (2014) 113–116.
- [129] N.S. Lipman, L.R. Jackson, F. Weis-Garcia, L.J. Trudel, Monoclonal versus polyclonal antibodies: distinguishing characteristics, applications, and information resources, *ILAR J.* 46 (2005) 258–68.
- [130] K. McClatchey, *Clinical laboratory medicine*, 2002.
- [131] H.X. Chen, J.M. Busnel, A.L. Gassner, G. Peltre, X.X. Zhang, H.H. Girault, Capillary electrophoresis immunoassay using magnetic beads, *Electrophoresis.* 29 (2008) 3414–3421.
- [132] B.D. Plouffe, S.K. Murthy, L.H. Lewis, Fundamentals and application of magnetic particles in cell isolation and enrichment: a review, *Reports Prog. Physics. Phys. Soc.* 78 (2015) 1–38.
- [133] B. Kaboord, M. Perr, Isolation of proteins and protein complexes by immunoprecipitation, *Methods Mol. Biol.* 424 (2008) 349–364.
- [134] H. Mischak, J.P. Schanstra, CE-MS in biomarker discovery, validation, and clinical application, *Proteomics - Clin. Appl.* 5 (2011) 9–23.
- [135] R. Haselberg, G.J. de Jong, G.W. Somsen, CE-MS for the analysis of intact proteins 2010-2012, *Electrophoresis.* 34 (2013) 99–112.
- [136] J. Hernández-Borges, C. Neusüss, A. Cifuentes, M. Pelzing, On-line capillary electrophoresis-mass spectrometry for the analysis of biomolecules, *Electrophoresis.* 25

- (2004) 2257–2281.
- [137] D. Heiger, High Performance Capillary Electrophoresis: A Primer, Agilent Technologies Inc. (2010).
- [138] P. Schmitt-Kopplin, M. Frommberger, Capillary electrophoresis-mass spectrometry: 15 years of developments and applications, *Electrophoresis*. 24 (2003) 3837–3867.
- [139] P. Schmitt-Kopplin, A. Fekete, The CE-way of thinking: “all is relative!”, in: *Methods Mol. Biol.* (2016) 3–19.
- [140] R. Bischoff, B. Barroso, LC-MS in proteomics and glycoprotein analysis, *Recent Appl. LC-MS*. (2002) 1–8.
- [141] S. Fanali, P. Haddad, C. Poole, P. Schoenmakers, D. Lloyd, *Liquid chromatography: fundamentals and instrumentation*, Elsevier Inc. (2013).
- [142] K. Robinson, F.J. Robinson, *Liquid chromatography*, in: *Contemp. Instrum. Anal.* (2000) 628–671.
- [143] J.G. Dorsey, K.A. Dill, The molecular mechanism of retention in reversed-phase liquid chromatography, *Chem. Rev.* 89 (1989) 331–346.
- [144] Y.Y. Zhao, R.C. Lin, UPLC-MSE application in disease biomarker discovery: the discoveries in proteomics to metabolomics, *Chem. Biol. Interact.* 215 (2014) 7–16.
- [145] 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.
- [146] S. Hua, C. Lebrilla, H.J. An, Application of nano-LC-based glycomics towards biomarker discovery, *Bioanalysis*. 3 (2011) 2573–2585.
- [147] J. Abian, A.J. Oosterkamp, E. Gelpí, Comparison of conventional, narrow bore and capillary liquid chromatography/mass spectrometry for electrospray ionization mass spectrometry: practical considerations, *J. Mass Spectrom.* 34 (1999) 244–254.
- [148] K. Tomer, M.A. Moseley, L.J. Deterding, C.E. Parker, *Capillary liquid chromatography/mass spectrometry*, *Mass Spectrom. Rev.* 13 (1994) 431–457.

- [149] K. Dettmer, P.A. Aronov, B.D. Hammock, Mass spectrometry based metabolomics, *Mass Spectrom. Rev.* 26 (2007) 51–78.
- [150] F.W. McLafferty, Tandem mass spectrometry, *Science*. 214 (1981) 280–287.
- [151] B. Domon, R. Aebersold, Mass spectrometry and protein analysis, *Science*. 312 (2006) 212–217.
- [152] A. El-Aneed, A. Cohen, J. Banoub, Mass spectrometry, review of the basics: electrospray, MALDI, and commonly used mass analyzers, *Appl. Spectrosc. Rev.* 44 (2009) 210–230.
- [153] M. Dole, L.L. Mack, R.L. Hines, R.C. Mobley, L.D. Ferguson, Molecular beams of macroions, *J. Chem. Phys.* 49 (1968) 2240–2249.
- [154] P. Kebarle, U. Verkerk, Electrospray: from ions in solution to ions in the gas phase, what we know now, *Mass Spectrom. Rev.* 28 (2009) 898–917.
- [155] M. Wilm, Principles of electrospray ionization, *Mol. Cell. Proteomics*. 10 (2011) 1–8.
- [156] M. Wilm, M. Mann, Analytical properties of the nanoelectrospray ion source, *Anal. Chem.* 68 (1996) 1–8.
- [157] R. Ramautar, A.A. Heemskerk, P.J. Hensbergen, A.M. Deelder, J.M. Busnel, O.A. Mayboroda, CE–MS for proteomics: advances in interface development and application, *J. Proteomics*. 75 (2012) 3814–3828.
- [158] R.D. Smith, J.A. Olivares, N.T. Nguyen, H.R. Udseth, Capillary zone electrophoresis-mass spectrometry using an electrospray ionization interface, *Anal. Chem.* 60 (1988) 436–441.
- [159] C. Wenz, C. Barbas, Á. Lopez-Gonzalvez, A. Garcia, F. Benavente, V. Sanz-Nebot, T. Blanc, G. Freckleton, P. Britz-McKibbin, M. Shanmuganathan, F. De L’Escaille, J. Far, R. Haselberg, S. Huang, C. Huhn, M. Pattky, D. Michels, S. Mou, F. Yang, C. Neusüss, N. Tromsdorf, E.E.K. Baidoo, J.D. Keasling, S.S. Park, Interlaboratory study to evaluate the robustness of capillary electrophoresis-mass spectrometry for peptide mapping, *J. Sep. Sci.* 38 (2015) 3262–3270.
- [160] V. Sanz-Nebot, E. Balaguer, F. Benavente, J. Barbosa, Comparison of sheathless and

- sheath-flow electrospray interfaces for the capillary spectrometry analysis of peptides, *Electrophoresis*. 26 (2005) 1457–1465.
- [161] M.C. Tseng, Y.R. Chen, G.R. Her, A beveled tip sheath liquid interface for capillary electrophoresis-electrospray ionization-mass spectrometry, *Electrophoresis*. 25 (2004) 2084–2089.
- [162] F.A. Li, M.C. Wu, G.R. Her, Development of a multiplexed interface for capillary electrophoresis-electrospray ion trap mass spectrometry, *Anal. Chem.* 78 (2006) 5316–5321.
- [163] I. Kohler, Agilent Technical Note, publication number 5990-9716EN, Agilent Technologies Inc. (2012).
- [164] 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.
- [165] A.D. Zamfir, Recent advances in sheathless interfacing of capillary electrophoresis and electrospray ionization mass spectrometry, *J. Chromatogr. A*. 1159 (2007) 2–13.
- [166] 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.
- [167] R. Haselberg, G.J. de Jong, G.W. Somsen, Capillary electrophoresis-mass spectrometry of intact basic proteins using polybrene-dextran sulfate-polybrene-coated capillaries: system optimization and performance, *Anal. Chim. Acta*. 678 (2010) 128–134.
- [168] S. Medina-Casanellas, E. Domínguez-Vega, F. Benavente, V. Sanz-Nebot, G.W. Somsen, G.J. de Jong, Low-picomolar analysis of peptides by on-line coupling of fritless solid-phase extraction to sheathless capillary electrophoresis-mass spectrometry, *J. Chromatogr. A*. 1328 (2014) 1–6.
- [169] E.D. Lee, W. Muck, J.D. Henion, T.R. Covey, Liquid junction coupling for capillary zone electrophoresis ion spray mass-spectrometry, *Biol. Mass Spectrom.* 18 (1989) 844–850.
- [170] M.M. Wolff, W.E. Stephens, A pulsed mass spectrometer with time dispersion, *Rev. Sci.*

- Instrum. 24 (1953) 691.
- [171] R.S. Brown, J.J. Lennon, Mass resolution improvement by incorporation of pulsed ion extraction in a matrix-assisted laser desorption/ionization linear time-of-flight mass spectrometer, *Anal. Chem.* 67 (1995) 1998–2003.
- [172] R.J. Cotter, Time-of-flight mass spectrometry for the structural analysis of biological molecules, *Anal. Chem.* 64 (1992) 1027A–1039A.
- [173] A.G. Marshall, C.L. Hendrickson, High-resolution mass spectrometers, *Annu. Rev. Anal. Chem.* 1 (2008) 579–599.
- [174] 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.
- [175] M. Guilhaus, D. Selby, V. Mlynski, Orthogonal acceleration time-of-flight mass spectrometry, *Mass Spectrom. Rev.* 19 (2000) 65–107.
- [176] G.L. Glush, D.J. Burinsky, Hybrid mass spectrometers for tandem mass spectrometry, *J. Am. Soc. Mass Spectrom.* 19 (2008) 161–172.
- [177] H.R. Morris, T. Paxton, A. Dell, J. Langhorne, M. Berg, R.S. Bordoli, J. Hoyes, R.H. Bateman, High sensitivity collisionally-activated decomposition tandem mass spectrometry on a novel quadrupole/orthogonal-acceleration time-of-flight mass spectrometer, *Rapid Commun. Mass Spectrom.* 10 (1996) 889–896.
- [178] A. Krutchinsky, I.V. Chernushevich, V.L. Spicer, W. Ens, K.G. Standing, Collisional damping interface for an electrospray ionization time-of-flight mass spectrometer, *J. Am. Soc. Mass Spectrom.* 9 (1998) 569–579.
- [179] 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.
- [180] 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.
- [181] 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) 281–94.
- [182] C.B. Shumate, H.H. Hill, Coronaspray nebulization and ionization of liquid samples for ion mobility spectrometry, *Anal. Chem.* 61 (1989) 601–606.
- [183] J.A. McLean, B.T. Ruotolo, K.J. Gillig, D.H. Russell, Ion mobility-mass spectrometry: a new paradigm for proteomics, *Int. J. Mass Spectrom.* 240 (2005) 301–315.
- [184] S.J. Valentine, M.D. Plasencia, X. Liu, M. Krishnan, S. Naylor, H.R. Udseth, R.D. Smith, D.E. Clemmer, Toward plasma proteome profiling with ion mobility-mass spectrometry, *J. Proteome Res.* 5 (2006) 2977–2984.
- [185] P. Dwivedi, P. Wu, S.J. Klopsch, G.J. Puzon, L. Xun, H.H. Hill, Metabolic profiling by ion mobility mass spectrometry (IMMS), *Metabolomics.* 4 (2008) 63–80.
- [186] 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.
- [187] B.T. Ruotolo, S.J. Hyung, P.M. Robinson, K. Giles, R.H. Bateman, C.V. Robinson, Ion mobility-mass spectrometry reveals long-lived, unfolded intermediates in the dissociation of protein complexes, *Angew. Chem. Int. Ed. Engl.* 46 (2007) 8001–8004.
- [188] M. Borges-Alvarez, F. Benavente, M. Vilaseca, J. Barbosa, V. Sanz-Nebot, Characterization of superoxide dismutase 1 (SOD-1) by electrospray ionization-ion mobility mass spectrometry, *J. Mass Spectrom.* 48 (2013) 60–67.
- [189] A.B. Kanu, P. Dwivedi, M. Tam, L. Matz, H.H.J. Herbert, Ion-mobility-mass spectrometry, *J. Mass Spectrom.* 43 (2008) 1–22.
- [190] A.A. Shvartsburg, R.D. Smith, Fundamentals of traveling wave ion mobility spectrometry, *Anal. Chem.* 80 (2008) 9689–9699.
- [191] 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.
- [192] Z. Malá, P. Gebauer, P. Boček, Recent progress in analytical capillary isotachopheresis,

- Electrophoresis. 36 (2015) 2–14.
- [193] Z. Malá, P. Gebauer, P. Boček, Analytical capillary isotachopheresis after 50 years of development: recent progress 2014-2016, *Electrophoresis*. 38 (2017) 9–19.
- [194] S. Medina-Casanellas, F. Benavente, J. Barbosa, V. Sanz-Nebot, Transient isotachopheresis in on-line solid phase extraction capillary electrophoresis time-of-flight-mass spectrometry for peptide analysis in human plasma, *Electrophoresis*. 32 (2011) 1750–1759.
- [195] P. Puig, F. Borrull, M. Calull, C. Aguilar, Sorbent preconcentration procedures coupled to capillary electrophoresis for environmental and biological applications, *Anal. Chim. Acta*. 616 (2008) 1–18.
- [196] M.C. Breadmore, R.M. Tubaon, A.I. Shallan, S.C. Phung, A.S. Abdul Keyon, D. Gstoettenmayr, P. Prapatpong, A.A. Alhusban, L. Ranjbar, H.H. See, M. Dawod, J.P. Quirino, Recent advances in enhancing the sensitivity of electrophoresis and electrochromatography in capillaries and microchips (2008-2010), *Electrophoresis*. 32 (2011) 127–148.
- [197] N.A. Guzman, T. Blanc, T.M. Phillips, Immunoaffinity capillary electrophoresis as a powerful strategy for the quantification of low-abundance biomarkers, drugs, and metabolites in biological matrices, *Electrophoresis*. 29 (2008) 3259–3278.
- [198] N.A. Guzman, D.E. Guzman, An emerging micro-scale immuno-analytical diagnostic tool to see the unseen. Holding promise for precision medicine and P4 medicine, *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* 1021 (2016) 14–29.
- [199] F.W.A. Tempels, W.J.M. Underberg, G.W. Somsen, G.J. de Jong, Design and applications of coupled SPE-CE, *Electrophoresis*. 29 (2008) 108–128.
- [200] R. Ramautar, G.W. Somsen, G.J. de Jong, Developments in coupled solid-phase extraction-capillary electrophoresis 2013-2015, *Electrophoresis*. 37 (2016) 35–44.
- [201] F. Benavente, S. Medina-Casanellas, J. Barbosa, V. Sanz-Nebot, Investigation of commercial sorbents for the analysis of opioid peptides in human plasma by on-line SPE-CE, *J. Sep. Sci.* 33 (2010) 1294–1304.

- [202] J. Pawliszyn, Sample preparation: quo vadis?, *Anal. Chem.* 75 (2003) 2543–2558.
- [203] E. Hernández, F. Benavente, V. Sanz-Nebot, J. Barbosa, Evaluation of on-line solid phase extraction-capillary electrophoresis-electrospray-mass spectrometry for the analysis of neuropeptides in human plasma, *Electrophoresis*. 29 (2008) 3366–3376.
- [204] F. Benavente, M.C. Vescina, E. Hernández, V. Sanz-Nebot, J. Barbosa, N.A. Guzman, Lowering the concentration limits of detection by on-line solid-phase extraction-capillary electrophoresis-electrospray mass spectrometry, *J. Chromatogr. A*. 1140 (2007) 205–212.
- [205] N.A. Guzman, M.A. Trebilcock, J.P. Advis, The use of a concentration step to collect urinary components separated by capillary electrophoresis and further characterization of collected analytes by mass spectrometry, *J. Liq. Chromatogr.* 14 (1991) 997–1015.
- [206] N.A. Guzman, Immunoaffinity capillary electrophoresis applications of clinical and pharmaceutical relevance, *Anal. Bioanal. Chem.* 378 (2004) 37–39.
- [207] N.A. Guzman, T.M. Phillips, Immunoaffinity capillary electrophoresis: a new versatile tool for determining protein biomarkers in inflammatory processes, *Electrophoresis*. 32 (2011) 1565–1578.
- [208] 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*. 717 (2012) 134–142.
- [209] S. Medina-Casanellas, F. Benavente, J. Barbosa, V. Sanz-Nebot, Preparation and evaluation of an immunoaffinity sorbent with Fab' antibody fragments for the analysis of opioid peptides by on-line immunoaffinity solid-phase extraction capillary electrophoresis-mass spectrometry, *Anal. Chim. Acta*. 789 (2013) 91–99.
- [210] S. Medina-Casanellas, F. Benavente, E. Giménez, J. Barbosa, V. Sanz-Nebot, On-line immunoaffinity solid-phase extraction capillary electrophoresis mass spectrometry for the analysis of large biomolecules: A preliminary report, *Electrophoresis*. 35 (2014) 2130–2136.

- [211] R. Peró-Gascón, L. Pont, F. Benavente, J. Barbosa, V. Sanz-Nebot, Analysis of serum transthyretin by on-line immunoaffinity solid-phase extraction capillary electrophoresis mass spectrometry using magnetic beads, *Electrophoresis*. 37 (2016) 1220–1231.
- [212] N.A. Guzman, Determination of immunoreactive gonadotropin-releasing hormone in serum and urine by on-line immunoaffinity capillary electrophoresis coupled to mass spectrometry, *J. Chromatogr. B Biomed. Sci. Appl.* 749 (2000) 197–213.
- [213] E. Giménez, F. Benavente, C. de Bolós, E. Nicolás, J. Barbosa, V. Sanz-Nebot, Analysis of recombinant human erythropoietin and novel erythropoiesis stimulating protein digests by immunoaffinity capillary electrophoresis-mass spectrometry, *J. Chromatogr. A*. 1216 (2009) 2574–2582.
- [214] M. Nisnevitch, M.A. Firer, The solid phase in affinity chromatography: strategies for antibody attachment, *J. Biochem. Biophys. Methods*. 49 (2001) 467–480.
- [215] J.E. Schiel, R. Mallik, S. Soman, K.S. Joseph, D.S. Hage, Application of silica support in affinity chromatography, *J. Sep. Sci.* 29 (2006) 719–737.
- [216] J. Jaumot, R. Gargallo, A. De Juan, R. Tauler, A graphical user-friendly interface for MCR-ALS: a new tool for multivariate curve resolution in MATLAB, *Chemom. Intell. Lab. Syst.* 76 (2005) 101–110.
- [217] R. Tauler, Calculation of maximum and minimum band boundaries of feasible solutions for species profiles obtained by multivariate curve resolution, *J. Chemom.* 15 (2001) 627–646.
- [218] J. Saurina, S. Hernández-Cassou, R. Tauler, A. Izquierdo-Ridorsa, Multivariate resolution of rank-deficient spectrophotometric data from first-order kinetic decomposition reactions, *J. Chemom.* 12 (1998) 183–203.
- [219] J. Saurina, Multivariate curve resolution based on alternating least squares in capillary electrophoresis, in: *Chemom. Methods Capill. Electrophor.* (2009) 199–226.
- [220] M. Maeder, Evolving factor analysis for the resolution of overlapping chromatographic peaks, *Anal. Chem.* 59 (1987) 527–530.
- [221] W. Windig, J. Guilment, Interactive self-modeling mixture analysis, *Anal. Chem.* 63

- (1991) 1425–1432.
- [222] R. Tauler, A. Smilde, B. Kowalski, Selectivity, local rank, three-way data analysis and ambiguity in multivariate curve resolution, *J. Chemom.* 9 (1995) 31–58.
- [223] A. De Juan, R. Tauler, Chemometrics applied to unravel multicomponent processes and mixtures: revisiting latest trends in multivariate resolution, *Anal. Chim. Acta.* 500 (2003) 195–210.
- [224] S. Wold, K. Esbensen, P. Geladi, Principal component analysis, *Chemom. Intell. Lab. Syst.* 2 (1987) 37–52.
- [225] I. Jolliffe, Principal component analysis and factor analysis, in: *Springer Ser. Stat.* (2002) 150–166.
- [226] I.T. Jolliffe, J. Cadima, Principal component analysis: a review and recent developments, *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* 374 (2016) 1–16.
- [227] K.L. Sainani, Introduction to principal components analysis, *PM&R.* 6 (2014) 275–278.
- [228] S. Wold, A. Ruhe, H. Wold, W.J. Dunn, The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses, *SIAM J. Sci. Stat. Comput.* 5 (1984) 735–743.
- [229] S. Wold, M. Sjöström, L. Eriksson, PLS-regression: a basic tool of chemometrics, *Chemom. Intell. Lab. Syst.* 58 (2001) 109–130.
- [230] P. Geladi, B.R. Kowalski, Partial least-squares regression: a tutorial, *Anal. Chim. Acta.* 185 (1986) 1–17.
- [231] T. Rajalahti, O.M. Kvalheim, Multivariate data analysis in pharmaceuticals: a tutorial review, *Int. J. Pharm.* 417 (2011) 280–290.
- [232] R.G. Brereton, G.R. Lloyd, Partial least squares discriminant analysis: taking the magic away, *J. Chemom.* 28 (2014) 213–225.
- [233] 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.
- [234] M. Borges-Alvarez, F. Benavente, J. Barbosa, V. Sanz-Nebot, Separation and

- characterization of superoxide dismutase 1 (SOD-1) from human erythrocytes by capillary electrophoresis time-of-flight mass spectrometry, *Electrophoresis*. 33 (2012) 2561–2569.
- [235] O.B. Suhr, Y. Ando, P.I. Ohlsson, A. Olofsson, K. Andersson, E. Lundgren, M. Ando, G. Holmgren, Investigation into thiol conjugation of transthyretin in hereditary transthyretin amyloidosis, *Eur. J. Clin. Invest.* 28 (1998) 687–692.
- [236] A. Biroccio, P. Del Boccio, M. Panella, S. Bernardini, C. Di Ilio, D. Gambi, P. Stanzione, P. Sacchetta, G. Bernardi, A. Martorana, G. Federici, A. Stefani, A. Urbani, Differential post-translational modifications of transthyretin in Alzheimer's disease: a study of the cerebral spinal fluid, *Proteomics*. 6 (2006) 2305–2313.
- [237] K. Poulsen, J.M. Bahl, A.H. Simonsen, S.G. Hasselbalch, N.H. Heegaard, Distinct transthyretin oxidation isoform profile in spinal fluid from patients with Alzheimer's disease and mild cognitive impairment, *Clin. Proteomics*. 11 (2014) 1–9.
- [238] P.J. Hogg, Disulfide bonds as switches for protein function, *Trends Biochem. Sci.* 28 (2003) 210–214.
- [239] I.N. Serdyuk, N.R. Zaccai, J. Zaccai, *Methods in molecular biophysics*, Cambridge University Press. (2007).
- [240] C.S. Jørgensen, L.R. Ryder, A. Steinø, P. Højrup, J. Hansen, N.H. Beyer, N.H.H. Heegaard, G. Houen, Dimerization and oligomerization of the chaperone calreticulin, *FEBS J.* 270 (2003) 4140–4148.
- [241] N.H.H. Heegaard, T.J.D. Jørgensen, N. Rozlosnik, D.B. Corlin, J.S. Pedersen, A.G. Tempesta, P. Roepstorff, R. Bauer, M.H. Nissen, Unfolding, aggregation, and seeded amyloid formation of lysine-58-cleaved beta2-microglobulin, *Biochemistry*. 44 (2005) 4397–4407.
- [242] J.M.A. Gavina, P. Britz-McKibbin, Protein unfolding and conformational studies by capillary electrophoresis, *Curr. Anal. Chem.* 3 (2007) 17–31.
- [243] G.G. Mironov, C.M. Clouthier, A. Akbar, J.W. Keillor, M. V Berezovski, Simultaneous analysis of enzyme structure and activity by kinetic capillary electrophoresis-MS, *Nat.*

- Chem. Biol. 12 (2016) 918–922.
- [244] D.S. Dalpathado, H. Desaire, Glycopeptide analysis by mass spectrometry, *Analyst*. 133 (2008) 731–738.
- [245] G. Zauner, C.A.M. Koeleman, A.M. Deelder, M. Wuhler, Protein glycosylation analysis by HILIC-LCMS of Proteinase K-generated N- and O-glycopeptides, *J. Sep. Sci.* 33 (2010) 903–910.
- [246] C. Temporini, E. Perani, E. Calleri, L. Dolcini, D. Lubda, G. Caccialanza, G. Massolini, Pronase-immobilized enzyme reactor: an approach for automation in glycoprotein analysis by LC/LC-ESI/MSn, *Anal. Chem.* 79 (2007) 355–363.
- [247] A. Barroso, E. Giménez, F. Benavente, J. Barbosa, V. Sanz-Nebot, Modelling the electrophoretic migration behaviour of peptides and glycopeptides from glycoprotein digests in capillary electrophoresis-mass spectrometry, *Anal. Chim. Acta.* 854 (2015) 169–177.
- [248] G. Palermo, P. Piraino, H.D. Zucht, Performance of PLS regression coefficients in selecting variables for each response of a multivariate PLS for omics-type data, *Adv. Appl. Bioinforma. Chem.* 2 (2009) 57–70.
- [249] M. Li, L. Song, X. Qin, Glycan changes: cancer metastasis and anti-cancer vaccines, *J. Biosci.* 35 (2010) 665–673.
- [250] S.S. Pinho, C.A. Reis, Glycosylation in cancer: mechanisms and clinical implications, *Nat. Rev. Cancer.* 15 (2015) 540–555.
- [251] M.C. Plucinsky, W.M. Riley, J.J. Prorok, J.A. Alhadeff, Total and lipid-associated serum sialic acid levels in cancer patients with different primary sites and differing degrees of metastatic involvement, *Cancer.* 58 (1986) 2680–2685.
- [252] B.N. Vajaria, K.R. Patel, R. Begum, P.S. Patel, Sialylation: an avenue to target cancer cells, *Pathol. Oncol. Res.* 22 (2016) 443–447.
- [253] E. Miyoshi, K. Moriwaki, T. Nakagawa, Biological function of fucosylation in cancer biology, *J. Biochem.* 143 (2008) 725–729.
- [254] E. Miyoshi, K. Moriwaki, N. Terao, C.C. Tan, M. Terao, T. Nakagawa, H. Matsumoto,

- S. Shinzaki, Y. Kamada, Fucosylation is a promising target for cancer diagnosis and therapy, *Biomolecules*. 2 (2012) 34–45.
- [255] R. Ziganshin, G. Arapidi, I. Azarkin, E. Zaryadieva, D. Alexeev, V. Govorun, V. Ivanov, New method for peptide desorption from abundant blood proteins for plasma/serum peptidome analyses by mass spectrometry, *J. Proteomics*. 74 (2011) 595–606.
- [256] C.A. Smith, G.O. Maille, E.J. Want, C. Qin, S.A. Trauger, T.R. Brandon, D.E. Custodio, R. Abagyan, G. Siuzdak, METLIN: a metabolite mass spectral database, *Ther. Drug Monit.* 27 (2005) 747–751.
- [257] K.A. Schug, W. Lindner, Isomeric discrimination of arginine-containing dipeptides using electrospray ionization-ion trap mass spectrometry and the kinetic method, *J. Am. Soc. Mass Spectrom.* 15 (2004) 840–847.
- [258] K.L. Chang, L.S. New, M. Mal, C.W. Goh, C.C. Aw, E.R. Browne, E.C.Y. Chan, Metabolic profiling of 3-nitropropionic acid early-stage Huntington’s disease rat model using gas chromatography time-of-flight mass spectrometry, *J. Proteome Res.* 10 (2011) 2079–2087.
- [259] A.J. Kee, R.C. Smith, A.S. Gross, D.C. Madsen, B. Rowe, The effect of dipeptide structure on dipeptide and amino acid clearance in rats, *Metabolism*. 43 (1994) 1373–1378.
- [260] I. Kardys, M.P.M. de Maat, C.C.W. Klaver, D.D.G. Despriet, A.G. Uitterlinden, A. Hofman, P.T.V. de Jong, J.C.M. Witteman, Usefulness of Combining Complement Factor H and C-Reactive Protein Genetic Profiles for Predicting Myocardial Infarction (from the Rotterdam Study), *Am. J. Cardiol.* 100 (2007) 646–648.
- [261] J. Ali-Torres, L. Rodríguez-Santiago, M. Sodupe, A. Rauk, Structures and stabilities of $\text{Fe}^{2+/3+}$ complexes relevant to Alzheimer’s disease: an ab initio study, *J. Phys. Chem. A*. 115 (2011) 12523–12530.
- [262] S. Ginés, I.S. Seong, E. Fossale, E. Ivanova, F. Trettel, J.F. Gusella, V.C. Wheeler, F. Persichetti, M.E. MacDonald, Specific progressive cAMP reduction implicates energy deficit in presymptomatic Huntington’s disease knock-in mice, *Hum. Mol. Genet.* 12

- (2003) 497–508.
- [263] H. Wang, P.J. Lim, M. Karbowski, M.J. Monteiro, Effects of overexpression of Huntingtin proteins on mitochondrial integrity, *Hum. Mol. Genet.* 18 (2009) 737–752.
- [264] U. Shirendeb, A.P. Reddy, M. Manczak, M.J. Calkins, P. Mao, D.A. Tagle, P.H. Reddy, Abnormal mitochondrial dynamics, mitochondrial loss and mutant huntingtin oligomers in Huntington’s disease: implications for selective neuronal damage, *Hum. Mol. Genet.* 20 (2011) 1438–1455.
- [265] A. Petersén, S. Gabery, Hypothalamic and limbic system changes in Huntington’s disease, *J. Hunt.* 1 (2012) 5–16.

