

THESE

présentée pour l'obtention du grade de
DOCTEUR DE L'UNIVERSITE LOUIS PASTEUR DE STRASBOURG

par

Audrey BEDNARCZYK

NOUVELLES METHODOLOGIES EN PROTEOMIQUE POUR UNE CARACTERISATION FINE DES PROTEINES

Soutenue le 24 octobre 2008 devant la commission d'examen :

Dr. Alain VAN DORSSELAER

Prof. Jean-Marc JELTSCH

Prof. Christian HUBER

Dr. Patrick JOUIN

Dr. Nükhet CAVUSOGLU

Directeur de thèse

Président et rapporteur interne

Rapporteur externe

Rapporteur externe

Examineur

A Chouchou,

A mon père,

A ma mère,

« Il faut savoir s'instruire dans la gaieté. Le savoir triste est un savoir mort.
L'intelligence est joie. » (Voltaire)

« Ce qui s'apprend sans peine ne vaut rien et ne demeure pas. » (René Barjavel)

Remerciements

Dans un premier temps, je tiens à exprimer ma profonde gratitude à Alain Van Dorsselaer pour m'avoir accueillie au sein de son équipe et m'avoir offert la possibilité d'aller au bout de ma thèse dans les meilleures conditions possibles. Merci pour la confiance que vous m'avez accordée et pour m'avoir donné l'opportunité d'assister à de nombreux congrès qui m'ont permis d'élargir ma vision et de me perfectionner dans le domaine de la spectrométrie de masse.

Je tiens également à remercier sincèrement Dominique Julien, George Hussler et Nükhet Cavusoglu sans lesquels cette thèse n'aurait pas été possible. Merci pour les nombreuses discussions que nous avons pu avoir toujours dans la bonne humeur et merci à vous d'avoir été là pour moi. Vous le valez bien !!! Je ne verrai plus les cheveux de la même manière dorénavant.

J'exprime toute ma reconnaissance à Jean-Marc Jeltsch, Christian Huber et Patrick Jouin pour avoir accepté d'être les membres de mon jury.

Ma gratitude va également à Christine Schaeffer-Reiss pour m'avoir encadrée dès mon arrivée au laboratoire et m'avoir fait faire mes premiers pas en chromatographie liquide. Merci à toi d'avoir toujours laissé la porte de ton bureau ouverte et d'être disponible à tout moment.

Merci également à toutes les personnes que j'ai eu l'occasion de rencontrer dans le cadre de ce travail de thèse et plus particulièrement Elsa Wagner-Rousset et Franck Zerbib.

Je tiens également à exprimer ma reconnaissance à tous les membres de l'équipe qui m'ont apporté leur soutien et leur contribution durant ce travail de thèse : Danièle Thiersé pour m'avoir fait découvrir l'électrophorèse sur gel et m'avoir soutenue dans les moments difficiles (vive le step !!!), Jean-Marc Strub pour son aide au moindres soucis instrumentaux, Sarah Sanglier pour sa bonne humeur et ses conseils, Hélène Diemer pour son aide précieuse apportée dans plusieurs projets et sa patience (vive l'ISO !!), Fabrice Varrier pour avoir soigné mon ordinateur dans les moments difficiles et les superbes grillades qu'il fait, Fabrice Bertile pour m'avoir aidée à décoder le langage des biologistes (ciao petit prince), Véronique Trimbour pour son soutien indispensable dans ce monde qu'on appelle l'administration et Cyril Colas pour son aide de dernière minute (vive les plans) .

Mes remerciements s'adressent bien sûr à tous les étudiants du laboratoire que j'ai pu rencontrer et découvrir pendant toutes ces années : un grand merci aux anciens Christine Carapito, François Delalande, Laurent Miguet, Guillaume Chevreux et Haiko Herschbach pour m'avoir si bien intégrée dans l'équipe et m'avoir aidée dans mes premiers pas en spectrométrie de masse, mes collègues de bureau Sébastien Gallien, Cédric Atmanene (le non-covalent c'est vraiment un autre monde !!!), Daniel Ayoub, Nicolas Barthélémy (tu le vaux bien maintenant !!!), Thierry Wasselin pour leur bonne humeur et les nombreux débats scientifiques et moins scientifiques (vive le foot !!!), les « précieux » Guillaume Béchade et Véronique Delval-Dubois pour leur disponibilité et les pauses bien méritées, Marie-Laure Lortz pour son aide dans certains projets lors de son stage de DEA, Laetitia Fouillen et Agnès Hovasse pour tous les moments mémorables passés devant les nanoLC-Chip-MS/MS et les nombreux éclats de rire.

Je tiens à remercier tout particulièrement mes parents, ma sœur, ma famille et mes proches qui ont toujours cru en moi et qui n'ont cessé de me soutenir dans les moments difficiles.

Enfin, j'adresse un immense merci à toi Chouchou car sans toi ...je n'en serais pas là.

Et merci à tous ceux que j'aurais pu oublier.

ABREVIATIONS

Liste des abréviations les plus utilisées

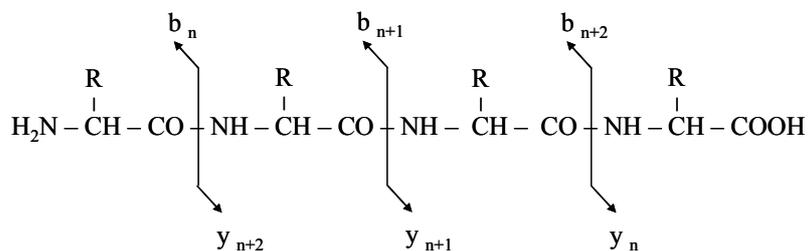
DNA	Desoxyribonucleic Acid
RNA	Ribonucleic Acid
BSA	Bovin Serum Albumin
CID	Collision Induced Dissociation
Da	Dalton
ECD	Electron Capture Dissociation
ESI	Electrospray Ionisation
ETD	Electron Transfert Dissociation
eV	Electronvolt
FSH	Follicle stimulating hormone
FT-ICR	Fourier Transform – Ion Cyclotron Resonance
Gel 1D	Monodimensional Gel Electrophoresis
Gel 2D	Bidimensional Gel Electrophoresis
HPLC	High Performance Liquid Chromatography
IF	Intermediate Filaments
IPG	Immobilized pH Gradient
IT	Ion Trap
KAP	Keratin Associated Proteins
KIF	Intermediate Filaments Keratins
LC	Liquid Chromatography
LC-MS	Liquid Chromatography coupled to Mass Spectrometry
m/z	Mass to Charge ratio
MAb	Monoclonal Antibody
MALDI	Matrix Assisted Laser Desorption/Ionisation
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
NP	Normal Phase
PFF	Peptide Fragment Fingerprint
pI	Isoelectric Point
MM	Molecular Mass
PMF	Peptide Mass Fingerprint
ppm	Part per million
PTM	Post-Translational Modification
Q	Quadrupole
RP	Reversed Phase
SDS-PAGE	Sodium Dodecyl Sulfate – Polyacrylamide Gel Electrophoresis
TOF	Time Of Flight

Liste des acides aminés

De Hoffmann E., Charette J., Stroobant V., Spectrométrie de masse, Ed. Masson, Paris, 1994, p. 220.

Acides aminés	Codes		Masse monoisotopique du résidu ^a (Da)
	Trois lettres	Une lettre	
Glycine	Gly	G	57,022
Alanine	Ala	A	71,037
Sérine	Ser	S	87,032
Proline	Pro	P	97,053
Valine	Val	V	99,068
Thréonine	Thr	T	101,048
Cystéine	Cys	C	103,009
Isoleucine	Ile	I	113,084
Leucine	Leu	L	113,084
Asparagine	Asn	N	114,043
Acide aspartique	Asp	D	115,027
Glutamine	Gln	Q	128,059
Lysine	Lys	K	128,095
Acide glutamique	Glu	E	129,043
Méthionine	Met	M	131,041
Histidine	His	H	137,059
Phénylalanine	Phe	F	147,068
Arginine	Arg	R	156,101
Tyrosine	Tyr	Y	163,063
Tryptophane	Trp	W	186,079

^a Différence de masse observée entre deux ions fragments consécutifs de même type (par exemple les ions b et y) permettant de déterminer la nature de l'enchaînement des acides aminés formant une séquence.



Avec R le groupement variable selon la nature de l'acide aminé.

$$\text{Masse}(b_{n+1}) - \text{Masse}(b_n) = \text{Masse d'un résidu.}$$

PLAN DE THESE

PLAN GENERAL

INTRODUCTION GENERALE	1
-----------------------------	---

PARTIE INTRODUCTION BIBLIOGRAPHIQUE

INTRODUCTION.....	5
CHAPITRE 1: L'ANALYSE PROTEOMIQUE.....	9
I. La spectrométrie de masse : un outil de base pour la caractérisation de protéines	10
II. Les stratégies d'identification par spectrométrie de masse des protéines	32
III. Les stratégies de séparations pour caractériser les protéines	42
CHAPITRE 2: LES NOUVEAUX BESOINS : LA QUANTIFICATION ET L'ETUDE DE MODIFICATIONS POST-TRADUCTIONNELLES.....	53
I. La quantification	53
II. L'étude de modifications post-traductionnelles (PTMs) : cas de la glycosylation	55
III. Conclusion	67
CONCLUSION.....	69
BIBLIOGRAPHIE.....	71

PARTIE RESULTATS

CHAPITRE 1 : CARACTERISATION FINE DES GLYCOFORMES D'UNE HORMONE DE FECONDITE, LA « FOLLICLE STIMULATING HORMONE » (FSH)	81
I. Introduction	82
II. Résultats : optimisations méthodologiques sur un lot de r-hFSH de référence	86
III. Application des méthodes de quantification à des lots de composition en glycoformes variable	100
IV. Conclusion.....	110
BIBLIOGRAPHIE.....	113
CHAPITRE 2 : CARACTERISATION FINE DES N-GLYCOSYLATIONS: STRATEGIES BASEES SUR L'ANALYSE NANOLC-MS/MS DE N-GLYCOPEPTIDES	115
I. Méthodologie : Analyse de glycopeptides par nanoLC-Chip-MS/MS sur colonnes en carbone graphite.....	116
II. Résultats : Exemple de trois applications.....	127
III. Conclusion.....	155
BIBLIOGRAPHIE.....	157
CHAPITRE 3 : ETUDE DU PROTEOME DU CHEVEU HUMAIN	161
I. Introduction	161
II. Méthodologie	169
III. Résultats	173
IV. Discussion	188
V. Conclusion	192
BIBLIOGRAPHIE.....	195
CONCLUSION GENERALE	201

PLAN DETAILLE

INTRODUCTION GENERALE.....	1
----------------------------	---

PARTIE INTRODUCTION BIBLIOGRAPHIQUE

INTRODUCTION.....	5
-------------------	---

<u>CHAPITRE 1: L'ANALYSE PROTEOMIQUE.....</u>	<u>9</u>
---	----------

I. La spectrométrie de masse : un outil de base pour la caractérisation de protéines	10
---	-----------

I.1. Les sources d'ionisation	11
-------------------------------------	----

I.1.1. La source électrospray (ESI)	11
---	----

a) Production de gouttelettes chargées	12
--	----

b) Fissions des gouttelettes chargées en gouttelettes filles.....	12
---	----

c) Transfert des ions dessolvatés.....	13
--	----

d) La source nanoESI.....	14
---------------------------	----

I.1.2. La source MALDI	14
------------------------------	----

I.2. Les analyseurs	16
---------------------------	----

I.2.1. L'analyseur Quadripolaire (Q)	17
--	----

I.2.2. L'analyseur à Temps de Vol (TOF).....	18
--	----

I.2.3. L'analyseur hybride : le Q-TOF	20
---	----

I.2.4. L'analyseur Trappe Ionique (IT)	20
--	----

a) Piégeage des ions	22
----------------------------	----

b) Région de stabilité des trajectoires des ions dans la trappe	23
---	----

c) Ejection des ions de la trappe	24
---	----

d) Analyse MS ⁿ	24
----------------------------------	----

e) Perspectives.....	25
----------------------	----

I.3. Les modes de fragmentation pour l'analyse des peptides	26
---	----

I.3.1. Nomenclature de fragmentation des peptides	26
---	----

I.3.2. La fragmentation en mode CID (Collision Induced Dissociation)	27
--	----

I.3.3. La fragmentation en mode ETD (Electron Transfer Dissociation).....	29
---	----

I.4. Conclusion.....	31
----------------------	----

II. Les stratégies d'identification par spectrométrie de masse des protéines	32
---	-----------

II.1. Identification par « Peptide Mass Fingerprint » (PMF)	32
---	----

II.2. Identification par nanoHPLC-MS/MS : les modes d'interprétation des spectres MS/MS	34
---	----

II.2.1. Interprétation des spectres MS/MS par comparaison avec des listes de masses théoriques issues des banques de séquences	34
--	----

a) Le « Peptide Sequence Tag »	34
--------------------------------------	----

b) Le « Peptide Fragment Fingerprint » (PFF)	35
--	----

II.2.3. Interprétation des spectres MS/MS par séquençage <i>de novo</i>	36
---	----

II.2.4. Les outils bioinformatiques pour l'interprétation des spectres de fragmentation.....	38
--	----

II.3. Les banques de séquences utilisées pour l'identification de protéines.....	38
II.3.1. Les banques de séquences nucléotidiques.....	38
II.3.2. Les banques de séquences protéiques.....	39
a) Les banques généralistes de dépôts.....	39
b) Les banques de séquences corrigées.....	39
c) UniProt (The Universal Protein Resource): une nouvelle génération de bases de données protéiques.....	40
II.4. Conclusion.....	41
III. Les stratégies de séparations pour caractériser les protéines.....	42
III.1. Les techniques de séparation des protéines : l'électrophorèse.....	43
III.1.1. L'électrophorèse bidimensionnelle sur gel (gel 2D).....	43
a) Première dimension.....	44
b) Seconde dimension.....	45
c) Détection des protéines par marquage.....	45
d) Les limitations.....	46
III.1.2. Les alternatives.....	47
a) L'électrophorèse monodimensionnelle (gel 1D).....	47
b) L'électrophorèse Off-Gel.....	47
III.2. Les techniques de séparation des peptides : la chromatographie liquide.....	48
III.2.1. Le système nanoHPLC-MS/MS.....	49
III.2.2. La chromatographie multidimensionnelle.....	50
III.3 Conclusion.....	52
CHAPITRE 2: LES NOUVEAUX BESOINS : LA QUANTIFICATION ET L'ETUDE DE MODIFICATIONS POST-TRADUCTIONNELLES.....	53
I. La quantification.....	53
II. L'étude de modifications post-traductionnelles (PTMs) : cas de la glycosylation.....	55
II.1. Introduction.....	55
II.2. Cas de la N-glycosylation.....	57
II.2.1. Introduction.....	57
II.2.2. Nature et structures des glycosylations.....	57
II.2.3. Stratégies pour la caractérisation des N-glycoprotéines.....	59
II.2.4. Apport de l'analyse MS pour déterminer le site de glycosylation et étudier l'hétérogénéité globale.....	60
a) Analyse MS des N-glycoprotéines entières.....	60
b) Analyse MS des N-glycopeptides.....	61
c) Analyse MS des protéines et des peptides déglycosylés.....	62
II.2.5. Apport de la MS ⁿ pour la caractérisation structurale.....	64
a) Nomenclature de fragmentation des polysaccharides.....	64
b) Analyse MS ⁿ des N-glycanes.....	64
c) Analyse MS ⁿ des N-glycopeptides.....	65
III. Conclusion.....	67
CONCLUSION.....	69
BIBLIOGRAPHIE.....	71

PARTIE RESULTATS

CHAPITRE 1 : CARACTERISATION FINE DES GLYCOFORMES D'UNE HORMONE DE FECONDITE, LA « FOLLICULE STIMULATING HORMONE » (FSH) 81

I. Introduction 82

I.1. Structure de la FSH 82

I.2. Rôle de la FSH et de ses glycosylations dans le cycle menstruel de la femme 83

I.3. Contexte de l'étude et choix de la stratégie 85

II. Résultats : optimisations méthodologiques sur un lot de r-hFSH de référence 86

II.1. Etude de la r-hFSH entière de référence par spectrométrie de masse 86

II.1.1. Méthodologie 86

II.1.2. Résultats : Analyse ESI-MS de la r-hFSH entière 87

a) Masses théoriques des glycoformes de la sous-unité alpha 87

b) Analyses par ESI-MS 88

c) Quantification relative des glycoformes de la sous-unité alpha de la r-hFSH entière 90

II.2. Etudes des glycopeptides de digestion de la r-hFSH de référence par LC-MS 93

II.2.1. Méthodologie 93

II.2.2. Résultats : Analyse LC-MS du digeste de r-hFSH de référence 94

a) Masses théoriques des N-glycopeptides de la sous-unité alpha 94

b) Analyse LC-MS 95

c) Quantification des glycopeptides de la sous-unité alpha de la r-hFSH 97

III. Application des méthodes de quantification à des lots de composition en glycoformes variable.... 100

III.1. Expérience 1 : quantification relative de lots enrichis en glycoformes acides et basiques 100

III.1.1. Quantification par analyses directes ESI-MS des fractions acides et basiques 101

III.1.2. Quantification par analyse LC-MS des mélanges de peptides issus de la digestion des fractions acides et basiques 103

III.2. Expérience 2 : quantification relative de lots enrichis en glycoformes selon l'antennarité 105

III.2.1. Analyses directes par ESI-MS des fractions purifiées selon l'antennarité 105

III.2.2. Analyses LC-MS des mélanges de peptides issus de la digestion des fractions purifiées selon l'antennarité 107

IV. Conclusion 110

BIBLIOGRAPHIE 113

CHAPITRE 2 : CARACTERISATION FINE DES N-GLYCOSYLATIONS: STRATEGIES BASEES SUR L'ANALYSE NANOLC-MS/MS DE GLYCOPEPTIDES 115

I. Méthodologie : Analyse de glycopeptides par nanoLC-Chip-MS/MS sur colonnes en carbone graphite 116

I.1. Le système nanoLC-Chip-MS/MS 116

I.1.1. La chip et son interface, le cube 117

I.1.2. Capacité du système 118

a) Conditions expérimentales optimales pour les deux systèmes 119

b) Performances du système nanoLC-Chip-MS/MS	120
I.1.3. Conclusion.....	121
I.2. Les chips carbone graphite pour l'analyse de N-glycopeptides.	121
I.3. Méthodologie	122
II. Résultats : Exemple de trois applications.....	127
II.1. Analyse d'anticorps thérapeutiques : les MAbs	127
II.1.1. Contexte de l'étude	127
a) Une classe de MAbs : Les immunoglobulines G (IgG).....	127
b) Rôle de la N-glycosylation des MAbs.....	129
c) Objectif : caractérisation complète des MAbs.....	129
II.1.2. Méthodologie	129
a) Préparation des échantillons.....	129
b) Analyses LC-MS des MAbs intacts et réduits.....	130
c) Analyses par ESI-MS/MS des N-glycanes libres.....	130
d) Analyses des N-glycopeptides par nanoLC-Chip-MS/MS.....	131
II.1.3. Résultats publiés.....	131
II.1.4. Perspectives.....	132
ANNEXE 1: Publication	133
The way forward, enhanced characterization of therapeutic antibody glycosylation: Comparison of three level mass spectrometry-based strategies	
Elsa Wagner-Rousset, Audrey Bednarczyk, Marie-Claire Bussat, Olivier Colas, Nathalie Corvaia, Christine Schaeffer, Alain Van Dorsselaer, Alain Beck, <i>Journal of Chromatography B</i> , 2008.	
II.2. Le parasite <i>Toxoplasma gondii</i> : identification de protéines glycosylées	135
II.2.1. Contexte de l'étude	135
a) Le parasite <i>Toxoplasma Gondii</i>	135
b) Les N-glycoprotéines.....	136
c) Objectif : caractériser des protéines N-glycosylées	137
II.2.2. Méthodologie	137
a) Séparation des protéines	137
b) Identification des protéines par nanoLC-Chip-MS/MS sur colonnes de phase inverse C18	137
c) Identification de protéines N-glycosylées par nanoLC-Chip-MS/MS sur colonnes de carbone graphite	138
II.2.3. Résultats publiés.....	138
II.2.4. Conclusion et perspectives.....	139
ANNEXE 2: Publication.....	141
Proteomic and Glycomic Analysis of N-Glycosylated Structures Involved in <i>Toxoplasma gondii</i> - Host Cell Interactions	
Syvain Fauquenoy, Willy Morelle, Agnès Hovasse, Audrey, Bednarczyk, Christian Slomianny, Christine Schaeffer, Alain Van Dorsselaer and Stanislas Tomavo, <i>Molecular and Cellular Proteomics</i> , 2008, 7, 891-910.	
II.3. Le pollen d'olivier : Etude de la N-glycoprotéine allergène Ole e1 issue de différentes variétés cultivées d'olivier.	143
II.3.1. Contexte de l'étude	143
a) La protéine Ole e1	143
b) Rôle de la N-glycosylation.....	144
c) Objectif : caractériser les mutations de la protéine Ole e1 et les différents glycoformes dans différentes variétés cultivées d'olivier	145
II.3.2. Méthodologie	145
a) Séparation des protéines sur gel 2D.....	145
b) Identification des protéines par nanoLC-Chip-MS/MS sur colonnes de phase inverse C18	146

c) Identification de protéines N-glycosylées par nanoLC-Chip-MS/MS sur colonnes de carbone graphite	146
II.3.3. Résultats en cours de publication	147
a) Séparation des protéines sur gel 2D	147
b) Mutations identifiées par analyses nanoLC-Chip-MS/MS	148
c) N-glycopeptides identifiés par analyses nanoLC-Chip-MS/MS	151
II.3.4. Conclusion et perspectives	153
III. Conclusion	155
BIBLIOGRAPHIE	157
<u>CHAPITRE 3 : ETUDE DU PROTEOME DU CHEVEU HUMAIN</u>	161
I. Introduction	161
I.1. Le cheveu	162
I.1.1. Les filaments intermédiaires (IFs)	162
I.1.2. Les protéines de la matrice	165
I.1.3. Expression des protéines dans le follicule pileux	166
I.2. L'analyse protéomique des protéines du cheveu à maturation : un défi	167
I.3. Contexte de l'étude et choix de la stratégie	168
II. Méthodologie	169
II.1. Extraction des protéines	169
II.1.1. Extraction des protéines du cheveu (cortex et cuticule)	169
II.1.2. Extraction des protéines du cortex	169
II.2. Etude des protéines du cortex par gels 2D suivis d'analyses nanoLC-Chip-MS/MS	170
II.2.1. Séparation sur gels 2D	170
II.2.2. Digestion des protéines	170
II.2.3. Identification des protéines par nanoLC-Chip-MS/MS	170
II.3. Etude des deux extraits de protéines du cheveu par chromatographie liquide bidimensionnelle (SCX/RP/MS/MS des peptides)	171
II.3.1. Digestion des deux extraits de protéines (extrait total et extrait du cortex)	171
II.3.2. Première dimension : chromatographie par échange de cations (SCX)	171
II.3.3. Seconde dimension : analyses nanoLC-Chip-MS/MS sur phase inverse C18	172
II.3.4. Interprétation des données de masse	172
III. Résultats	173
III.1. Protéines identifiées dans les extraits du cortex après séparation des protéines sur gel 2D suivi d'analyses nanoLC-MS/MS	173
III.2. Identification des protéines par LC/nanoLC-Chip-MS/MS	179
III.2.1. Extrait total (cortex et cuticule)	179
a) Les protéines identifiées	179
b) Les modifications post-traductionnelles des kératines	181
III.2.2. Extrait du cortex	183
a) Les protéines identifiées dans le cortex	183
b) Les modifications post-traductionnelles des kératines du cortex	184
III.2.3. Comparaison des résultats extrait total vs extrait du cortex	185
IV. Discussion	188
V. Conclusion	192

BIBLIOGRAPHIE	195
ANNEXE 3 : Tableau complet des protéines identifiées par approche gel 2D suivi d'analyses nanoLC-MS/MS.....	197
ANNEXE 4 : Publication.....	199
Transglutaminase-3 Enzyme: A Putative Actor in Human Hair Shaft Scaffolding ? Sébastien Thibaut, Nükhet Cavüsoglü, Emmanuelle De Becker, Franck Zerbib, Audrey Bednarczyk, Christine Schaeffer, Alain Van Dorsselaer and Bruno A. Bernard , <i>Journal of Investigative Dermatology</i> , 2008.	
CONCLUSION GENERALE.....	201

INTRODUCTION GENERALE

INTRODUCTION GENERALE

Depuis les **années 1990**, l'**analyse protéomique par spectrométrie de masse**, qui permet l'identification massive des protéines, s'est particulièrement développée grâce aux progrès réalisés concernant les **sources d'ionisation MALDI** (Matrix Assisted Laser Desorption/Ionisation) ou **ESI** (Electrospray Ionisation) et les techniques de **séparation des mélanges complexes de protéines et peptides**. Des nouveaux **outils bioinformatiques** ont également permis une exploitation rapide et plus complète des données générées pour l'**identification de centaines voire de milliers de protéines**.

Aujourd'hui les méthodologies d'identification en analyse protéomique ont largement montré leur intérêt pour l'étude des processus biologiques. L'importance de la spectrométrie de masse en biologie a d'ailleurs été confirmée en étant récompensée par l'attribution en 2002 du prix Nobel à John B. Fenn et Koichi Tanaka pour le développement des sources d'ionisation ESI et MALDI.

Toutefois, à l'heure actuelle, les **nouveaux besoins** en analyse protéomique résident dans une identification toujours plus fine des protéines pour une meilleure compréhension des processus biologiques. Le véritable objectif est donc une **caractérisation complète de toutes les séquences peptidiques** ce qui inclut la mise en évidence de modifications post-traductionnelles (PTMs), de mutations, de liaisons intra- et intermoléculaires mais également l'étude de l'hétérogénéité de ces protéines. Pour atteindre cet objectif, il est essentiel de développer de **nouvelles méthodologies en analyse protéomique** qui nécessitent de prendre en compte la **préparation** des échantillons, les stratégies de **séparation**, le choix des **spectromètres de masse** mais également **les stratégies d'identifications utilisées**.

Dans ce contexte général, **l'objectif de mon travail de thèse** a consisté à mettre **au point de nouvelles méthodologies en protéomique pour une caractérisation fine des protéines** présentes dans divers extraits biologiques ainsi qu'à la **caractérisation de PTMs** avec plus particulièrement le cas des **N-glycosylations**.

Ce manuscrit de thèse est organisé de la manière suivante :

- Une partie « **Introduction bibliographique** » (**chapitres 1 et 2**) résume l'état de l'art en analyse protéomique lorsque j'ai débuté mon travail de thèse.
- Une partie « **Résultats** » (**4 chapitres**) présente **6 études** que j'ai réalisées durant ce travail de thèse.

La partie « **Introduction bibliographique** » est dédiée aux progrès qui ont été réalisés ces vingt dernières années dans le domaine de l'analyse protéomique.

Un **premier chapitre** présente les différents outils utilisés pour l'identification des protéines par spectrométrie de masse. Ceci inclut :

- Les développements instrumentaux (sources, analyseurs, modes de fragmentation).
- Les modes d'interprétation des données de masse.
- Les techniques séparatives utilisées en amont et compatibles avec la spectrométrie de masse (gels d'électrophorèse, chromatographie liquide multidimensionnelle).

Un **second chapitre** est consacré aux nouveaux besoins en analyse protéomique qui sont la quantification et l'étude de modifications post-traductionnelles avec plus particulièrement le cas des N-glycosylations qui sera traité en détail. En effet, celle-ci est une des PTMs les plus répandues dans les protéines, et son étude est importante puisque cette PTM a un effet sur l'activité et la fonction de protéines.

La partie « **Résultats** » de ce travail de thèse rassemble **cinq études**, pour lesquelles, selon le contexte et les problématiques, les stratégies d'analyses ont différé (analyses des protéines entières, analyse de peptides, séparations sur gels d'électrophorèse, séparation par chromatographie liquide). De nombreux développements méthodologiques ont donc été nécessaires durant ce travail de thèse. L'ensemble de ces études se regroupe dans les **3 chapitres** suivants :

- **Chapitre 1 : La caractérisation fine des glycoformes d'une hormone de fécondité, la « Follicle Stimulating Hormone » (FSH) et étude de leur hétérogénéité**

La FSH est une glycoprotéine impliquée dans la croissance des follicules ovariens dont la nature et la distribution de ces glycoformes varient durant le cycle menstruel. Des traitements à base d'injection de FSH ont été élaborés et sont aujourd'hui utilisés pour combattre l'infertilité. Néanmoins, récemment l'importance dans le cycle menstruel de la micro-hétérogénéité de la N-glycosylation a été mise en évidence. De ce fait, il est apparu qu'il fallait disposer de méthodes analytiques permettant de déterminer la nature des glycoformes de la FSH et de les quantifier. Dans ce but, nous avons travaillé à la mise au point d'une méthode de quantification relative des glycoformes d'une FSH recombinante présentant des N-

glycosylations sur deux positions clefs impliquées dans l'interaction avec le récepteur (Asn⁵² et Asn⁷⁸ de la sous-unité alpha). Pour cela, nous avons choisi de partir sur deux stratégies complémentaires : (1) **l'analyse de la protéine entière par ESI-MS** afin d'étudier son hétérogénéité globale et (2) **l'analyse par LC-MS des N-glycopeptides** issus de la digestion de la protéine afin d'étudier la micro-hétérogénéité sur chaque site de glycosylation de la FSH.

- **Chapitre 2 : La caractérisation fine de N-glycoprotéines : stratégie basée sur l'analyse nanoLC-Chip-MS/MS de N-glycopeptides.**

L'analyse LC-MS/MS de N-glycopeptides nécessite souvent en amont des étapes d'enrichissement, de marquage et de purification pour favoriser la détection et la caractérisation des N-glycopeptides. Afin de s'affranchir de toutes ces étapes qui peuvent engendrer la perte d'échantillon, nous avons basé notre approche sur l'utilisation d'une nouvelle technologie de nano-chromatographie, **la LC-Chip couplée à la MS/MS**. L'utilisation de **phase stationnaire de carbone graphite** combinée à cette technologie a apporté la sélectivité chromatographique indispensable pour une analyse par spectrométrie de masse des peptides N-glycosylés. Nous avons utilisé cette approche dans les trois études suivantes :

- **Caractérisation des N-glycosylations d'anticorps monoclonaux (MAbs)** à visée thérapeutique.
- **Identification de protéines N-glycosylées** du parasite *Toxoplasma gondii* à partir de l'analyse de **bandes de gel 1D**.
- **Etude différentielle** des formes allergènes de la N-glycoprotéine Ole e1 (**distribution relative des N-glycosylations**) des pollens issus de différentes variétés d'olivier à partir de l'analyse de **spots de gels 2D**.

- **Chapitre 3 : Etude du protéome du cheveu humain**

La nature des protéines présentes dans le follicule du cheveu sont aujourd'hui très bien connues. En revanche, la manière dont celles-ci évoluent au cours de la maturation du cheveu (expression, modifications qu'elles subissent) est encore mal définie. De ce fait, l'objectif de cette étude, était de mettre au point des méthodes analytiques pour la **caractérisation de l'ensemble des protéines de la tige capillaire**, et cela afin de mieux comprendre la structure du cheveu en termes de **localisation des protéines, modifications** des protéines et **interactions** protéines-protéines.

Néanmoins, l'étude des protéines du cheveu par les approches classiques utilisées en analyse protéomique reste difficile.

En effet, l'ensemble forme un **complexe filamenteux très stable** par des **liaisons inter- et intramoléculaires** entre **protéines aux propriétés physico-chimiques extrêmes**. De ce fait, il s'agira d'**adapter les méthodes d'extraction et de solubilisation aux méthodes de séparations** (gels d'électrophorèse 1D, 2D et chromatographies liquides) tout en restant **compatible** avec l'analyse par **spectrométrie de masse**.

De plus, ces protéines présentent de **fortes homologies de séquences**. Ainsi lors de l'interprétation des données de masse, il s'agira d'une part d'utiliser **plusieurs outils bioinformatiques** et d'autre part **d'interpréter manuellement** les données de MS/MS afin d'identifier avec certitude les protéines présentes dans les extraits.

Les activités biologiques des protéines pouvant considérablement varier en fonction des modifications structurales qu'elles subissent, il est très important de développer des méthodologies pour caractériser les biomolécules dans leur intégrité.

Ainsi, l'ensemble de ce travail de thèse a été réalisé dans le cadre de la chimie et de la détermination de structure des protéines.

INTRODUCTION BIBLIOGRAPHIQUE

CHAPITRE I

L'analyse protéomique

CHAPITRE II

Les nouveaux besoins : la quantification et l'étude de PTMs, cas de la N-glycosylation

INTRODUCTION

Le terme « **protéome** », introduit pour la première fois dans les années 1996 par Wilkins [Wilkins et al., 1996], définit l'ensemble des protéines présentes dans un milieu biologique. Le protéome se définissait alors comme étant le complément protéique exprimé par un génome en considérant un état « statique ». Or selon l'état physiologique des cellules et des tissus, le type de protéines exprimées, leur abondance, leur état de modification peut varier [Aebersold and Goodlett, 2001].

En effet, une protéine est une macromolécule formée d'une ou plusieurs chaînes d'acides aminés qui résulte de la traduction de l'ARN messager lui-même issu de la transcription de l'ADN. La synthèse de protéines résulte donc directement de l'information génétique fournie par l'organisme. Selon leur structure, leur fonction va varier. Les protéines peuvent remplir des rôles divers dans les phénomènes tels que la catalyse, le transport, la communication, la signalisation, la reconnaissance ou encore dans la structure. Le protéome d'une cellule selon son état pouvant se constituer de 1000 à 10000 différents types de protéines, leur caractérisation est devenue primordiale à ce jour pour mieux comprendre les processus biologiques [Aebersold and Goodlett, 2001].

Le terme « **protéomique** » fut alors introduit pour définir l'analyse de ces protéines exprimées dans une cellule ou un tissu à un moment donné. Une première définition fut donnée en 1998 par Anderson et Anderson : « utilisation de la quantification du niveau protéique comme mesure de l'expression génique afin de caractériser des processus biologiques et de déchiffrer les mécanismes du contrôle de l'expression des gènes » [Anderson and Anderson, 1998].

Toutefois, la définition du terme « protéomique » a été longtemps discutée et aujourd'hui une définition adéquate est la suivante : « l'étude de l'ensemble des protéines exprimées dans une cellule

incluant leurs isoformes, les modifications qu'elles peuvent subir, leurs interactions avec d'autres protéines, leur description structurale et les complexes qu'elles peuvent former, donc de ce fait tout ce qui est post-génomique » [Tyers and Mann, 2003].

Jusque dans les années 1980 la méthode de choix pour l'analyse des protéines était le **séquençage par dégradation d'Edman** [Roepstorff and Kristiansen, 1974], [Edman, 1949], [Edman and Begg, 1967]. Or cette technique nécessite d'avoir des protéines purifiées et en quantité non négligeable afin d'obtenir la séquence en acides aminés complète de ces protéines. De plus, elle ne permet pas de mettre en évidence les modifications que peuvent subir ces protéines.

Cependant, dans les années 1990, de nombreux progrès furent apportés dans les instruments de **spectrométrie de masses** et les techniques associées entraînant une révolution dans la chimie des protéines ainsi que dans leur analyse (figure 1):

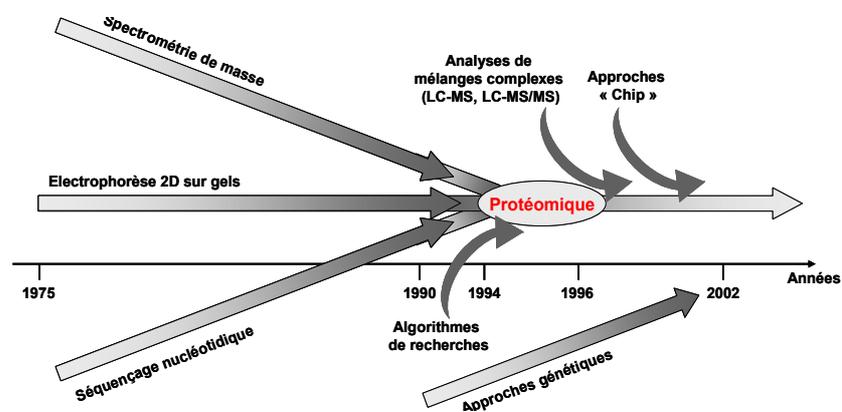


Figure 1 : Evolutions instrumentales depuis 1975 ayant contribué à l'émergence de l'analyse protéomique d'après [Patterson and Aebersold, 2003].

- ❖ Dans un premier temps, de nouvelles **sources d'ionisations** (MALDI et ESI) furent développées permettant l'ionisation de macromolécules.
- ❖ De plus, bon nombre de progrès furent apportées aux **techniques de séparation** visant à décomplexifier des mélanges de protéines et de peptides (gels 2D, chromatographie liquide) [Issaq, 2001], [Issaq, 2003].
- ❖ Enfin, la croissance exponentielle dans **l'annotation des génomes** alimentant les banques de données protéiques ainsi que le développement de nouveaux **outils bioinformatiques** ont permis une exploitation rapide et plus complète des données générées pour l'identification de centaines voire de milliers de protéines.

Toutefois, à l'heure actuelle, l'identification des protéines dans un milieu n'est plus suffisante et les **nouveaux besoins** en analyse protéomique résident dans une identification toujours plus fine des protéines pour une meilleure compréhension des processus biologiques. Le véritable objectif est donc une **caractérisation complète de toutes les séquences peptidiques** ce qui inclut entre autre la mise en évidence de modifications post-traductionnelles (PTMs), de mutations, de liaisons intra- et

intermoléculaires mais également l'étude de l'hétérogénéité de ces protéines. Pour atteindre cet objectif très ambitieux, il est nécessaire **d'améliorer les performances des analyses par spectrométrie de masse** (sensibilité, résolution, précision de masse, mode de fragmentation) mais également de développer de **nouvelles méthodologies de préparation et de séparation** des échantillons.

Cette « **Introduction bibliographique** » vise à faire un point sur les progrès effectués en analyse protéomique ces 20 dernières années. Elle se divisera en deux chapitres :

- Le **Chapitre 1** visera à faire le point sur les techniques actuellement utilisées pour **l'analyse protéomique** et leurs avancées récentes (la spectrométrie de masse, les stratégies d'identification, les stratégies de séparation).
- Le **Chapitre 2** sera dédié aux progrès apportés pour la **quantification** et pour la **caractérisation de modifications post-traductionnelles** et plus particulièrement le cas de la **N-glycosylation**.

CHAPITRE 1

L'ANALYSE PROTEOMIQUE

Ce premier chapitre est consacré à définir les différentes étapes de l'analyse protéomique basée sur l'utilisation de la spectrométrie de masse qui est un instrument indispensable pour la caractérisation des protéines. L'analyse protéomique peut se décrire en trois grandes étapes distinctes (figure 1) :

1. Séparation / digestion des protéines / extraction des peptides
2. Analyse des peptides par spectrométrie de masse (mesure de masse précise ou MS/MS)
3. Interprétation des données de masse pour l'identification des protéines (PMF, PFF, séquençage *de novo*)

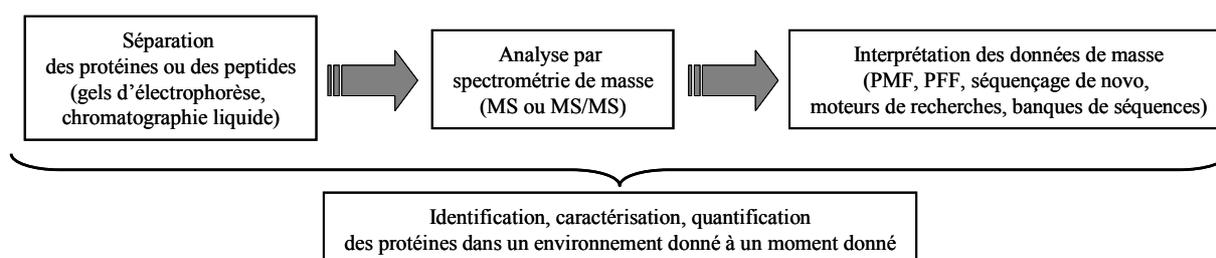


Figure 1 : Les grandes étapes de l'analyse protéomique.

Nous verrons dans un premier temps quels ont été les développements en spectrométrie de masse pour l'analyse des protéines, notamment concernant les sources d'ionisation ESI et MALDI, les différents analyseurs classiquement utilisés et quel type de données de masse peut-on générer (mesure de masse précise, les spectres de fragmentation générés par CID et ETD).

Dans un second temps, nous verrons quelles stratégies d'interprétation des données de masse sont couramment utilisées et comment exploiter ces données de masses à travers les divers logiciels bioinformatiques disponibles.

Enfin, nous aborderons les différentes stratégies pour l'analyse d'extraits complexes en discutant les différentes techniques séparatives disponibles à l'heure actuelle et comment les utiliser à bon escient selon la nature des échantillons (séparation sur gel, chromatographie liquide, analyse de protéines entières ou analyses de peptides).

I. La spectrométrie de masse : un outil de base pour la caractérisation de protéines

La spectrométrie de masse a longtemps été dédiée uniquement à l'analyse de petites molécules thermiquement stables cela étant dû au manque de techniques d'ionisation douce permettant de transférer les molécules ionisées de la phase condensée vers la phase gazeuse sans entraîner une fragmentation de celles-ci.

Depuis les années 80, la spectrométrie de masse est devenue un outil indispensable pour l'investigation de molécules biologiques tel que les protéines. Cela est notamment dû au développement de la source FAB (Fast Atom Bombardment) qui permet pour la première fois d'ioniser des molécules polaires de poids moléculaire élevé.

En effet, Barber et al. décrivent en 1981 le spectre de masse obtenu par FAB d'un peptide d'une masse de 1318 Da, la Met-Lys-Bradykinine. S'en suivit le développement simultané de deux nouvelles sources d'ionisation, l'Ionisation par Electrospray (ESI) et l'Ionisation/Désorption Laser Assistée par Matrice (MALDI). C'est en 1988 que Hillenkamp montra qu'il était possible d'ioniser par MALDI des molécules d'un poids supérieur à 60 000 Da. La même année, Fenn et al., montrèrent le premier spectre de masse multichargée de la myoglobine obtenu par ESI [Biemann, 1992].

Enfin, de nombreuses améliorations furent apportés aux analyseurs de masse notamment sur des analyseurs tels que les analyseurs de résonance cyclotronique des ions à transformée de Fourier FT-ICR, les temps de vol TOF ou encore sur les trappes ioniques IT [Wood et al., 1995] qui font partie des analyseurs les plus utilisés actuellement.

Le principe d'un spectromètre de masse est basé sur la séparation en phase gazeuse de molécules chargées (ions) en fonction de leur rapport masse/charge (m/z). Il se constitue de trois parties (figure 2):

- La source d'ionisation qui consiste à volatiliser et à ioniser les molécules
- L'analyseur qui permet de collecter et de séparer les molécules ionisées selon leur rapport masse sur charge (m/z).
- Le détecteur qui va permettre de transformer les ions en courant qui sera mesuré

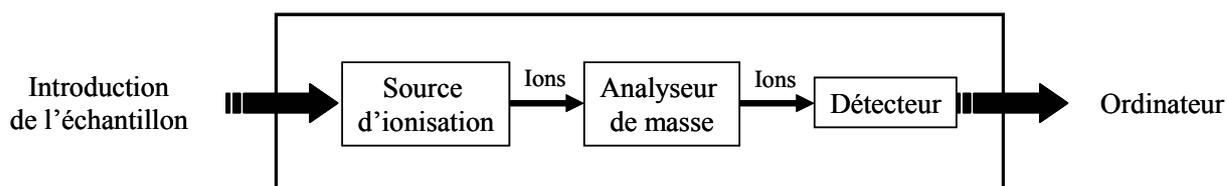


Figure 2 : Schéma des composants d'un spectromètre de masse.

Les performances d'un spectromètre de masse se définissent selon les paramètres suivants : la précision de masse, la résolution, la sensibilité et la gamme dynamique.

Au cours de ma thèse, j'ai eu l'occasion de travailler sur les spectromètres de masse suivants : un MALDI-TOF (Ultraflex de Bruker Daltonics), des trappes ioniques (HCTplus, HCTultra de Bruker Daltonics), un TOF (MicroTOF de Bruker Daltonics) et un Q-TOF (Q-TOF II de Waters). Je m'emploierai donc à décrire essentiellement les éléments de ces spectromètres de masse.

I.1. Les sources d'ionisation

I.1.1. La source électrospray (ESI)

L'analyse par spectrométrie de masse à source électrospray de molécules de poids moléculaires élevés fut pour la première fois suggérée dans les années 1960 par Dole [Dole et al., 1968].

Cependant ce n'est que quelques années plus tard, lorsque Fenn reprit ces travaux que la source ESI montra tout son intérêt pour l'étude de protéines notamment en mettant en évidence les états de charges multiples des molécules. Ainsi, ces sources peuvent être utilisées avec des analyseurs de gammes de m/z limitées et permettent l'analyse de molécules de haut poids moléculaires [Yamashita and Fenn, 1984].

De plus ce mode d'ionisation permet l'analyse directe de molécules en phase liquide et donc peut être couplée à des systèmes chromatographiques primordiaux pour l'analyse de mélanges complexes.

Le mécanisme électrospray se fait à pression atmosphérique sous l'effet d'un champ électrique très élevé qui entraîne la polarisation du liquide et la séparation des charges. Il se divise en 3 étapes :

- Production de gouttelettes chargées
- Fission des gouttelettes chargées
- Transfert des ions dessolvatés en phase gazeuse

A ce jour le mécanisme de l'électrospray est assez bien connu même s'il réside quelques ambiguïtés concernant l'émission des ions en phase gazeuse à partir de gouttelettes chargées.

a) Production de gouttelettes chargées

La source électrospray peut se voir comme une cellule électrolytique [Blades, 1991]. L'analyte est introduit via un capillaire. En appliquant entre ce capillaire et la contre-électrode une différence de potentiel de 2-5 kV, on génère un fort champ électrique (10^6 V.m^{-1}) qui va entraîner la polarisation du liquide et séparer les ions positifs des ions négatifs. A l'extrémité du capillaire, les charges vont s'accumuler et être attirées vers la contre-électrode entraînant un allongement du liquide en pointe dans sa direction. Il va se former un cône que l'on appelle « cône de Taylor ». A la pointe de ce cône, les gouttelettes ayant un excès de charge positive ou négative vont être émises [Kearle, 2000a] (figure 3).

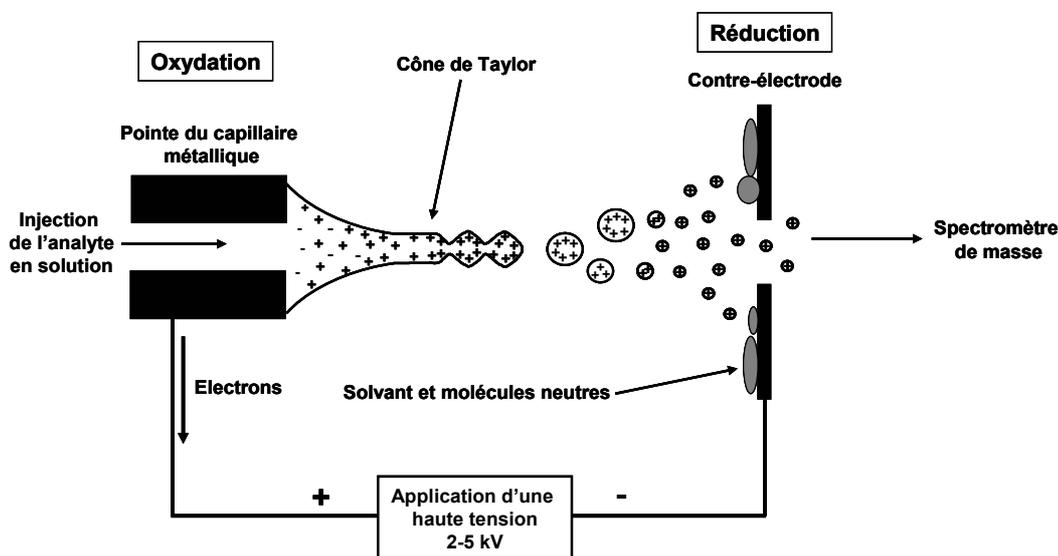


Figure 3 : Schéma du processus d'ionisation électrospray dans le cas d'une ionisation positive montrant sa nature électrolytique [Cech and Enke, 2001].

Afin d'améliorer la production de gouttelettes chargées et d'augmenter la stabilité du spray, il est nécessaire d'avoir une assistance pneumatique, le gaz de nébulisation qui est un gaz inerte tel que l'azote par exemple [Ikononou et al., 1991].

b) Fissions des gouttelettes chargées en gouttelettes filles

Sous l'effet de la température, le solvant entourant les gouttelettes chargées va s'évaporer entraînant une diminution de leur taille, donc une diminution de leur rayon et une augmentation de la densité de charge. De ce fait, pour un rayon R et une charge Q donnée, les forces dues aux répulsions électrostatiques des charges vont être égales à la tension de surface du solvant qui permet de garder les

gouttelettes fixées entre elles, on l'appelle « la limite de stabilité de Rayleigh ». Cette limite se définit selon l'équation de Rayleigh où :

$$Q^2 = f(\gamma R_R^3) \quad \text{avec} \quad Q : \text{charge de la gouttelette}$$

$$\gamma : \text{tension de surface du solvant}$$

$$R_R : \text{valeur critique de Rayleigh du rayon } R$$

Avec l'évaporation continue du solvant, le rayon des gouttelettes va diminuer tout en gardant le nombre de charges constant. L'équilibre entre les forces, n'est alors plus respecté, la limite de stabilité de Rayleigh est dépassée. Lorsque cette limite est dépassée, il va y avoir fission coulombienne des gouttelettes en plus petites gouttelettes qui auront elles-mêmes une densité de charge trop élevée entraînant des fissions en chaîne [Cole, 2000].

c) Transfert des ions dessolvatés

A l'heure actuelle deux modèles sont envisagés pour décrire l'émission des ions en phase gazeuse : le modèle de la charge résiduelle (CRM) attribué à Dole [Dole et al., 1968] et le modèle de l'évaporation des ions (IEM) développé par Iribarne et Thomson [Iribarne, 1976].

Le modèle de la charge résiduelle (CRM) : L'augmentation de la densité de charge due à l'évaporation du solvant va entraîner la division des gouttelettes mères en gouttelettes de plus en plus petites jusqu'à éventuellement ne plus avoir que des ions uniques.

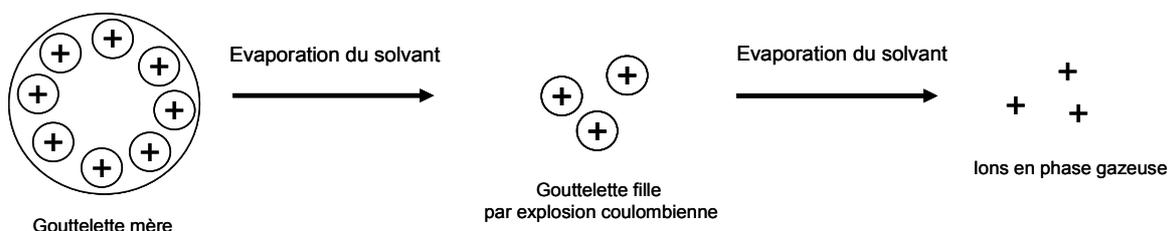


Figure 4 : Mécanisme de Dole (CRM).

Le modèle de l'évaporation des ions (IEM) : L'augmentation de densité de charge due à l'évaporation du solvant va générer des répulsions coulombiennes de forces supérieures à la tension de surface du liquide qui vont entraîner la désorption directe des ions de la gouttelette mère.

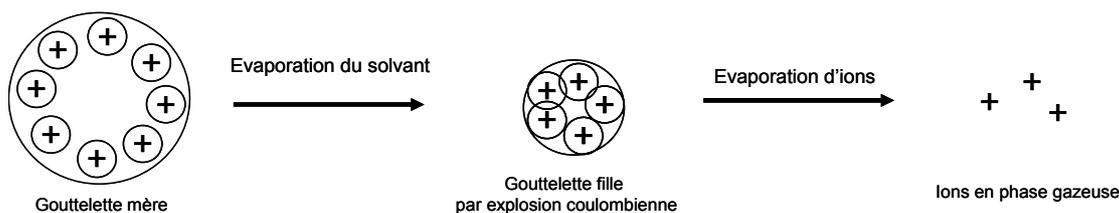


Figure 5 : Mécanisme d'Iribarne et Thomson (IEM).

Actuellement ces deux modèles sont admis. Toutefois, d'après des expériences, il semblerait que les protéines non dénaturées soient soumises au modèle de Dole et que des molécules de plus petites tailles ou ne présentant pas un encombrement stérique trop important seraient plutôt soumises au modèle d'Iribarne et Thomson [Kearle, 2000b].

d) La source nanoESI

Cette source fut développée suite aux difficultés rencontrées pour l'analyse d'échantillons biologiques. En effet, la source ESI est une source « concentration dépendante » et non « débit dépendante ». C'est à dire que la sensibilité de détection des ions sera dépendante de la concentration de l'analyte infusé.

Or les échantillons biologiques ne sont souvent disponibles qu'en faible quantité. C'est pourquoi il a fallu mettre au point des nouvelles sources ESI pouvant travailler à des débits plus faibles permettant ainsi de concentrer les échantillons biologiques pour gagner en sensibilité lors de leur détection.

La première source ESI à débit plus faible développée fut la source micro-électrospray permettant de travailler à des débits d'environ 1 $\mu\text{L}/\text{min}$ [Emmett and Caprioli, 1994]. S'en suivit la mise au point de la source nano-électrospray pouvant descendre jusqu'à des débits de 20 nL/min [Wilm and Mann, 1994], [Wilm and Mann, 1996].

Ce type de source génère des gouttelettes de taille inférieure à celles obtenues par ESI et de ce fait l'assistance pneumatique permettant de faciliter leur volatilisation n'est plus nécessaire. De plus elles ont su montrer leurs avantages puisqu'elles permettent non seulement de gagner en sensibilité mais également de gagner en précision de mesure de masse de protéines entières [Oosterkamp et al., 1998], [Shevchenko et al., 1996].

I.1.2. La source MALDI

La désorption/ionisation laser assistée par matrice (MALDI) fut développée simultanément fin des années 1980 par Karas, Hillenkamp et Tanaka [Karas and Hillenkamp, 1988], [Karas et al., 1987], [Tanaka et al., 1988].

Elle a su montrer tout son intérêt pour l'analyse de biomolécules en étant couplée à un analyseur à temps de vol (TOF). En effet, elle a permis d'analyser des molécules de poids moléculaires supérieurs à 350 kDa et cela avec une sensibilité de l'ordre de la femtomole [Caprioli et al., 1997].

Cette technique consiste à irradier par un faisceau laser à une longueur d'onde donnée un dépôt cristallin constitué d'une matrice organique et de l'analyte, le tout se déroulant sous un vide poussé (10^{-7} mbar).

Même si les phénomènes qui se produisent lors de ce processus ne sont pas encore totalement établis, le modèle le plus admis se divise en trois étapes [Zenobi, 1998]. Celui-ci sera brièvement décrit ci-dessous et illustré par la figure 6 (De Hoffman et al., 1994).

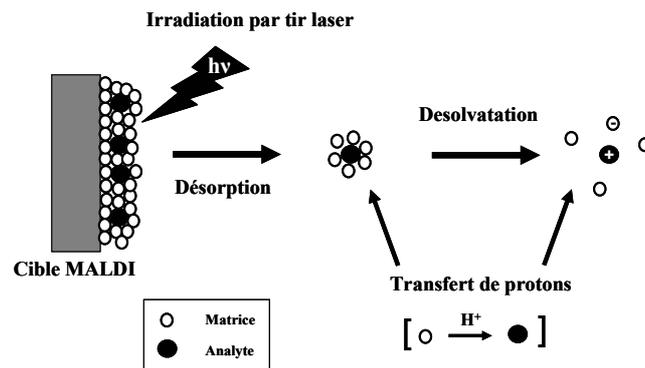


Figure 6 : Mécanisme d'ionisation MALDI [De Hoffmann et al, 1994].

- **Excitation des molécules de matrice** : Suite au tir laser sur le dépôt, les molécules de matrice vont être excitées par absorption résonnante de photons. L'excès d'énergie emmagasiné par ces molécules va être redistribué à l'analyte ou bien entraîner la dégradation des cristaux de matrice.
- **Emission des ions en phase gazeuse** : Cette redistribution va entraîner un abaissement des forces de cohésion des cristaux et provoquer l'éjection des molécules de matrices et d'analyte en phase gazeuse donnant naissance à un jet de gaz en forme de plume.
- **Ionisation des molécules en phase gazeuse** : C'est au sein de ce jet, qu'il semblerait que l'analyte soit ionisé. Toutefois cette étape du processus est encore mal définie et deux hypothèses sont envisagées [Knochenmuss and Zenobi, 2003] :
 - L'ionisation se produirait par transfert d'un proton lors des collisions entre molécules de matrices et analytes.
 - L'ionisation résulterait d'une réaction acide/base entre molécules de matrices et analytes.

De ce fait les molécules ionisées sont pour la plupart monochargées.

Ce type de technique d'ionisation présente l'avantage d'être très sensible (de l'ordre de la femtomole pour un mélange de peptides).

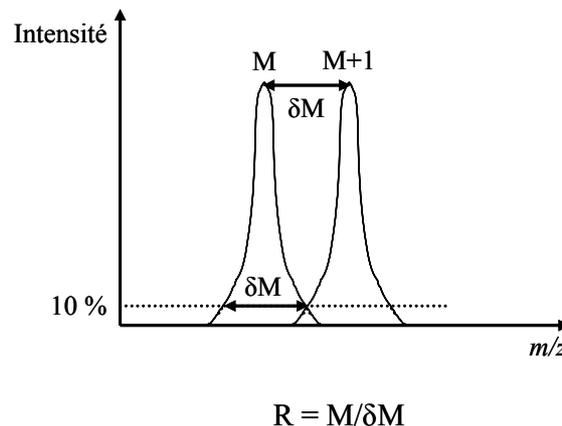
Plusieurs paramètres entrent en jeu pour une ionisation efficace :

- la longueur d'onde du laser (UV ou IR)
- le choix de la matrice (DHB, SA, 4HCCA, ...)
- le choix du mode de dépôt (couche mince, sandwich ou goutte séchée)

Parmi ces paramètres, le choix de la matrice et le mode de dépôt sont primordiaux et sont souvent étroitement liés à la nature des échantillons à analyser (protéines ou peptides de digestion, tampon de solubilisation, milieu acide ou basique).

I.2. Les analyseurs

L'analyseur constitue la partie centrale d'un spectromètre de masse et permet de définir ces performances. Parmi les principaux paramètres d'un spectromètre de masse qui définissent ses qualités, on trouve la sensibilité et la précision de mesure qui dépendent en partie de la résolution de l'instrument. La résolution est définie dans la figure 7.



La résolution R correspond à l'aptitude d'un analyseur à séparer une masse M d'une masse M+ δM (on utilise souvent une valeur de $\delta M = 1$ qui correspond à la séparation des pics isotopiques). Elle se calcule en faisant le rapport entre la masse d'un pic et sa largeur δM à 10 % de son intensité globale. On admet que deux pics sont résolus quand l'intensité de la vallée entre ces deux pics est égale à 10% de l'intensité du pic le plus faible. On appelle cela la résolution à 10 % de vallée. Actuellement, on utilise couramment la résolution FWHM (Full Width at Half Maximum), on prend alors la valeur δ à 50 % de l'intensité globale du pic.

Figure 7 : Définition de la résolution [De Hoffmann et al, 1994].

Les autres caractéristiques d'un spectromètre de masse sont la gamme de m/z analysable, la précision de masse, la sensibilité et son aptitude à générer des spectres de masse très informatifs notamment dans le cas la fragmentation des peptides.

Actuellement, quatre types d'analyseurs sont couramment utilisés en analyse protéomique :

- Le quadripôle (Q)
- Le temps de vol (TOF)
- La trappe ionique (IT)
- Le cyclotron à ions à transformée de Fourier (FT-ICR)

Chacun de ces analyseurs, aux performances différentes, présente des avantages et des inconvénients notamment selon la nature de l'étude menée. De plus ils peuvent être utilisés individuellement ou être couplés comme c'est le cas par exemple pour le Q-TOF, le TOF-TOF ou le triple quadripôle [Domon and Aebersold, 2006].

I.2.1. L'analyseur Quadripolaire (Q)

L'analyseur quadripolaire, couramment utilisé en protéomique, est généralement associé à une source électrospray et peut être couplé à d'autres analyseurs (Q-Q-Q et Q-TOF). Son principe a été décrit par Paul et Steinweden en 1953 [Paul and Steinwedel, 1953]. La séparation des ions s'y fait en appliquant un champ électrique sous l'effet duquel certains ions de valeurs m/z auront une trajectoire stable. Il sépare donc les ions en fonction de la stabilité de leur trajectoire [De Hoffmann et al., 1994].

Un quadripôle est constitué de quatre barres parallèles de section hyperbolique ou cylindrique. Chacune de ces barres va être soumise à un potentiel Φ_0 (figure 8).

Tensions appliquées aux barres :

$$+\Phi_0 = U - V\cos\omega t$$

$$-\Phi_0 = -U + V\cos\omega t$$

Avec ω : fréquence angulaire

U : tension continue

$V\cos\omega t$: tension alternative

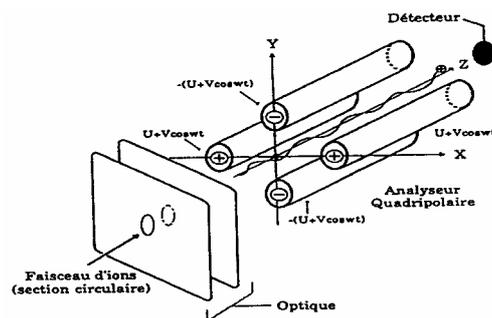


Figure 8 : Schéma d'un analyseur quadripolaire. La trajectoire des ions se fait selon l'axe z [Campana, 1980].

Ce potentiel comporte une composante continue U et une composante sinusoïdale $V\cos\omega t$, qui est la radiofréquence, où V est l'amplitude de la radiofréquence, ω la fréquence angulaire.

Les deux électrodes opposées sont reliées entre elles et soumises au même potentiel Φ_0 . Quant aux deux autres électrodes adjacentes, celles-ci sont portées aux mêmes potentiels mais de valeurs opposées ($-\Phi_0$). Cela a pour effet de créer un champ électrique.

Lorsque les ions, qui suivent une trajectoire selon l'axe z , entrent dans le quadripôle avec une certaine vitesse, ils vont être soumis selon les axes x et y aux accélérations résultantes des forces dues au champ électrique. De ce fait leur trajectoire va être déstabilisée ou stabilisée et ainsi les laisser traverser le quadripôle. Ces trajectoires sont régies par les équations de Mathieu qui ont été établies en 1866.

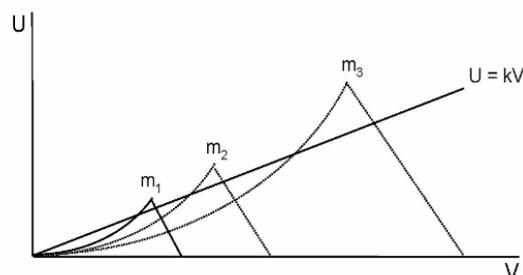


Figure 9 : Zone de stabilité des ions en fonction de U et de V pour des ions de masses différentes (m_1 , m_2 et m_3).

La résolution de ces équations permet de déterminer la zone de stabilité des ions de masses différentes en fonction de U et de V . En changeant linéairement U en fonction de V , on pourra détecter ces ions séparément (figure 9).

I.2.2. L'analyseur à Temps de Vol (TOF)

Historiquement, le concept de séparer les ions en fonction de leur m/z via un TOF fut proposé pour la première fois en 1948 par AE Cameron et DF Eggers, mais ne fut approfondi que dans les années 50 [Wiley and McLaren, 1955].

Toutefois ce n'est qu'avec l'apparition de la source MALDI dans les années 80, que cet analyseur a pu faire ses preuves en analyse protéomique [Karas and Hillenkamp, 1988]. Nous décrivons ici son principe dans le cas d'un couplage avec une source MALDI.

Le TOF se constitue de deux parties : (1) une zone d'accélération et (2) une zone de libre champ, le temps de vol (figure 10).

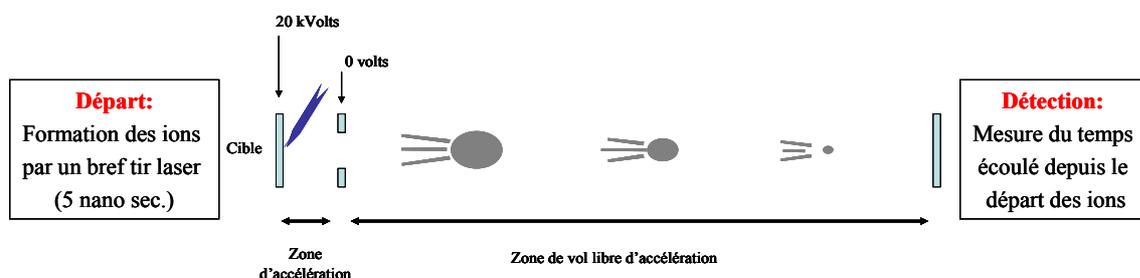


Figure 10 : Principe de l'analyseur à temps de vol couplé à une source MALDI.

Les ions accélérés par un champ électrique E vont pénétrer dans le tube de vol avec une certaine énergie cinétique E_c donc une certaine vitesse qui dépend du nombre de charges z de l'ion ainsi que de sa masse m . Par exemple, pour un même état de charge, un ion de masse plus élevée passera plus de temps dans le tube de vol.

Le rapport m/z peut alors être déterminé en mesurant le temps qu'il faut à des ions pour voyager d'un point à un autre dans une zone libre de champ. Ce calcul se fait selon plusieurs équations :

Un ion de masse m et de charge z accéléré dans un champ E aura une certaine énergie cinétique E_c

$$E_c = zeE = \frac{1}{2} mv^2 \quad (e : \text{charge de l'électron } 1,602 \cdot 10^{-19} \text{ C})$$

De ce fait on peut en déduire la vitesse de cet ion

$$v = (2zeE/m)^{1/2}$$

Le temps que mettra un ion pour parcourir une distance L dans le tube de vol sera :

$$t = L (2zeE/m)^{1/2}$$

Cependant, les ions issus de la source MALDI ne sont pas forcément générés au même moment et au même endroit, et de ce fait des ions de même m/z n'auront pas toujours la même énergie cinétique. Cela a pour effet de diminuer les capacités de l'analyseur à séparer des ions de m/z proches et donc de diminuer la résolution.

Toutefois avec le développement de l'extraction retardée et du réflecteur électrostatique, la résolution et la sensibilité du TOF pu être considérablement améliorée. Nous décrivons brièvement leurs principes qui sont également illustré dans la figure 11 :

- ❖ **L'extraction retardée (Delayed Extraction DE)** consiste à ajouter une lentille supplémentaire avant l'analyseur, la lentille DE. Cette lentille sur laquelle on applique un potentiel légèrement plus élevé que celui présent dans la source va freiner les ions avant qu'ils n'entrent dans le TOF. Ceci permet de refocaliser les ions et de leur faire prendre un départ simultané dans le tube de vol [Vestal et al., 1995].
- ❖ **Le réflecteur électrostatique** consiste à ajouter une série de lentilles à l'extrémité du tube de vol qui vont agir comme un miroir électrostatique. En appliquant sur ces lentilles des potentiels croissants, les ions vont être progressivement freinés jusqu'à faire demi-tour pour voyager vers le détecteur. La distance à parcourir étant doublée, les ions seront mieux séparés. De même, on va pouvoir refocaliser les ions de mêmes m/z qui lors de leur entrée dans le tube de vol n'ont pas forcément la même énergie cinétique. En augmentant le temps de vol, on diminue cet écart entre deux mêmes ions. Cela a pour effet d'augmenter considérablement la résolution [Mamyryn et al., 1973].

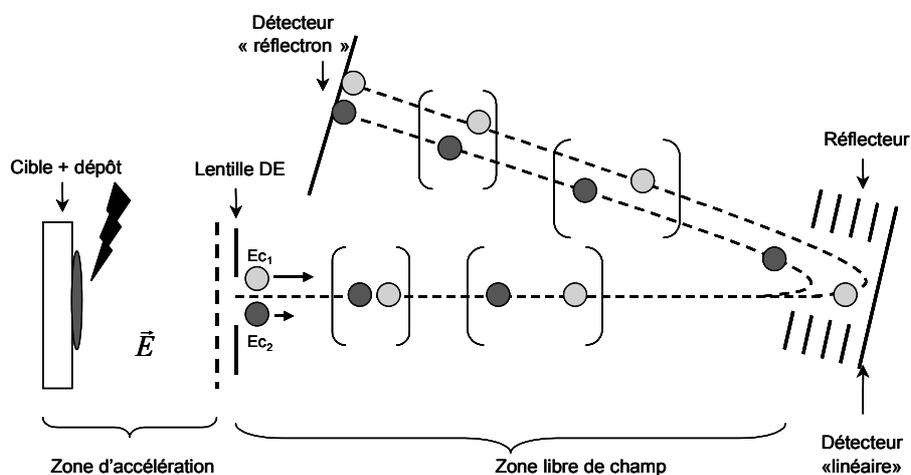


Figure 11 : Principe d'un temps de vol avec extraction retardée et réflectron couplé à une source MALDI [Wagner, 2004].

Aujourd'hui, les spectromètres de masse MALDI-TOF peuvent atteindre des résolutions allant jusqu'à 40000 et en fait un instrument de choix pour des mesures de masses précises à 10 ppm près. De même, ce type d'analyseur offre une large gamme de m/z analysable allant jusqu'à 100 000 m/z [Tanaka et al., 1988]. Enfin, il présente l'avantage d'être très sensible (50 fmol d'un mélange de peptide).

I.2.3. L'analyseur hybride : le Q-TOF

Cet instrument se constitue de deux analyseurs, un quadripôle suivi d'un temps de vol. Les deux analyseurs sont séparés par une cellule de collision [Chernushevich et al., 2001]. La figure 12 représente un exemple de schéma d'un Q-TOF.

L'analyseur Q peut être utilisé de deux manières :

- ❖ **Guide d'ions** : En abaissant la valeur de U à 0 sur les barres du quadripôle, tous les ions seront transmis vers le TOF qui jouera le rôle de l'analyseur. Dans ce cas, le quadripôle joue uniquement le rôle de guide d'ion. Cependant, la transmission des ions étant dépendante de l'amplitude de la radiofréquence, celle-ci sera successivement modifiée en cours d'analyse pour laisser passer différentes gammes de m/z .
- ❖ **Rôle d'analyseur** : il est utilisé pour sélectionner un ion spécifique et l'envoyer dans la cellule de collision afin qu'il soit fragmenté. Le TOF permettra ensuite d'analyser ces ions fragments. On appelle cela la spectrométrie de masse en tandem.

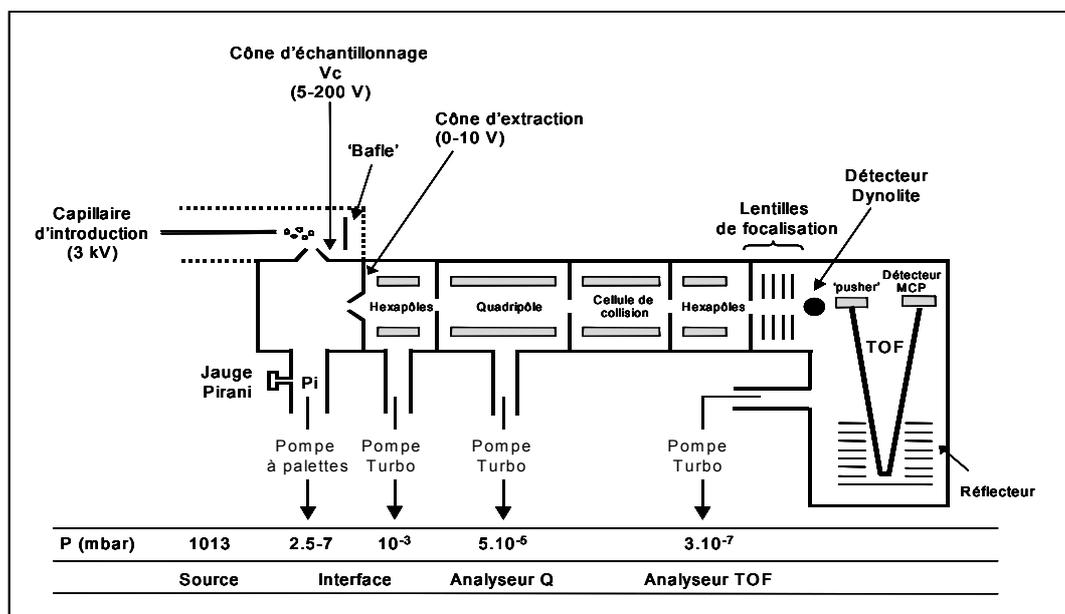


Figure 12 : Exemple d'un spectromètre de masse en tandem, le ESI-Q-TOF II (Micromass).

I.2.4. L'analyseur Trappe Ionique (IT)

Dans leur livre *Quadrupole Storage Mass Spectrometry* édité en 1989, RE March et RJ Hugues suggère que la « technologie trappe ionique » peut être résumée historiquement en étant divisée en trois périodes majeures [Stafford, 2002]:

- ❖ **1953 à 1962, la détection sélective des ions**: L'analyseur à trappe ionique découla directement des travaux réalisés par Paul et Steinwedel en 1960 [Paul, 1990] (US Patent 2 939 952) qui furent récompensés par le prix Nobel en physique en 1989 attribué à

Wolfgang Paul. Ils furent les premiers à monter une méthode pour l'analyse de masse en piégeant des ions dans une trappe à ions quadripolaire. Ils s'employèrent ensuite à détecter et à mesurer ces ions.

- ❖ **1968 à 1976, le stockage sélectif des ions :** Cette ère fut particulièrement marquée par l'ensemble des travaux réalisés par PH Dawson et NR Whetten en 1968, JFJ Todd en 1972 et RE March en 1976, dédiés à la mise au point d'une méthode permettant de stocker et d'éjecter un ion de masse donnée à la fois dans la trappe en jouant sur les champs appliqués sur les électrodes.
- ❖ **1984, l'éjection sélective des ions :** Les travaux débutés en 1979 et réalisés par G Stafford et d'autres scientifiques tel que JEP Syka, WE Reynolds et PE Kelley aboutirent à la mise sur le marché par Thermo Finnigan en 1983 de la première trappe comme analyseur de spectromètre de masse, le Ion Trap Detector (ITD) [Stafford et al., 1984] couplé à une chromatographie gazeuse.

Suite à cette commercialisation et au développement dans les années 1990 de la source ESI, de nombreuses études furent menées pour développer de nouvelles trappes à ions et améliorer les performances de cet analyseur. C'est le cas de Finnigan et Bruker-Franzen qui mirent sur le marché les premières trappes à sources externes (LCQ, GCQ, ESQUIRE) permettant d'injecter des ions générés à l'extérieur de la trappe et pouvant de ce fait être couplé à des systèmes de chromatographie liquide [Quarmby and Yost, 1999].

Toutes ces améliorations apportées à cet analyseur et son principe sont largement décrits dans bon nombres d'articles publiés par March [March, 2000], [March, 1998], [March, 1997]. Nous reverrons ici brièvement son mode de fonctionnement qui consiste à piéger des ions dans un champ électrique tridimensionnel.

La trappe ionique se constitue de 3 électrodes, deux électrodes chapeaux de géométrie hyperbolique et une électrode annulaire en forme de diabolo (figure 13). Les ions y entrent et en sortent par des orifices situés sur les électrodes chapeaux. L'électrode chapeau par laquelle entrent les ions ne contient qu'un orifice alors que celle de sortie contient plusieurs petits orifices pour faciliter leur sortie.

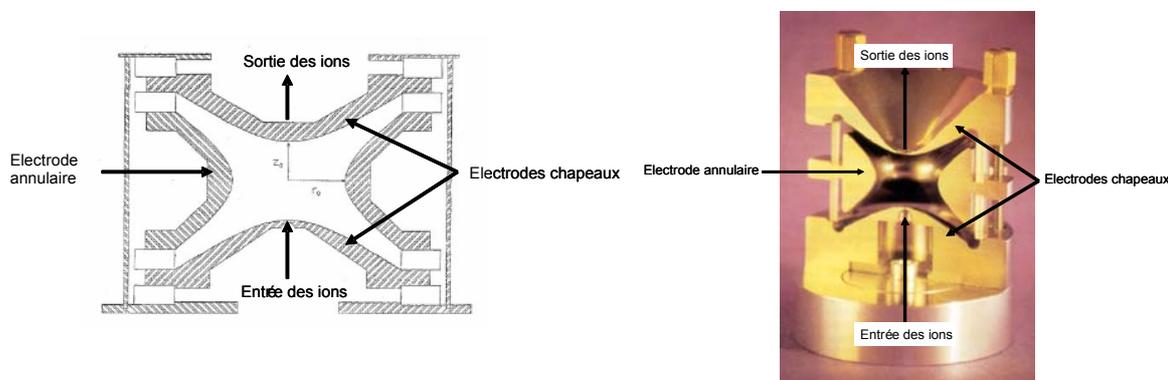


Figure 13 : Schéma d'un analyseur trappe ionique [March, 2000].

a) Piégeage des ions

Comme dans le cas du quadripôle, lorsque les ions pénètrent dans la trappe, ils vont être soumis à un champ quadripolaire généré en appliquant une radiofréquence $V \cos \omega t$ sur l'électrode annulaire qui va permettre de les piéger.

Ce champ quadripolaire peut être assimilé à un puits de potentiel de largeur $2z_0$ au fond duquel les ions sont piégés. La profondeur du puits de potentiel est définie par l'équation suivante :

$$D_z = \frac{V^2}{m/z (r_0^2 + 2z_0^2) \omega_{RF}^2}$$

D_z : profondeur du puits de potentiel

V : amplitude de radiofréquence

ω_{RF} : fréquence angulaire

r_0, z_0 : distances minimales qui séparent les électrodes du centre de la trappe

Son effet sur des ions de m/z différents peut être comparé à un verre à vin contenant des liquides de densités différentes (figure 14).

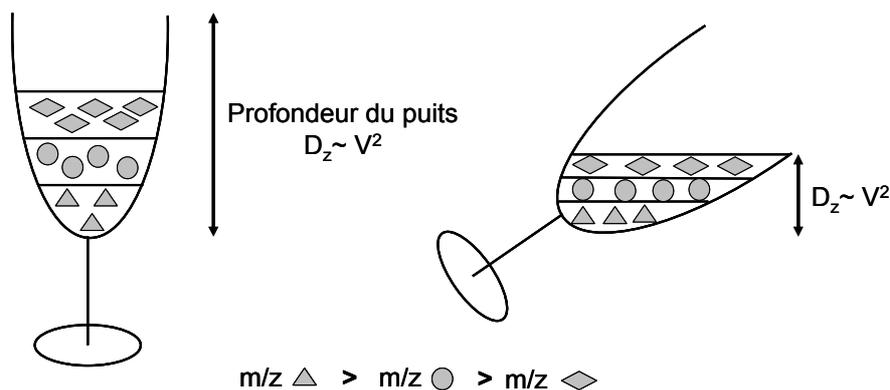


Figure 14 : Représentation schématique du puits de potentiel: il peut être comparé à un verre de vin contenant trois liquides de densités différentes, ces liquides représentent des ions de m/z différents [March, 1997].

La profondeur du puits de potentiel est directement liée à l'amplitude de la radiofréquence V . Plus celle-ci va augmenter, plus le puits de potentiel sera profond et meilleur sera le piégeage des ions. De même, la profondeur du puits étant inversement proportionnelle au rapport m/z d'un ion, les ions de m/z élevés seront mieux piégés que les ions de m/z plus faibles.

L'amplitude de la radiofréquence V a donc une influence directe sur la gamme de m/z analysée. En effet, pour une valeur de V fixe, certains ions de faibles m/z auront une fréquence trop élevée pour qu'ils soient piégés, on appelle cette limite le cut-off.

Leur trajectoire tridimensionnelle résulte également des équations de Mathieu et suit une courbe de Lissajou (figure 15). Cette trajectoire peut se comparer à un huit tridimensionnel.

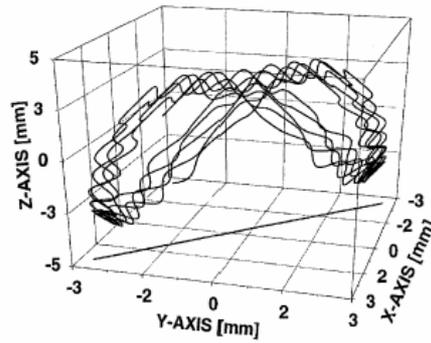
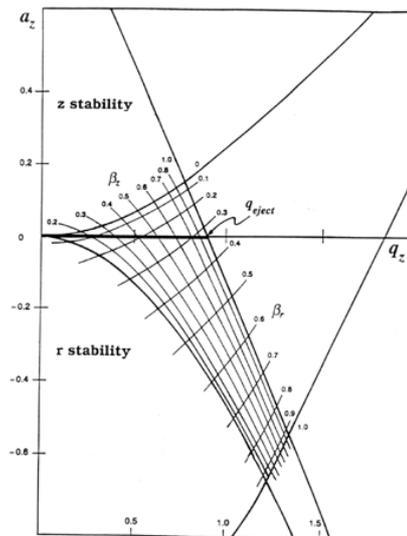


Figure 15 : Trajectoire des ions dans une trappe ionique déduite des équations de Mathieu [March, 1997].

Afin d'améliorer l'efficacité du piégeage, un gaz, l'hélium, est introduit dans la trappe à une pression d'environ 10^{-3} mbar. Ce gaz va entrer en collision avec les molécules ionisées, ce qui aura pour effet de diminuer leur énergie cinétique et de ce fait de les refocaliser au centre de la trappe.

b) Région de stabilité des trajectoires des ions dans la trappe

Comme dans le cas du quadripôle linéaire, ce sont les équations de Mathieu qui déterminent les zones de stabilité des trajectoires des ions dans l'espace. Ces équations permettent de déterminer les diagrammes de stabilité pour chacun des deux axes (z et r). Pour qu'un ion soit piégé dans la trappe, il faut que sa trajectoire soit stable selon les deux axes. En superposant les deux diagrammes de stabilité des ions selon l'axe z et l'axe r (les deux dépendant de a et de q) on observe une zone commune à ces deux diagramme et qui de ce fait correspond au diagramme de stabilité selon les deux axes z et r (figure 16).



$$a_z = \frac{-16U}{m/z(r_0^2 + 2z_0^2)\omega_{RF}^2}$$

$$a_r = \frac{-a}{2}$$

$$q_z = \frac{8V}{m/z(r_0^2 + 2z_0^2)\omega_{RF}^2}$$

$$q_r = \frac{-q}{2}$$

Figure 16 : Diagramme de stabilité des ions dans l'espace (a_z , q_z) selon les deux axes r et z.

La trappe fonctionnant en mode RF only pour piéger les ions, c'est-à-dire qu'on ne fait intervenir qu'une composante $V \cos \omega t$, le facteur a_z devient nul et de ce fait le diagramme de stabilité ne dépend plus que de q_z . On obtient alors une droite avec un point critique ($q_z = q_{\text{eject}}$) qui correspond à la valeur critique pour laquelle les ions seront éjectés ou non, $q_{\text{eject}} = 0,908$.

Pour $q_z < 0,908$: les ions ont une trajectoire stable et sont piégés.

Pour $q_z > 0,908$: les ions ont une trajectoire instable.

De même cette valeur correspond également à la valeur limite pour laquelle des ions de faible m/z peuvent être piégés, c'est le cut-off de faible masse. En effet pour une valeur de radiofréquence donnée, des ions de faible m/z auront une valeur de $q_z > 0,908$ et de ce fait seront éjectés.

c) *Ejection des ions de la trappe*

L'éjection des ions peut se faire de deux manières :

- **Ejection simple** : celle-ci consiste à effectuer un balayage croissant de la radiofréquence afin d'éjecter les ions par ordre croissant de m/z . Or étant limité par la valeur de la tension applicable (20kV), il y aura des ions de m/z élevés pour lesquels l'amplitude maximale de la tension alternative ne sera pas suffisante pour les éjecter. Cela a pour effet de limiter la gamme de m/z analysable.
- **Ejection résonante** [March, 1997]: afin d'augmenter la gamme de m/z analysable, on va mettre les ions en résonance pour forcer leur éjection. Pour cela, on transfère de l'énergie cinétique aux ions suivant l'axe z en appliquant une tension alternative sur les électrodes chapeaux. Les ions qui étaient alors confinés au centre de la trappe verront leur mouvement oscillatoire périodique augmenter et seront de ce fait éjectés. De même chaque ion ayant une fréquence d'oscillation propre, il sera possible de mettre en résonance chacun d'eux individuellement. Enfin, l'éjection résonante permet également d'augmenter la résolution et la précision de masse. En effet, lorsque la quantité d'ions accumulés est trop importante, le champ quadripolaire est perturbé entraînant une éjection retardée des ions, on appelle cela les phénomènes d'espace-charge. Ces phénomènes se traduisent par une baisse de la résolution et de la précision de mesure de l'instrument.

d) *Analyse MSⁿ*

La trappe permettant de contrôler les trajectoires des ions et de ce fait de les éjecter sélectivement, cet analyseur peut être utilisé pour fragmenter des ions soit en deux étapes MS/MS ou soit encore en de multiples étapes MS³, MS⁴, ..., MSⁿ. En effet, il est possible d'isoler sélectivement un ion dans la trappe puis de le fragmenter. Pour cela il faudra éjecter les ions de m/z inférieurs et de m/z supérieurs à l'ion de m/z sélectionné. On procède de la manière suivante :

- **Ejection des ions de $m/z < m/z$ de l'ion précurseur** : on effectue un balayage croissant de l'amplitude de radiofréquence appliquée sur l'électrode annulaire jusqu'à atteindre la limite de l'amplitude nécessaire pour éjecter l'ion précurseur.

- **Ejection des ions de $m/z > m/z$ de l'ion précurseur** : on effectue un balayage en fréquence en partant d'une fréquence proche de la fréquence de résonance de l'ion précurseur.

Afin d'affiner l'isolation de l'ion précurseur, on effectue des balayages en fréquences proches de la fréquence de résonance de celui-ci.

Une fois l'ion isolé, il s'agit de l'exciter et de le fragmenter. Pour cela, l'ion précurseur est placé en résonance en ajustant la fréquence de la tension alternative sur l'électrode chapeau de sortie à la valeur de la fréquence de résonance propre de l'ion précurseur. Les ions ainsi dopés en énergie cinétique vont entrer en collision avec les molécules de gaz. Cela a pour effet de convertir l'énergie cinétique en énergie interne vibrationnelle et d'entraîner la fragmentation des molécules. Afin de ne pas éjecter les ions fragments, l'amplitude de résonance est maintenue constante et inférieure à la valeur de l'amplitude d'éjection des ions. Enfin, l'énergie transmise aux ions précurseurs n'étant délivrée qu'à leur seule fréquence de résonance, les ions fragments générés ne seront pas fragmentés une seconde fois.

Les ions fragments générés peuvent alors être soit éjectés dans leur totalité ou bien en répétant les opérations décrites ci-dessus, il sera possible de sélectionner un ion issu de la première fragmentation, de l'isoler et de le fragmenter pour obtenir le maximum d'informations quant à la structure de la molécule. Toutefois inclure plusieurs étapes de fragmentation nécessite une certaine quantité de matériel.

e) Perspectives

Suites à toutes les améliorations apportées à cet analyseur, celui-ci est devenu un outil de choix pour l'analyse protéomique. En effet, robuste, sensible et peu coûteux, de nombreux constructeurs se sont appliqués à augmenter les performances de cet analyseur notamment en augmentant la capacité de stockage des ions dans la trappe. De nouvelles générations d'instruments ont été développées permettant de gagner en sensibilité, en résolution et en précision de mesure de masses, c'est le cas par exemple de l'orbitrap en 1999 et des trappes 2D en 2002 [Ouyang et al., 2007], [Ouyang et al., 2004], [Makarov et al., 2006], [Makarov, 2000].

Toutefois, cet analyseur nécessite de nombreuses optimisations pour obtenir des spectres de masse et des spectres de fragmentation de qualité. En effet, il s'agira d'optimiser les paramètres suivants : le cut-off, éviter les phénomènes d'espace charge en contrôlant le nombre d'ions accumulés dans la trappe, le choix du mode de balayage pour gagner en sensibilité ou gagner en résolution, le choix des critères de fragmentation (nombre d'ions précurseurs, nombre de MS/MS, amplitude de fragmentation, vitesse de balayage, nombre d'ions accumulés dans la trappe) et bien d'autres. De même, cet analyseur étant essentiellement utilisé en couplage LC, de nombreuses optimisations sont également nécessaires pour obtenir le maximum d'informations et de la meilleure qualité possible selon le type de chromatographie couplée. Afin d'illustrer toutes ces optimisations, je vous invite à consulter l'étude menée au laboratoire dans le cadre d'une thèse sur l'optimisation du couplage nanoHPLC-trappe ionique effectuée par Christine Carapito [Carapito, 2006] avec une trappe Esquire 3000+ (Bruker Daltonics) et une trappe HCTplus (Bruker Daltonics).

I.3. Les modes de fragmentation pour l'analyse des peptides

De nos jours, la spectrométrie de masse en tandem est devenue un outil incontournable pour la détermination de la structure en acides aminés d'une protéine mais surtout pour une caractérisation complète incluant l'identification de modifications post-traductionnelles ou de liaisons intermoléculaires entre protéines. Or pour pouvoir exploiter ces spectres de fragmentation, il est nécessaire de connaître les mécanismes de fragmentation des peptides.

A l'heure actuelle, les deux modes de fragmentations les plus utilisés sont la dissociation induite ou activée par collisions (CID ou CAD) et la dissociation par capture d'électrons (ECD). Ce mode de fragmentation utilisé sur des spectromètres de masse FT-ICR a été récemment adapté pour des analyseurs de type trappe ionique et se nomme la dissociation par transfert d'électron (ETD). Chacun d'eux permet de générer préférentiellement un certain type d'ions fragments.

I.3.1. Nomenclature de fragmentation des peptides

Les premiers spectres de fragmentation de peptides furent obtenus dans les années 1980 sur des spectromètres de masse à sources FAB. Roepstorff et Fohlman proposèrent alors une première nomenclature pour la fragmentation des peptides [Roepstorff and Fohlman, 1984]. Celle-ci fut simplifiée et complétée quelques années plus tard par Biemann (figure 17) [Biemann, 1988], [Biemann, 1990].

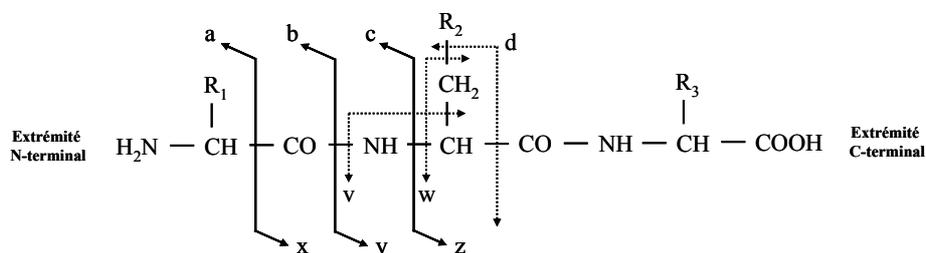


Figure 17: Nomenclature de fragmentation d'un peptide constitué de trois acides aminés (R1, R2 et R3) selon Biemann [Biemann, 1990].

Deux types d'ions fragments sont observés lors d'une fragmentation d'un peptide à basse énergie:

- les ions contenant l'extrémité N-terminal du peptide (a, b et c)
- les ions contenant l'extrémité C-terminal du peptide (x, y, z)

Dans le cas d'une fragmentation à haute énergie, il y aura également fragmentation des chaînes latérales pour donner des ions de type d, v et w en plus des autres.

Le plus couramment, les peptides protonés sont dissociés à basse énergie. Les ions générés majoritairement sont alors de type b et y puisque la liaison covalente la plus labile se situe au niveau de la liaison peptidique CO-NH. En prenant une série d'ions b et y successifs, il sera possible de déterminer la nature des acides aminés en calculant les différences de masses entre deux ions

fragments consécutifs qui correspondra à la masse d'un acide aminé. Selon la qualité des spectres de fragmentation des peptides, il est alors possible de remonter à leur composition en acides aminés.

Toutefois, dans le cas de certains acides aminés, il ne sera pas possible de trancher. C'est le cas par exemple de la leucine et de l'isoleucine qui ont la même masse.

De même, selon la précision de mesure de masse du spectromètre de masse, il ne sera pas possible de trancher dans le cas d'une lysine et d'une glutamine qui ne diffèrent en masse que de 0,036 Da.

Enfin, dans le cas de modifications post-traductionnelles de certains acides aminés, l'interprétation de spectres de fragmentation sera beaucoup moins aisée. C'est le cas par exemple d'une lysine diméthylée ($M = 156,149$ Da) qui aura une masse proche d'une arginine ($M = 156,101$ Da).

Néanmoins, ces difficultés peuvent être contournées en utilisant des spectromètres de masse aux résolutions élevées de $R = 10\ 000$ à $20\ 000$.

I.3.2. La fragmentation en mode CID (Collision Induced Dissociation)

Ce mode de fragmentation permet de dissocier des ions en les activant énergétiquement. Pour cela, on utilise typiquement un gaz inerte (par exemple de l'argon ou de l'hélium) qui va entrer en collision avec les molécules ionisées et leur conférer une énergie qui va les exciter et entraîner leur dissociation au niveau des liaisons covalentes. Selon la valeur de l'énergie de fragmentation mise en jeu, le type d'ions fragments observés seront différents. Il existe deux types de fragmentations CID :

- **Fragmentation CID basse énergie** qui met en jeu des énergies de fragmentation de quelques eV. C'est le cas des spectromètres de masse de type Q-TOF et trappe ionique constitués de sources d'ionisation douce (ESI). Les dissociations se feront préférentiellement au niveau des liaisons covalentes les plus fragiles. Pour un peptide, on générera essentiellement des ions fragments de type b et y.
- **Fragmentation CID haute énergie** qui met en jeu des énergies de fragmentation de l'ordre du keV. Ce type de fragmentation est observé en utilisant des spectromètres de masse de type TOF-TOF ou à secteurs magnétiques. Pour un peptide, on générera, en plus des autres séries d'ions fragments, des ions fragments au niveau des chaînes latérales (d, v et w).

A l'heure actuelle, le modèle le plus admis concernant le chemin de fragmentation à basse énergie des peptides protonés est « **le modèle du proton mobile** » [Paizs and Suhai, 2005]. Celui-ci est le résultat de plusieurs études menées entre autres par Wysocki, Harrison, Gaskell et Boyd dans les années 1990.

D'après ce modèle, les peptides protonés activés à basse énergie fragmentent par des réactions dirigées par la charge. Lorsque l'énergie interne d'un ion augmente jusqu'à l'exciter, les protons vont migrer vers les sites de protonation les moins favorable énergétiquement, et de ce fait entraîner une fragmentation [Dongré et al., 1996].

Toutefois, l'efficacité de fragmentation d'un peptide dépend de nombreux paramètres tels que sa composition en acides aminés (présence ou absence d'acides aminés basiques), sa séquence, sa taille, son état de charge, l'énergie de fragmentation utilisée et bien d'autres [Wysocki et al., 2000].

En effet de manière générale, les peptides comportent plusieurs sites de protonation (N-terminal, azote et oxygène de la liaison amide, groupements des chaînes latérales) conduisant à de nombreux isomères. Ces isomères se divisent en deux classes de peptides protonés :

- Un ou plusieurs sites de protonation sont favorisé énergétiquement et ou cinétiquement par rapport à d'autres sites, cela conduit à la « séquestration » du proton. De ce fait, beaucoup plus d'énergie sera nécessaire pour délocaliser ce proton sur d'autres sites et entraîner la fragmentation. C'est le cas par exemple de peptides monochargés contenant une arginine en position C-terminal.
- La plupart des sites de protonations sont accessibles dans des gammes d'énergies plus faibles. La fragmentation de ces isomères est donc plus aisée. Parmi cette classe, on compte majoritairement les peptides doublement chargés issus de la digestion à la trypsine. Ce sont ceux qui sont majoritairement observés et préférentiellement soumis à la fragmentation.

Dans le cas des peptides doublement chargés issus de la digestion trypsique, les charges sont préférentiellement portées par l'extrémité N-terminale et par la chaîne latérale de la lysine et de l'arginine en position C-terminale. Le transfert du proton porté par l'extrémité N-terminale nécessitant un apport d'énergie plus faible que pour celui positionné sur la lysine ou l'argine, les ions fragments préférentiellement générés seront de type b et y (figure 18).

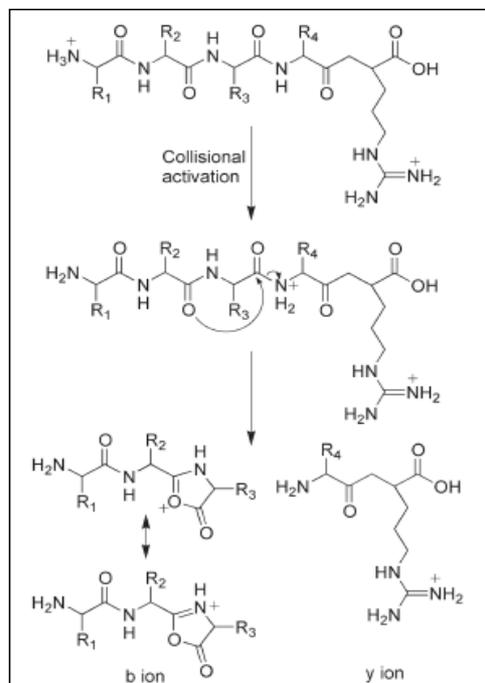


Figure 18 : Chemin de fragmentation d'un ion doublement chargé correspondant à un peptide issu de la digestion trypsique avec une arginine en position C-terminale et constitué de 5 acides aminés et conduisant à des fragments de type b et y [Lane, 2005].

I.3.3. La fragmentation en mode ETD (Electron Transfer Dissociation)

Ce mode de fragmentation découle directement du mode de fragmentation ECD (Electron Capture Dissociation) mis en place sur un spectromètre de masse FT-ICR pour la première fois en 1998 par Mc Lafferty et ses collaborateurs [Zubarev et al., 1998].

Le principe de la fragmentation ECD consiste à confiner des peptides ou des protéines protonées dans la trappe Penning du FT-ICR et de les exposer à des électrons thermiques. La capture de cet électron par le peptide protoné étant exothermique, la fragmentation de celui-ci ne se fera préférentiellement au niveau des liaisons peptidiques mais pourra se produire de manière aléatoire. Il présente ainsi l'avantage de générer d'autres séries d'ions (c et z) mais également de conserver les PTMs [Cooper et al., 2005].

Toutefois cette technique très prometteuse n'étant compatible qu'avec des spectromètres de masse FT-ICR, elle fut adaptée quelques années plus tard pour des trappes ioniques [Syka et al., 2004] notamment sur des trappes Bruker Daltonics et Agilent (figure 19).

A la différence de la fragmentation ECD, le principe de l'ETD est basé sur le transfert d'électrons par l'intermédiaire d'un réactif, un radical anion qui joue le rôle de transporteur de l'électron vers le peptide protoné. Le mécanisme est le suivant :

- Dans un premier temps, les peptides protonés sont piégés et isolés dans la trappe.
- Les radicaux anions (du fluoranthène par exemple), générés à l'extérieur de la trappe par ionisation chimique, sont alors introduits et accumulés dans la trappe en quantité suffisante pour induire une fragmentation des peptides protonés.

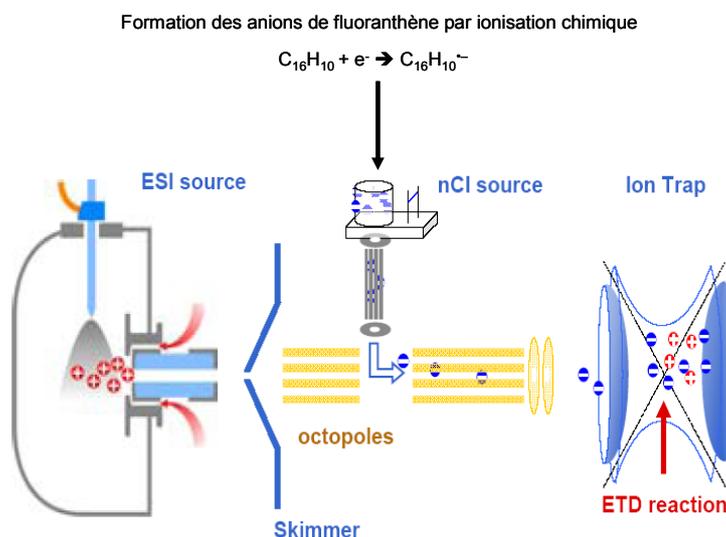
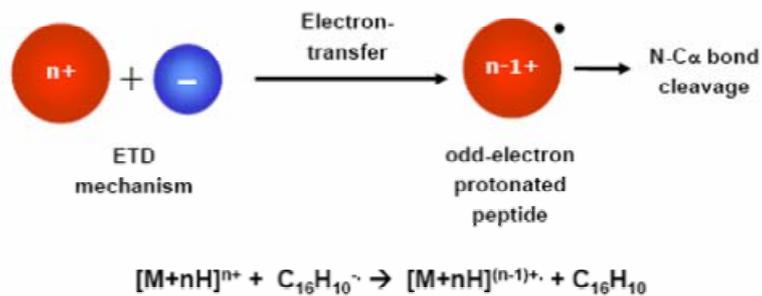


Figure 19 : Schéma de la trappe Bruker Daltonics équipée avec une source ETD.

- Les anions vont alors entrer en collision avec les peptides protonés entraînant ainsi un transfert d'électron sur celui-ci. Le peptide se fragmente de la même manière qu'en ECD, on génère des séries d'ions c et z (figure 20).

Schéma du transfert de l'électron vers le peptide protoné



Chemin de fragmentation ETD des peptides

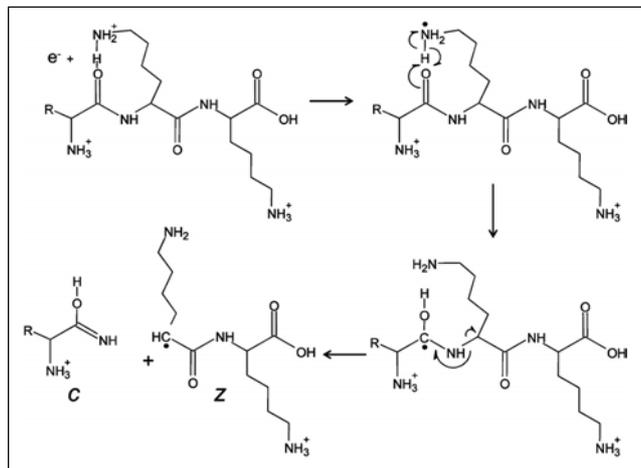


Figure 20 : Mécanisme ETD entraînant la fragmentation des peptides. Chemin de fragmentation ETD d'un peptide constitué de 3 acides aminés permettant de générer des ions fragments de type c et z [Syka et al., 2004].

Ce mode de fragmentation adapté sur des trappes ioniques, donc beaucoup plus facilement accessible, offre les avantages suivants :

- La fragmentation n'ayant pas de coupure préférentielle, le type d'ions fragments générés est différents et les spectres de fragmentations peuvent être plus riches et plus informatifs que les spectres obtenus par CID.
- La possibilité de caractériser des modifications post-traductionnelles (phosphorylation, glycosylation), en effet celles-ci sont conservées puisqu'il n'y a pas de fragmentation préférentielle.
- La fragmentation de peptides de taille élevée voire de protéines.

Néanmoins, l'interprétation de ces spectres de fragmentations très riches peut s'avérer rapidement compliquée. En effet, la fragmentation n'étant pas préférentielle, le nombre d'ions fragments va considérablement augmenter entraînant ainsi une perte de sensibilité.

De plus, ce mode de fragmentation peut entraîner la formation d'ions fragments multichargés. De ce fait, afin de déterminer les états de charges des ions fragments multichargés lors de l'interprétation, il sera nécessaire d'avoir une bonne résolution isotopique. Cette augmentation en résolution entraînera une perte en sensibilité.

I.4. Conclusion

Depuis la fin des années 1980, marquées par les premiers spectres de masse de macromolécules biologiques obtenus grâce au développement des sources ESI et MALDI, la spectrométrie de masse a pris une place centrale dans l'analyse protéomique [Cravatt et al., 2007], [Aebersold and Mann, 2003]. Celle-ci permet d'identifier, de caractériser et de quantifier les protéines dans un extrait.

Néanmoins l'analyse d'extraits toujours plus complexes a rapidement conduit à améliorer les caractéristiques des spectromètres de masse et à développer des spectromètres de masse hybrides.

A ce jour, les 4 types d'analyseurs (Q, TOF, IT, FT-ICR) peuvent être couplés à des sources ESI et/ou des sources MALDI. De plus, la spectrométrie de masse en tandem s'est également développée (Q-Q-TOF, les TOF-TOF, les Q-Q-Q et les Q-Q-LIT).

Chacun de ces systèmes présente des caractéristiques qui sont primordiales selon la nature des échantillons et le contexte biologique des études. Celles-ci sont décrites dans le tableau 1 [Domon and Aebersold, 2006].

	IT-LIT	Q-Q-TOF	TOF-TOF	FT-ICR	Q-Q-Q	Q-Q-LIT
Précision de masse	Faible	Bon	Bon	Excellent	Moyen	Moyen
Résolution	Faible	Bon	Elevé	Très élevé	Faible	Faible
Sensibilité	Bon		Elevé	Moyen	Elevé	Elevé
Gamme dynamique	Faible	Moyen	Moyen	Moyen	Elevé	Elevé
ESI	√	√		√	√	√
MALDI	(√)	(√)	√			
MS/MS	√	√	√	√	√	√
Autre	Séq. MS/MS			Ions précurseurs, pertes de neutres, MRM		
Identification	++	++	++	+++	+	+
Quantification	+	+++	++	++	+++	+++
Détection de modifications	+++	++	+++	++	++	++
Capacité	+	+	+	+		+++

Tableau 1 : Caractéristiques et performances des spectromètres de masse couramment utilisées [Domon and Aebersold, 2006].

Actuellement, la spectrométrie de masse peut être utilisée soit pour mesurer la masse précise de macromolécules biologiques, soit pour déterminer la séquence en acides aminés de ces macromolécules par fragmentation CID, ECD ou plus récemment ETD. Celles-ci ont conduit au développement de plusieurs stratégies d'identification où la bioinformatique tient une place importante.

II. Les stratégies d'identification par spectrométrie de masse des protéines

Nous avons vu précédemment que selon la nature du spectromètre de masse utilisé, la nature des données de spectrométrie de masse générées peut différer (nécessité de la fragmentation ou non, fragmentation possible ou non, quel type de fragmentation).

En effet, en utilisant des spectromètres précis en masse, il est possible de faire de l'identification en fonction uniquement des masses des ions. C'est le cas par exemple du MALDI-TOF ou encore du FT-ICR. Toutefois, utiliser ce type d'approche en protéomique peut rapidement être limitant, notamment dans le cas d'échantillons complexes ou encore dans le cas de recherche de modifications post-traductionnelles ou de mutation sur les chaînes protéiques.

Dans ces cas là, il est alors plus pertinent de travailler sur des approches de séquençage en utilisant les données de masse issues des spectres de fragmentation. Plus informative, cette technique permet de remonter à la composition en acide aminé mais également de mettre en évidence des modifications post-traductionnelles ou des mutations sur des protéines.

Cependant ces stratégies d'identifications par masse ne vaudraient rien sans les nombreux progrès apportés en bioinformatique avec le développement de nouveaux logiciels automatisés pour l'interprétation des données de masse conduisant à l'identification de protéines.

Nous verrons donc dans un premier temps quelles sont à l'heure actuelle les deux types de stratégies utilisées actuellement pour l'identification de protéines qui sont l'empreinte peptidique massique PMF (Peptide Mass Fingerprint) et le séquençage PFF (Peptide Fragment Fingerprint) puis nous verrons le mode de fonctionnement de ces logiciels bioinformatiques.

II.1. Identification par « Peptide Mass Fingerprint » (PMF)

Dans les années 1989, la méthode de choix pour identifier une protéine était le séquençage par dégradation d'Edman. Toutefois, ce type d'analyses d'une durée considérable (1 heure par acide aminé) et nécessitant des échantillons assez purs, fut vite mise de côté suite aux développements instrumentaux apportés en spectrométrie de masse. L'analyse par spectrométrie de masse permettant de travailler sur des échantillons plus complexes, le concept de l'identification par « empreinte peptidique massique » fut décrit dans les années 1990 par cinq groupes de chercheurs [Mann et al., 1993], [Henzel et al., 1993], [James et al., 1993], [Pappin et al., 1993], [Yates et al., 1993].

Généralement associée à une étape de séparation des protéines sur gels d'électrophorèse, cette stratégie consiste à (1) digérer les protéines en peptides avec une enzyme spécifique, (2) mesurer très précisément, pour un échantillon donné, l'ensemble des masses des peptides présents dans cet échantillon, (3) comparer via des moteurs de recherche cette liste de masses mesurées aux listes de masses théoriques générées *in silico* à partir des banques protéiques (figure 21). L'interprétation de ces données se fait à l'aide de moteurs de recherches automatisés tels que par exemple MASCOT.

Classiquement, les mesures de masses se font sur des instruments de type MALDI-TOF qui ont des résolutions élevées (> 10000) et qui sont précis en mesure de masse (10-20 ppm). De plus ce type d'analyses automatisables et rapides, présentent l'avantage d'être tolérantes aux sels et aux détergents. Afin de valider une identification par PMF, plusieurs critères sont à prendre en compte :

- Le pourcentage de recouvrement de la protéine, c'est à dire le nombre de peptides identifiés pour une protéine.
- La précision de masse sur chaque peptide identifié.

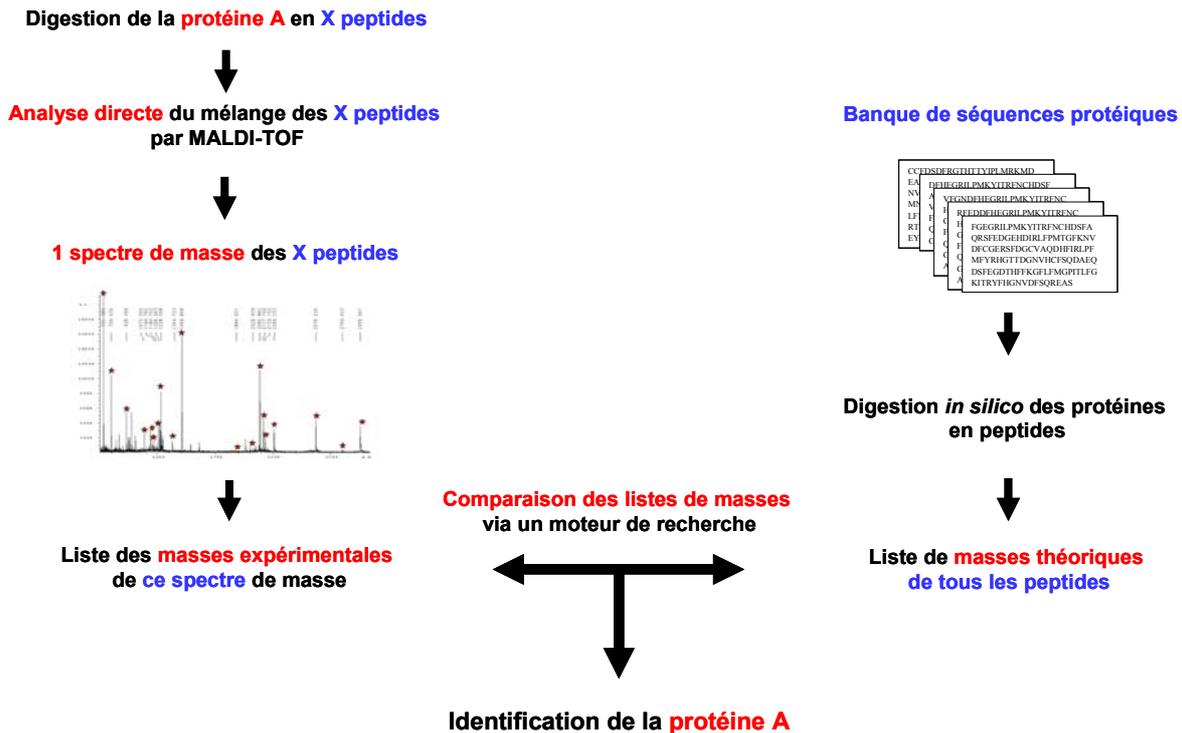


Figure 21 : Illustration de la stratégie d'identification par PMF à partir d'analyses MALDI-TOF-MS.

Toutefois cette technique bien adaptée pour de l'analyse à haut débit présente de nombreux inconvénients.

En effet, dans le cas d'échantillons complexes cette technique est peu adaptée puisque nous serons confrontés à des superpositions d'empreintes peptidiques massiques difficilement exploitables.

De plus, selon la nature des peptides générés par digestion enzymatique, il n'est pas exclus de se trouver dans des situations où les masses des peptides seront trop proches pour être attribuées à une protéine en particulier.

De même, la taille des banques protéiques ne faisant qu'augmenter, le nombre de peptides théoriques de masses proches peut lui aussi considérablement augmenter rendant l'identification par PMF de plus en plus complexe.

Enfin n'ayant aucune information de séquence, la stratégie PMF n'est pas adaptée pour la mise en évidence de modifications post-traductionnelles ou de mutations.

II.2. Identification par nanoHPLC-MS/MS : les modes d'interprétation des spectres MS/MS

L'utilisation du couplage nanoHPLC-MS/MS pour l'identification de protéines est à ce jour une des techniques les plus utilisées en analyse protéomique. Cette stratégie est souvent associée à (1) une étape de séparation des protéines sur gels d'électrophorèses, (2) une digestion des protéines avec une enzyme spécifique, (3) une séparation des peptides et fragmentation de ceux-ci par nanoHPLC-MS/MS puis (4) interprétation des spectres de fragmentation et comparaison aux banques de séquences. Elle présente de nombreux avantages par rapport à la stratégie PMF.

En effet, en ajoutant une étape de chromatographie liquide, cette stratégie permet l'analyse de mélanges de peptides beaucoup plus complexes puisque les peptides seront séparés avant d'entrer dans le spectromètre de masse réduisant ainsi considérablement une suppression de signal au niveau de l'ionisation. Une première étude utilisant cette stratégie fut menée en 1992 sur des mélanges complexes par Hunt et al [Hunt et al., 1992].

De même, cette étape de séparation permet d'effectuer de la fragmentation sur chacun de ces peptides protonés séparés mais également et surtout de pouvoir travailler en automatique. L'ensemble de ces spectres de fragmentation beaucoup plus informatifs permettront alors d'identifier des protéines plus aisément mais également de mettre en évidence la présence de modifications post-traductionnelles ou des mutations sur les séquences.

En 1994, Yates, McCormack et Eng proposent une stratégie d'identification similaire à la stratégie PMF mais cette fois-ci en utilisant les données générées par les spectres de fragmentation [Eng et al., 1994]. La même année, Mann et al développent le « Peptide Sequence Tag » qui consiste à déduire des séquences partielles en acides aminés en utilisant les différences de masses observées entre plusieurs fragments et de les comparer aux banques de données [Mann and Wilm, 1994].

Suite à cela de nouvelles stratégies d'identification par séquençage furent proposées. De même de nombreux moteurs de recherche pour l'interprétation des spectres de fragmentations furent développés [Nesvizhskii et al., 2007].

A ce jour les deux stratégies pour l'interprétation des spectres MS/MS les plus utilisées sont le « Peptide Fragment Fingerprint » (PFF) et le séquençage *de novo*. Nous décrirons ici brièvement leur principe ainsi que les moteurs de recherches dédiés à cet effet et leur mode de fonctionnement.

II.2.1. Interprétation des spectres MS/MS par comparaison avec des listes de masses théoriques issues des banques de séquences

a) Le « Peptide Sequence Tag »

Cette approche fut introduite pour la première fois en 1994 par Mann et Wilm [Mann and Wilm, 1994]. Son principe réside sur le fait que tout spectre de fragmentation de peptide contient toujours au moins une série d'ions fragments permettant de lire une partie de la séquence du peptide, on l'appelle le « tag ». Par comparaison entre la masse du peptide parent et les masses des ions fragments situés

aux deux extrémités de ce tag, il est alors possible de déduire manuellement les masses m_1 et m_2 correspondants aux acides aminés situés à chacune des extrémités de ce tag qui n'ont pas été lues dans le spectre de fragmentation. Le « Peptide Sequence Tag » se constitue donc de trois parties : la séquence partielle lue, le tag, et deux masses correspondants aux extrémités de ce tag non lues (m_1 et m_2) (figure 22).

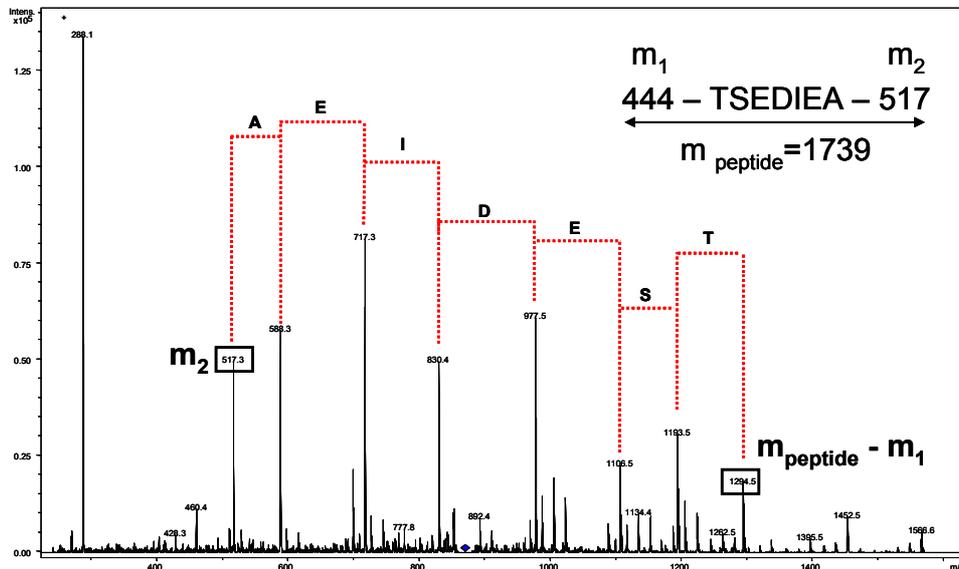


Figure 22 : Exemple de détermination du « Peptide Sequence Tag ».

Ces « Peptide Sequence Tags » peuvent alors être soumis à différents moteurs de recherche pour permettre une identification. Toutefois, cette approche nécessite un temps d'interprétation considérable et actuellement elle est de moins en moins utilisée.

En effet, en 1994, Mann et Wilm estimait qu'un tag de 3 acides aminés était suffisant pour identifier un peptide unique, or avec l'accroissement du nombre d'entrées dans les banques protéiques, celle-ci a été remplacée par la stratégie « Peptide Fragment Fingerprint » (PFF).

b) Le « Peptide Fragment Fingerprint » (PFF)

Le terme PFF fut introduit pour la première fois en 2004 par Blueggel et al. [Blueggel et al., 2004]. Cette approche est la plus couramment utilisée.

Elle est basée sur une comparaison des listes de masses des ions fragments mesurés dans chaque spectres de fragmentation avec des listes de masses de spectres de fragmentation théoriques des protéines présentes dans des banques ou des bibliothèques via des moteurs de recherche (figure 23). Il existe de nombreux moteurs de recherche [Nesvizhskii et al., 2007].

Nous ne décrivons ci-dessous que deux moteurs de recherche accessibles publiquement et les plus couramment utilisés:

- **MASCOT** (<http://matrixscience.com>)

Son algorithme est basé sur une approche probabilistique [Perkins et al., 1999]. Cela signifie que les masses des fragments mesurées sont comparées aux masses des fragments théoriques générés *in silico* à partir des peptides présents dans les banques. La probabilité d'identité est alors mesurée par un « score ».

- **SEQUEST** (<http://www.thermo.com>)

Son algorithme est basée sur un processus d'auto-corrélation [Eng et al., 1994]. Les peptides présents dans les banques sont utilisés pour construire des spectres de fragmentation théoriques qui seront corrélés avec les spectres de fragmentation expérimentaux.

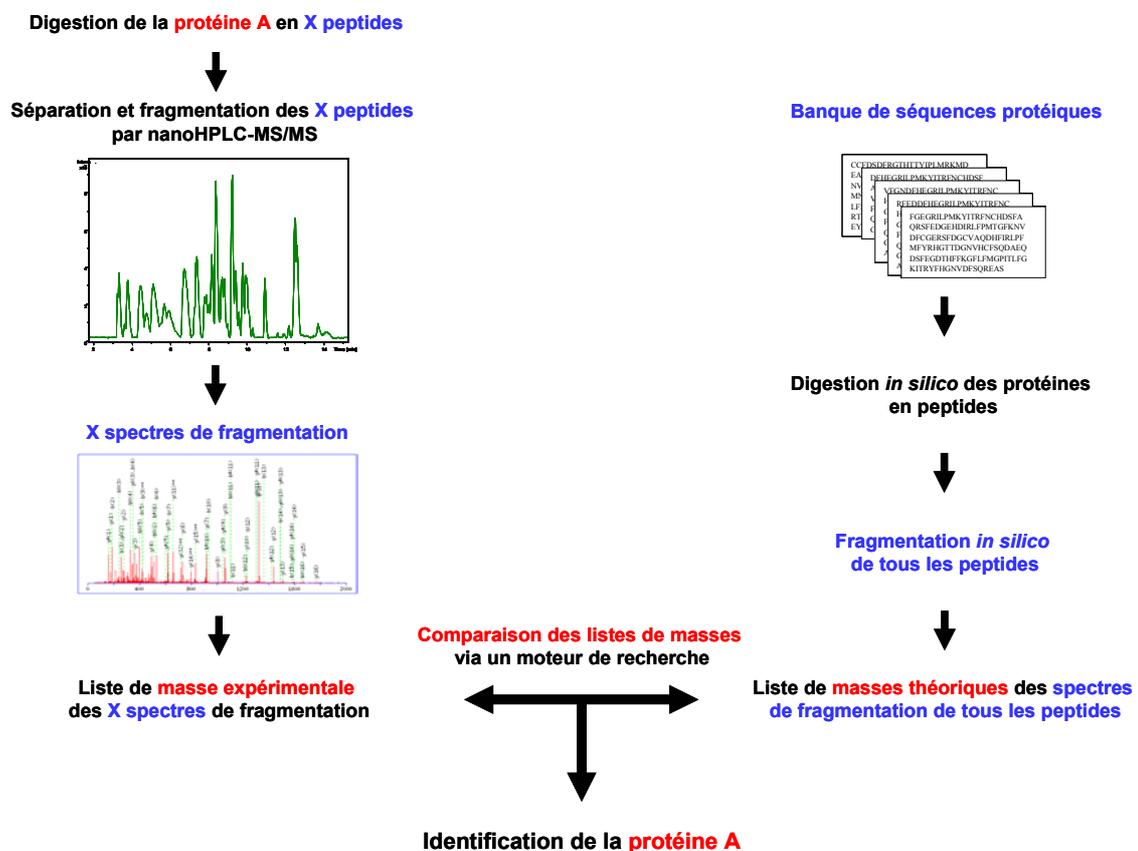


Figure 23 : Approche PFF : séquençage automatique par nanoHPLC-MS/MS.

II.2.3. Interprétation des spectres MS/MS par séquençage *de novo*

Une autre alternative est le séquençage *de novo*. Dans cette approche, les séquences en acides aminés sont directement lues sur les spectres expérimentaux. Initialement, ces lectures se faisaient manuellement et à l'aide des outils informatiques fournis par les constructeurs de spectromètres de masses (par exemple DataAnalysis de Bruker Daltonics).

Toutefois, plus récemment, de nombreux outils bioinformatiques furent développés afin d'automatiser ce séquençage directement dans les spectres. C'est le cas par exemple du logiciel Peaks développé par le Department of Computer Science, University of Western Ontario, London, Canada [Ma et al., 2003].

Les listes de masses des spectres de fragmentation sont soumises au logiciel qui va rechercher en automatique les possibilités de séquences en acides aminés pour chaque spectre de fragmentation. Ces recherches peuvent être affinées en fixant des paramètres tels que l'enzyme de digestion utilisée ou encore les modifications post-traductionnelles éventuellement présentes.

Celui-ci fait alors des propositions de séquences pour chacun des spectres MS/MS. L'ensemble de ces séquences sont ensuite soumises à des programmes de recherche de similarités de séquences, par exemple MS-BLAST [Shevchenko et al., 2001]. Ce type de programme va comparer les séquences proposées par Peaks aux séquences présentes dans les banques (figure 24).

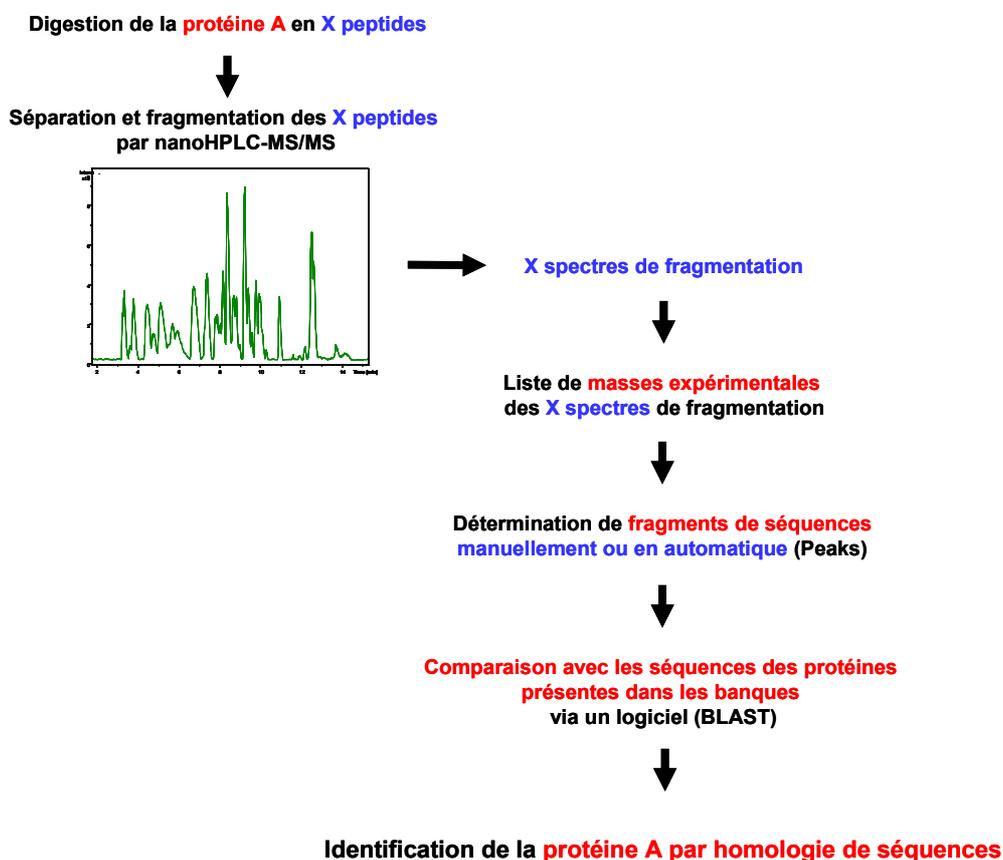


Figure 24 : Schéma de la stratégie pour le séquençage *de novo*.

Même si ce type d'approche nécessite d'avoir des spectres de très bonne qualité et que le temps d'interprétation est considérablement augmenté, elle peut souvent se révéler très utile.

En effet, l'approche « Peptide Fragment Fingerprint » peut rapidement être limitante notamment dans le cas d'études menées sur des organismes peu ou non séquencés ou encore dans le cas de recherches de modifications post-traductionnelles ou de mutations sur les chaînes d'acides aminés.

II.2.4. Les outils bioinformatiques pour l'interprétation des spectres de fragmentation

A l'heure actuelle, il existe de nombreux logiciels privés ou publics pour l'interprétation et la validation des données de spectres de fragmentations. L'ensemble de ces logiciels est décrit par Neshvizhskii dans un article paru fin 2007 dans Nature Methods [Neshvizhskii et al., 2007]. Parmi ces logiciels, on trouve par exemple:

- des outils de recherche dans les banques tels que MASCOT, SEQUEST ou Phenix
- des outils pour le séquençage *de novo* tels que Peaks ou PepNovo.
- des outils pour la validation statistique des identifications comme par exemple ProteinProphet, PeptideProphet et Scaffold
- des outils pour les approches de tags de séquences (Inspect, Popitam, GutenTag)
- des outils pour la comparaison de spectres (SpectraST, Biblispec)

II.3. Les banques de séquences utilisées pour l'identification de protéines

Avec les énormes progrès réalisés pour automatiser les stratégies de séquençage des génomes et le premier génome complet d'*Haemophilus influenzae* séquencé dans les années 1990 [Fleischmann et al., 1995], le nombre de projets de séquençage de génomes et le nombre de génomes complets séquencés n'a fait qu'augmenter depuis ces 10 dernières années. Afin de gérer toutes ces données, de nombreuses banques informatiques furent créées répertoriant toutes sortes de données biologiques telles que les structures chimiques, des cartographies physiques ou génétiques, les taux d'expression des gènes, et bien d'autres. Une nouvelle discipline était née, la bioinformatique.

Parmi ces banques, les plus nombreuses et les plus volumineuses sont les banques de séquences. On compte deux types de banques de séquences, les banques de séquences nucléotidiques et les banques de séquences protéiques. Celles-ci seront brièvement décrites ci-dessous.

II.3.1. Les banques de séquences nucléotidiques

A l'heure actuelle, trois banques de séquences nucléotidiques travaillent en étroite collaboration :

- EMBL créée en 1980 à l'EMBL [Kulikova et al., 2004]
- GenBank créée en 1982 au Los Alamos National Laboratory [Benson et al., 2003]
- DDBJ (DNA Data Bank of Japan) créée en 1986 au Centre d'Information Biologique de l'Institut National de Génétique au Japon (European Molecular Biology Laboratory) [Miyazaki et al., 2003].

Elles intègrent l'ensemble des séquences nucléotidiques primaires toutes origines, organismes et espèces moléculaires confondues. Leur but est de fournir des banques les plus complètes possible, de ce fait, toute nouvelle séquence produite est intégrée et il n'est pas exclu d'avoir affaire non seulement à des redondances mais également à une hétérogénéité des données non négligeable.

II.3.2. Les banques de séquences protéiques

L'ensemble des banques protéiques disponibles à ce jour peuvent être répertoriées selon deux catégories distinctes [Apweiler et al., 2004a] :

a) Les banques généralistes de dépôts

Elles sont constituées majoritairement de séquences de protéines issues de la traduction des gènes prédits sans vérification et avec des annotations fonctionnelles limitées. On y retrouve souvent des redondances.

- **GenPept (GenBank Gene Products Data Bank):** produite par le National Center of Biotechnology Information (NCBI), les séquences des protéines y étant répertoriées découlent directement de la translation des séquences nucléotidiques présentes dans les banques GenBank/EMBL/DDBJ et contiennent des annotations minimales. Chaque protéine peut y être représentée par plusieurs entrées sans que celles-ci soient regroupées en une entrée.
- **NCBI's Entrez Protein (Entrez Protein du National Center for Biotechnology Information) :** cette banque un peu particulière contient à la fois les séquences des protéines issues de la translation automatique des banques GenBank/EMBL/DDBJ et les séquences protéiques annotées de Swiss-Prot, de PIR (Protein Information Resource), de RefSeq et de PDB (Protein Data Bank). A la différence de GenPept, on peut trouver pour chaque entrée des informations complémentaires extraites des banques Swiss-Prot et PIR. Cette banque est donc très complète mais présente l'inconvénient d'être redondante et de présenter des erreurs dans les annotations.
- **RefSeq (Reference Sequence):** produite par NCBI, cette banque a pour projet de fournir un panel de séquences protéiques de référence non redondantes. Cette banque s'emploie à éviter toute redondance, à faire le lien entre les séquences nucléotidiques et protéiques, à faire des mises à jour régulières, à valider les données et à uniformiser les formats mais également à attribuer des numéros d'accès distincts. Cependant, la plupart des protéines y étant répertoriées le sont automatiquement et très peu de séquences sont vérifiées manuellement. De plus, elle n'existe que pour un certain nombre d'espèces (1100 virus et 150 bactéries) et très peu d'organismes.

b) Les banques de séquences corrigées

Les séquences répertoriées ont été vérifiées par des travaux de nettoyage, de tri, de documentation et d'annotation. Ce sont généralement ces banques qui sont utilisées en analyse protéomique. Parmi les banques de séquences corrigées, les plus couramment utilisées sont les suivantes :

- **PIR-PSD (Protein Information Resource Protein Sequence Database)** : elle fut créée en 1984 pour succéder à la banque National Biomedical Research Foundation Protein Sequence Database. Les séquences des protéines y sont organisées par familles et super-familles et annotées avec des données fonctionnelles, structurales, génétiques et bibliographiques. De plus pour chaque séquence on peut trouver des informations du type : nom de la protéine, classification, organisme, fonction, caractéristiques, références bibliographiques et bien d'autres.
- **Swiss-Prot** : créée en 1986 par Amos Bairoch, elle est l'une des mieux corrigée et des moins redondante. Chaque entrée est vérifiée par des analyses poussées et annotée par des biologistes en utilisant les données expérimentales extraites de la littérature et des programmes d'analyses de séquences pour confirmer manuellement les résultats. Ainsi, ces annotations contiennent également tout un ensemble d'informations concernant chaque protéine tel que leur fonctionnalité, leur nature, leur structure, la présence de modifications post-traductionnelles, la possibilité de mutations, leur similarité avec d'autres protéines, leur implication dans des maladies, zone d'expression, tissus dans lesquels on la trouve et bien d'autres.
- **TrEMBL** : Les travaux de vérification des séquences présentes dans Swiss-Prot prenant énormément de temps et le nombre de séquences protéiques à vérifier ne cessant d'augmenter, une banque complémentaire fut créée, la banque TrEMBL donnant rapidement accès aux utilisateurs aux nouvelles séquences protéiques. On y trouve d'une part toutes les séquences déduites informatiquement de la translation des séquences codantes de DDBJ/EMBL/GenBank qui ne sont pas présentes dans Swiss-Prot, mais également des séquences protéiques issues de publications ou directement déposées par les utilisateurs. Le travail principal de cette banque est (1) de combiner informatiquement toutes les séquences identiques afin d'éviter des redondances, et (2) d'améliorer automatiquement les informations contenues dans TrEMBL. Ce processus est basé sur un système standardisé de transfert des annotations des protéines en comparant ces séquences non annotées aux séquences annotées dans Swiss-Prot. Une fois ces vérifications faites, ces séquences pourront être traitées par des experts qui jugeront de la pertinence de ces pré-annotations et qui pourront se réserver le droit de les modifier ou encore de les supprimer.

c) UniProt (The Universal Protein Resource) : une nouvelle génération de bases de données protéiques

Début 2000, le National Institutes of Health autorise la fusion des banques Swiss-Prot, TrEMBL et PIR-PSD en une seule ressource, UniProt [Apweiler et al., 2004b]. Elle fournit dans une seule ressource des séquences protéiques annotées et des informations fonctionnelles. Elle se compose de trois types de bases de données :

- **UniProt Knowledgebase (UniProt)** : elle correspond à la combinaison des banques de séquences corrigées Swiss-Prot, TrEMBL et PIR-PSD. Toutes les séquences présentes dans l'une

ou l'autre de ces trois banques est ajouté dans UniProt tout en évitant des redondances. Elle se compose de deux parties : (1) une section « **UniProt/Swiss-Prot** » contenant les entrées annotées manuellement et vérifiées et (2) une section « **UniProt/TrEMBL** » contenant les entrées issues d'analyses statistiques et qui doivent encore être vérifiées et annotées manuellement.

- **UniProt Archive (UniParc)** : cette banque publique non redondante archive l'ensemble des séquences protéiques présentes dans les banques généralistes de dépôts et dans les banques de séquences corrigées (Swiss-Prot, PIR-PSD, TrEMBL, EMBL, Ensembl, International Protein Index, PDB, RefSeq, FlyBase, WormBase). Lorsqu'une séquence est répertoriée dans plusieurs banques ou est plusieurs fois répertoriée dans une même banque, UniParc s'emploie à ne la répertorier qu'une fois et lui attribue un identifiant. De plus elle fournit également d'autres types d'informations concernant chaque séquence (numéro d'accession dans la banque d'origine, versions des séquences, ...). Enfin elle donne également le statut de la séquence entrée dans la base de donnée « source » (active ou obsolète) permettant ainsi de savoir si la séquence existe toujours dans la base de données de départ ou non.
- **UniProt Reference (UniRef)** : Trois bases de données UniRef (UniRef100, UniRef90 et UniRef50) ont été créées à partir de UniProt Knowledgebase et UniParc pour former une base de données de séquences protéiques avec un contenu d'informations très élevé. **UniRef100** contient une collection de séquences non-redondantes regroupées par identité de séquence et taxonomie avec attribution à une source. Elle fusionne les séquences issues de différentes espèces. Les séquences et les fragments identiques sont représentés par une seule entrée qui permet de remonter aux informations suivantes par exemple : numéro d'accession, banques dans lesquelles elle est présente, taxonomie, bibliographie, **UniRef90** et **UniRef50** sont construites à partir de UniRef100. Elles se composent de séquences non-redondantes utilisées pour les recherches d'homologies de séquences. Toutes les entrées des différents organismes de sources ayant des homologies de séquences communes > 90% ou > 50% sont fusionnées en une seule entrées avec un lien vers leurs entrées dans UniProt Knowledgebase.

II.4. Conclusion

Nous avons pu voir dans ce chapitre qu'il existe différentes stratégies d'identification de protéines par spectrométrie de masse (figure 25) [Aebersold and Goodlett, 2001].

Toutefois selon la nature des échantillons biologiques et le contexte de l'étude menée le choix de la stratégie est primordial. En effet, plusieurs paramètres seront à prendre en compte :

- question posée (identification, caractérisation, recherche de modifications post-traductionnelles, mutation, quantification, ...)
- nature des échantillons (complexes ou non)
- mesure de masses précises ou fragmentation
- séquençage automatique ou séquençage de novo

- choix du moteur de recherche dans les banques
- choix de la banque (nucléotidique ou protéique, lesquelles)
- choix des critères de recherche (se limiter à un organisme ou rechercher dans une banque complète)

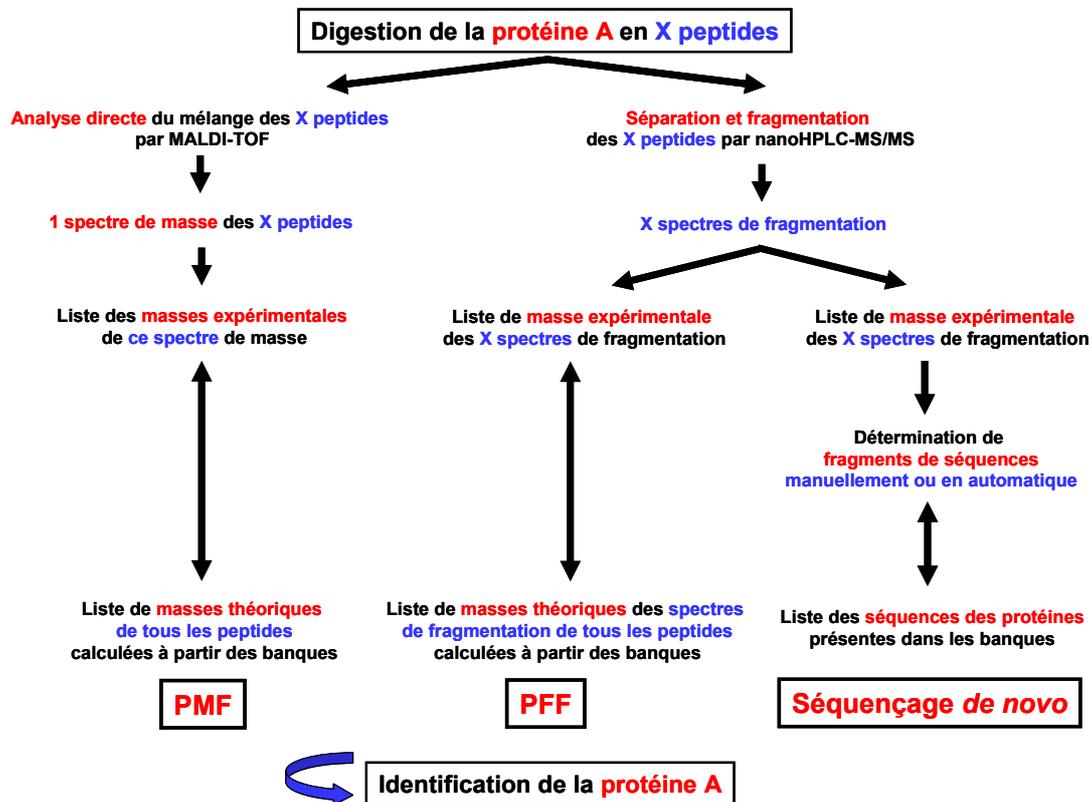


Figure 25 : Stratégies d'identification par spectrométrie de masse.

De plus, depuis quelques années, les directives de publication des données de masses dans des journaux scientifiques sont devenues de plus en plus strictes notamment dans le cas d'identification de protéines avec un peptide ou dans le cas d'identifications de PTMs isobares (acétylation vs. triméthylation, phosphorylation vs. sulfatation) [Bradshaw et al., 2005], [Wilkins et al., 2006].

C'est pourquoi, il n'est pas exclu de se demander si dans certains cas, il n'est pas plus judicieux de faire des recherches dans plusieurs banques ou encore de soumettre une analyse à différents moteurs de recherche.

III. Les stratégies de séparations pour caractériser les protéines

Avec les nombreux développements instrumentaux en spectrométrie de masse et les progrès apportés dans la validation et l'automatisation des interprétations de ces données de masse permettant

de pouvoir valider des identifications sans ambiguïté, l'analyse protéomique a fait un bond énorme dans l'étude de processus biologiques et une meilleure compréhension dans la fonction des gènes dans l'ère post-génomique [Cravatt et al., 2007], [Tyers and Mann, 2003], [Pandey and Mann, 2000].

En effet, il existe différentes stratégies d'identification des protéines par spectrométrie de masse (choix du spectromètre de masse, fragmentation ou non, PMF ou PFF, algorithme de recherche, accroissement du nombre et de la taille des banques de séquences).

Le choix de la stratégie est souvent dicté en fonction de la problématique abordée (identification de l'ensemble des protéines d'un extrait, identification de nouvelles protéines, quantification de certaines protéines,...) mais également de la nature des échantillons (complexes ou non, solubles ou non, ...) qui nécessitent souvent des étapes de séparations en amont.

De ce fait, afin de rendre possible l'ensemble de ces études, les techniques séparatives pour l'analyse de protéines furent également développées en parallèles tels que les gels d'électrophorèse et la chromatographie liquide [Issaq, 2001], [Issaq, 2003].

Toutefois selon la nature des protéines présentes dans ces échantillons, elles peuvent avoir des propriétés physico-chimiques incompatibles avec une technique ou l'autre, et il est souvent nécessaire de faire un choix ou encore d'utiliser la combinaison de plusieurs techniques pour pouvoir mener à bien une étude. A l'heure actuelle, l'étude de mélanges complexes peut se faire de deux manières :

- **Les gels d'électrophorèse**
- **La chromatographie liquide**

III.1. Les techniques de séparation des protéines : l'électrophorèse

III.1.1. L'électrophorèse bidimensionnelle sur gel (gel 2D)

Historiquement, la séparation de molécules sur des gels d'électrophorèse remonte aux années 1970 [Kenrick and Margolis, 1970], [Kaltschmidt and Wittmann, 1970]. Mais ce n'est que 5 années plus tard que furent publiés 5 articles par Klose, Scheele, Iborra et O'Farrell, dans lesquels on propose de séparer les protéines en combinant la focalisation isoélectrique et l'électrophorèse sur gels de polyacrylamide (SDS-PAGE) [Anderson and Anderson, 1996]. Depuis cette technique a été énormément améliorée et aujourd'hui elle fait partie des plus couramment utilisées pour séparer les protéines.

L'électrophorèse bidimensionnelle sur gel SDS-PAGE consiste à séparer les protéines selon leur charge (1^{ère} dimension du gel) et selon leur taille (2^{nde} dimension du gel). Une fois les protéines séparées, le gel est coloré pour pouvoir détecter les zones contenant les protéines. Ces zones sont alors découpées, les protéines sont réduites, alkylées et digérées à la trypsine. Les peptides sont extraits et analysés soit par MALDI-TOF ou soit par nanoHPLC-MS/MS.

L'ensemble de ces étapes sont très bien décrites par Rabilloud T. dans l'ouvrage « *Proteome Research : Two-dimensional Gel Electrophoresis and identification Methods* » édité par Rabilloud T. [Rabilloud, 2000]. Elles seront brièvement reprises ci-dessous.

a) Première dimension

Sur un strip de gel, sous l'effet d'un champ électrique, les protéines migrent via un gradient de pH vers leur point isoélectrique (pI) qui correspond au point où les protéines n'ont pas de charge. Ce pI dépend essentiellement de la nature des acides aminés la constituant (figure 26). Il existe actuellement deux manières de procéder :

- **CA-IEF (Carrier Ampholyte-Isoelectric focusing)** : La première consiste à séparer les protéines dans un gradient de pH généré en appliquant un champ électrique sur un gel contenant un mélange d'ampholytes transporteurs. Les ampholytes sont des molécules de faibles poids moléculaires contenant des groupements amino et carboxyl [An der Lan and Chrambach, 1985].
- **IPG (Immobilized pH Gradient)** : La seconde consiste à générer le gradient de pH par différents types de molécules chimiques, les immobilines. Les immobilines sont des dérivés d'acrylamides contenant des groupements amino et carboxyl. En copolymérisant avec le gel d'acrylamide, elles permettent de générer un gradient de pH immobilisé [Gorg et al., 1988], [Bjellqvist et al., 1982].

Chacune de ces techniques présente des avantages et des inconvénients, néanmoins, la technique actuellement la plus utilisée est l'IPG.

En effet, même si l'utilisation d'ampholytes permet d'obtenir des résolutions séparatives extrêmement élevées, cette technique présente l'inconvénient de n'être pas très reproductible spatialement du fait de l'instabilité du gradient de pH, ce qui n'est pas le cas de l'IPG.

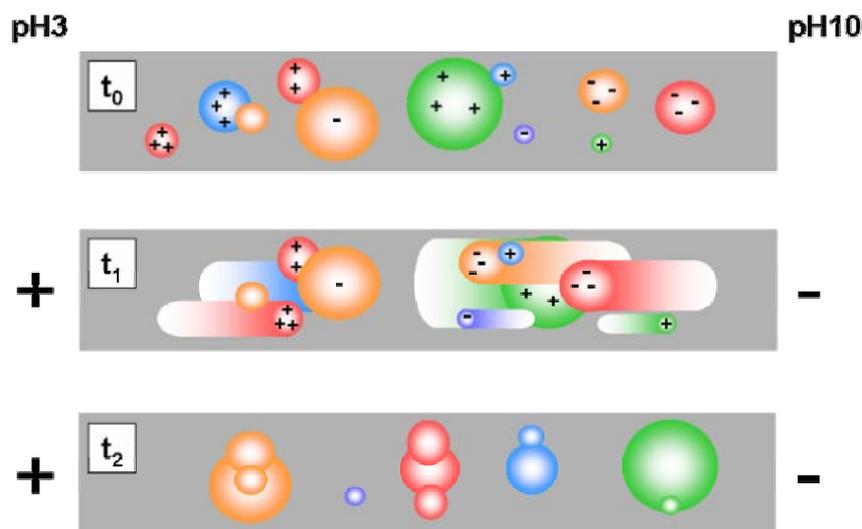


Figure 26 : 1^{ère} dimension du gel, séparation des protéines sur un strip IPG en fonction de leur charge à t_0 (pas de champ électrique), t_1 (application d'un champ électrique qui fait migrer les protéines en fonction de leur charge) et t_2 (temps final où toutes les protéines ont atteint leur point isoélectrique).

b) Seconde dimension

Une fois les protéines séparées selon leur charge sur un strip, celui-ci est récupéré pour être placé au sommet d'un gel de polyacrylamide de granulométrie variable en présence d'un agent dénaturant anionique, le sodium dodecyl sulfate (SDS). Le SDS interagit avec les protéines en entraînant leur dénaturation et en leur donnant une charge négative [Laemmli, 1970]. Cette charge négative va permettre d'accélérer les protéines dans un champ électrique tandis que la masse moléculaire détermine la distance de migration selon la réticulation du gel (figure 27).

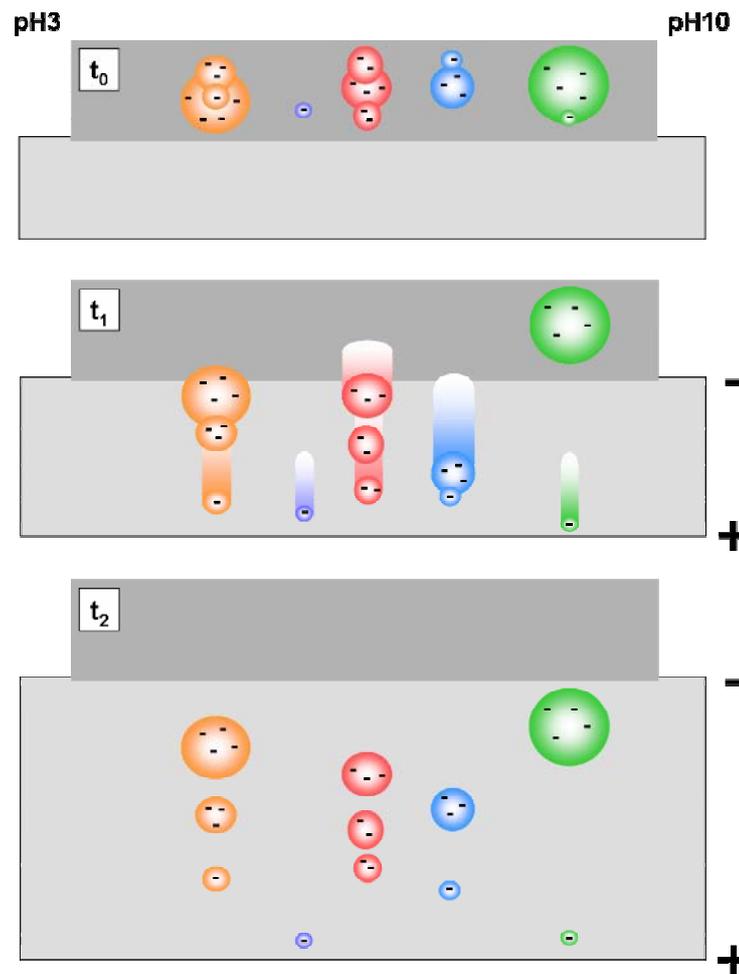


Figure 27 : 2nde dimension, séparation des protéines sur un gel SDS-PAGE en fonction de leur taille à t₀ (toutes les protéines sont négativement chargées), t₁ (les protéines migrent sous l'effet d'un champ électrique dans un gel réticulé) et t₂ (les protéines sont piégées en fonction de leur taille).

c) Détection des protéines par marquage

La sensibilité de détection de spots de protéines dépend essentiellement de la procédure de coloration utilisée. De nombreux types de colorants ont été développés et les plus couramment utilisés sont :

- **La coloration à l'argent** : le marquage se fait par affinité entre la protéine et le cation métallique. Celui-ci permet de détecter des quantités de 1-10 ng. Toutefois au-dessus d'une certaine concentration de protéines, cette coloration n'est plus linéaire. De plus, il a pu être montré qu'il peut y avoir des incompatibilités avec l'analyse par spectrométrie de masse [Gevaert and Vandekerckhove, 2000]. Toutefois, des études menées au laboratoire ont permis de mettre au point un protocole améliorant sa compatibilité avec la spectrométrie de masse [Richert et al., 2004].
- **Le bleu de coomassie** : le marquage se fait par des interactions non-covalentes entre le colorant et les groupements amino ou thiols des protéines [Neuhoff et al., 1988]. Même si celui-ci est beaucoup moins sensible (détection de quantités de l'ordre de 100 ng), ce marquage présente l'avantage d'être linéaire et peut donc être utilisé pour des études comparatives.
- **La fluorescence** : le principe de ce type de marquage, similaire à celui utilisant des colorants, consiste à lier un fluorophore à la protéine par des liaisons covalentes ou non. L'inconvénient de ce type de marquage dans la détection qui nécessite l'utilisation d'outils optiques très performants. Cependant, il est beaucoup plus sensible et permet de détecter des protéines jusqu'à l'attomole.

d) Les limitations

Toutefois, cette technique présente certaines limitations [Ong and Pandey, 2001], (Lopez, 2007) tels que:

- L'automatisation et le temps de réalisation des gels qui est assez long.
- La résolution des gels qui souvent ne permet pas de détecter, identifier et quantifier toutes les protéines présentes dans un échantillon [Gygi et al., 2000].
- La gamme dynamique de détection (limite de détection de 1ng): celle -ci est de 10^3 - 10^4 alors que dans une cellule on a des protéines minoritaire présentes à 100 copies par cellule alors que les plus abondantes sont présentes à 100 000 000 copies par cellule pour une gamme dynamique de 10^6 [Lopez, 2007].
- La visualisation des protéines membranaires ou hydrophobes [Santoni et al., 2000], [Braun et al., 2007].
- La visualisation des protéines de pI et de poids moléculaires extrêmes : classiquement, un gel de taille 18 cm x 18 cm, permet de séparer les protéines entre des pI 4-10 et des PM 10-100 kDa [Gorg et al., 2004]. De ce fait, il est difficile de détecter les protéines n'ayant pas ces caractéristiques.
- L'identification de modifications post-traductionnelles : en effet, il n'est pas rare d'identifier dans plusieurs spots la même protéine ceci étant lié à la présence de plusieurs formes de cette protéine. Malheureusement, lors de la digestion de ces protéines avec une enzyme de type trypsine, le nombre de peptides générés ne permet pas dans tous les cas de différencier ces formes.

Afin de pouvoir palier à ces limitations, des alternatives sont proposées [Lopez, 2007] et à ce jour, un gel 2D optimisé permet de détecter jusqu'à 5000 protéines avec une sensibilité allant jusqu'au nanogramme [Gorg et al., 2004], ce qui en fait une méthode de choix pour séparer les protéines avant analyse par spectrométrie de masse.

Toutefois, selon l'épaisseur du gel, sa taille, la qualité des réactifs utilisés, la qualité de la préparation d'échantillon et l'optimisation du système, la résolution d'un gel 2D peut considérablement varier. De ce fait des techniques alternatives existent.

III.1.2. Les alternatives

Comme nous avons pu le voir précédemment, la séparation sur gels d'électrophorèse bidimensionnelle est une des techniques les plus utilisée actuellement.

Or celle-ci peut rapidement être limitée notamment dans le cas d'échantillons insolubles dans les solvants et détergents utilisés (cas des protéines membranaires par exemple) lors de la préparation des échantillons. Toutefois il existe d'autres alternatives tels que les gels 1D et le OFFGEL.

a) L'électrophorèse monodimensionnelle (gel 1D)

Son principe est basé sur la séparation des protéines sur un gel SDS-PAGE uniquement en fonction de leur taille. Cette approche facile d'utilisation et rapide, est généralement utilisée pour fractionner des échantillons complexes avant leur analyse par spectrométrie de masse.

De manière générale, la séparation de protéines sur gels 1D est souvent suivie d'analyses sur des systèmes nanoLC-MS/MS. En effet, ne faisant intervenir qu'une dimension, les bandes résultantes peuvent souvent contenir plusieurs dizaines de protéines différentes mais également des isoformes d'une même protéine nécessitant de ce fait une seconde étape de séparation par chromatographie liquide [Miguet, 2006]. Ainsi de manière générale :

- (1) les protéines sont séparées en fonction de leur taille sur un gel 1D
- (2) les bandes de gels sont découpées
- (3) les protéines présentes dans chaque bande sont réduites, alkylés, digérées puis extraites
- (4) les extraits de peptides sont analysés par nanoLC-MS/MS.

b) L'électrophorèse Off-Gel

Plus récemment, un nouveau type de fractionnement fut développé, la focalisation isoélectrique Off-Gel. A la différence du fractionnement sur gel 1D, les protéines sont séparées en solution en fonction de leur point isoélectrique. Les extraits de protéines sont déposés dans des puits et sous l'effet d'un champ électrique les protéines vont migrer d'un puits à un autre via un strip de gel IPG [Michel et al., 2003b]. Lorsque les protéines ont atteint leur point isoélectrique, elles restent bloquées dans un puits (figure 28).

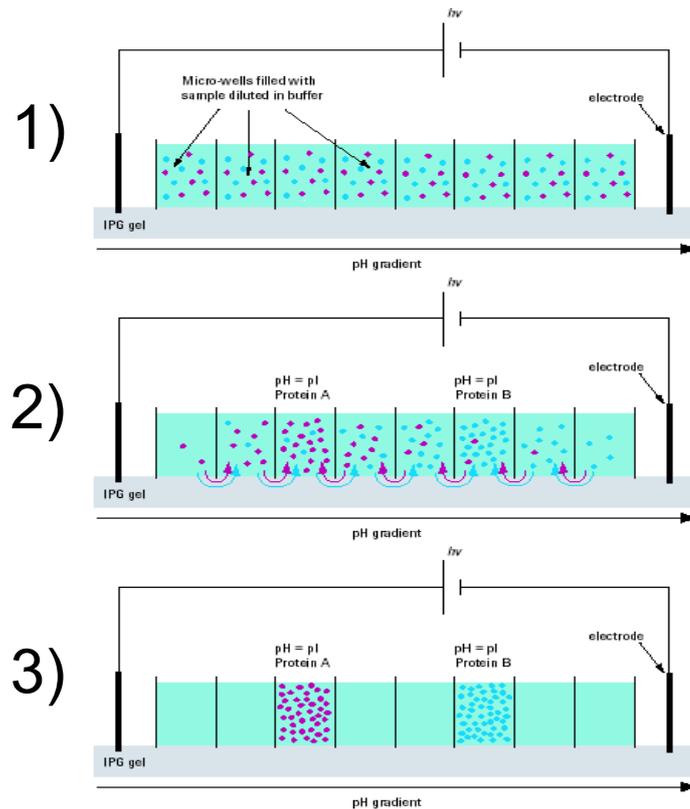


Figure 28 : Principe de la focalisation isoélectrique Off-Gel

(schéma du système 3100 OFFGEL d'Agilent Technologies) [Hörth et al., 2006].

- (1) les extraits sont déposés uniformément dans chaque puits, (2) sous l'effet d'un champ électrique les protéines vont migrer d'un puits à l'autre par gradient de pH via le strip IPG, (3) lorsque les protéines ont atteint leur point isoélectrique, elles restent piégées dans un puits.

Ce type de fractionnement présente donc l'avantage de pouvoir séparer les protéines en fonction de leur pI à 0,1 unités de pH près tout en récupérant les échantillons en solution [Michel et al., 2003a]. Ainsi il est ensuite éventuellement possible de refractionner les extraits par d'autres techniques (gel 1D par exemple). Enfin de récentes études ont montré son utilisation pour l'analyse de mélanges de peptides [Fraterman et al., 2007], [Horth et al., 2006].

III.2. Les techniques de séparation des peptides : la chromatographie liquide

Face aux limitations de l'électrophorèse bidimensionnelle (détection des protéines minoritaires, détection des protéines de masses trop faibles ou trop élevée, détection des protéines basiques ou hydrophobes/membranaires), il fut rapidement nécessaire de trouver d'autres techniques séparatives.

Ainsi, de nombreuses améliorations furent apportées aux systèmes de chromatographie liquide pour permettre de séparer des protéines et des peptides (type de systèmes, type de colonnes, phases de la colonne) mais également de gagner en résolution et en sensibilité (taille de particules, diamètres internes des colonnes, réduction des volumes morts dans les systèmes de chromatographie).

Actuellement deux types de techniques sont couramment utilisés, la nanoHPLC-MS/MS et la chromatographie bidimensionnelle.

III.2.1. Le système nanoHPLC-MS/MS

En général les systèmes de chromatographie liquide sont couplés à des spectromètres de masse à source ESI. La source ESI étant concentration dépendante et les échantillons biologiques étant souvent peu concentrés, il a fallu développer des systèmes de chromatographie liquide pouvant descendre à des débits de 100 à 300 nL/min. A ce jour la technique la plus utilisée est la nano chromatographie liquide haute performance (nanoHPLC) en phase inverse (RP) couplée à des spectromètres de masse.

Les échantillons étant souvent dilués au moment de l'injection, ce type de système se compose d'une colonne de chargement et d'une colonne analytique. La colonne de chargement permet de concentrer et de dessaler les échantillons avant leur séparation sur une colonne analytique par un gradient de solvants eau/acétonitrile. C'est en général, le choix de la phase utilisée (taille des particules, taille des pores, type de phase stationnaire) qui aura le plus d'impacte sur la sélectivité de séparation des protéines et des peptides. De manière générale, une colonne typiquement utilisée pour séparer les peptides est remplie avec des billes de silice sphériques modifiées à leur surface par des chaînes alkyles de type C18, dont la taille des particules va de 5 à 3 μm avec une porosité de 300 Å [Neverova and Van Eyk, 2005]. Classiquement, pour un système nanoHPLC-MS/MS, on utilise des colonnes de chargement et des colonnes analytiques de phase stationnaire C18 de 75 μm de diamètre interne avec des particules de 5 μm [Peng and Gygi, 2001]. La nanoHPLC-MS/MS est généralement associée à des étapes de fractionnement en amont de type électrophorèse mono- ou bidimensionnelle (figure 29).

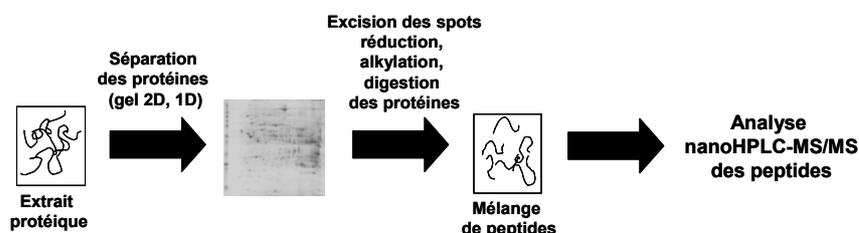


Figure 29 : Schéma de la stratégie séparation des protéines sur gel 2D suivie d'analyses nanoHPLC-MS/MS.

Cependant selon la complexité des échantillons, la détection peut rapidement être limitée. C'est pourquoi à ce jour, le déficit réside dans le développement de nouveaux systèmes de chromatographie liquide permettant de gagner en sensibilité, en résolution mais également en temps d'analyse. Ainsi Agilent Technologies développa récemment le système HPLC-Chip [Yin and Killeen, 2007], [Yin et al., 2005].

Enfin, plus récemment, le développement de colonnes avec des tailles de particules toujours plus petites permettant d'augmenter la résolution et la sensibilité furent développées et commencent à montrer leur utilité pour l'analyse de peptides [Liu et al., 2007]. Toutefois, ce type de colonne, permettant de travailler à des débits très élevés entraîne une augmentation de pression sur les systèmes

et afin de pouvoir palier à ce problèmes, des systèmes de chromatographie permettant de tenir des pressions très élevées furent développés (Ultra performances Liquid Chromatographie, UPLC™) [Kay et al., 2007].

III.2.2. La chromatographie multidimensionnelle

Afin de contourner les limites résolutive des approches basées sur la séparation des protéines sur gels d'électrophorèses suivies d'analyses MALDI-TOF-MS ou nanoHPLC-MS/MS sur phase inverse (identification des protéines peu abondantes, protéines ayant des pI et des poids moléculaires extrêmes, protéines membranaires), il fut rapidement nécessaire de trouver des alternatives pour l'analyse protéomique d'échantillons biologiques complexes.

Ainsi, en 1999, Yates et al proposent une nouvelle stratégie d'identification pour l'analyse directe d'échantillons complexes, la DALPC [Link et al., 1999], la chromatographie multidimensionnelle couplée à la spectrométrie de masse. Aujourd'hui on parle de stratégie MudPIT pour Multidimensional Protein Identification Technologies [Wolters et al., 2001]. Elle a ouvert la voie à la « protéomique shotgun » qui consiste à obtenir un profil global des protéines présentes dans un mélange par l'analyse directe de celui-ci [Wolters et al., 2001].

Son principe est basé sur la séparation de mélanges complexes de peptides selon leur charge et leur hydrophobicité. Les mélanges complexes de protéines sont directement digérés en peptides qui sont ensuite soumis à plusieurs étapes de chromatographie liquide. Actuellement il existe de nombreuses stratégies cependant, le plus couramment, les peptides acidifiés sont dans un premier temps fractionnés par passage sur une colonne échangeuse de cations (SCX), puis chacune de ces fractions est ensuite analysée par nanoHPLC-MS/MS en phase inverse (RP), c'est la SCX/RP/MS/MS. A ce jour, plusieurs approches ont été développées pour la réalisation de ce type d'analyses [Fournier et al., 2007] :

- **Approche off-line** : les peptides sont fractionnés dans un premier temps sur une colonne échangeuse de cations, récoltés, puis analysés par nanoHPLC-MS/MS sur colonne de phase inverse. Les deux systèmes n'étant pas couplés, cette approche est beaucoup plus tolérante aux sels et aux solvants utilisés et permet de ce fait d'optimiser chacune des deux étapes de chromatographie liquide. De plus, elle permet d'introduire un système de détection UV après la première étape de fractionnement permettant ainsi de choisir les fractions à collecter. Cependant, cette approche peut entraîner des pertes d'échantillons lors des étapes de préparation des échantillons entre les deux chromatographies.
- **Approche on-line** : les deux colonnes sont couplées via un système de valve contenant une colonne de chargement de phase inverse C18. Le mélange de peptides est injecté sur la colonne SCX. En augmentant la concentration de sels, les peptides vont être élués de la colonne SCX vers la colonne de chargement C18 qui permet de dessaler les peptides. Enfin, les peptides sont élués de la colonne de chargement vers la colonne analytique C18 où ils seront séparés en utilisant un gradient de solvants organiques (eau/acétonitrile). Cette approche est plus rapide et entraîne moins de pertes en échantillons que les systèmes off-line. Toutefois, la mise en place de ce type d'approche peut rapidement s'avérer complexe puisqu'il

nécessite l'utilisation de systèmes de valves dont la configuration est compliquée à mettre en place (figure 30). De plus, les peptides sont élués par paliers de sels et non par gradient, la séparation est donc moins efficace. Enfin, l'absence de système de détection UV entre les deux étapes oblige l'utilisateur à travailler en aveugle quant à la quantité de peptides présents dans chaque fraction.

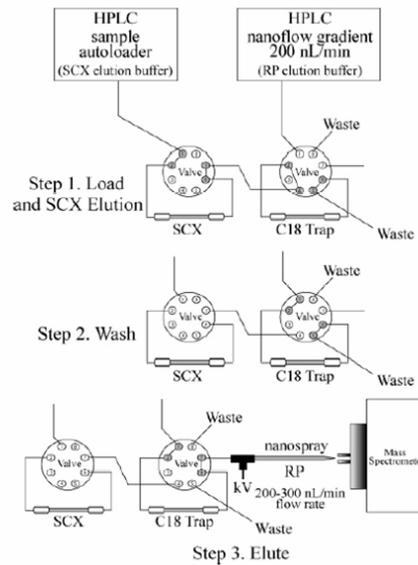


Figure 30 : Exemple de la configuration on-line [Fournier et al., 2007]

- **Approche MudPIT** : dans cette approche, on utilise une colonne microcapillaire de silice remplie en série avec de la phase SCX et de la phase inverse qui est directement connectée au spectromètre de masse et joue le rôle de la source ESI (figure 31). Les peptides sont chargés sur cette colonne bi-phasique et élués de la résine SCX par paliers de sels vers la résine RP pour être concentrés en tête de celle-ci. Entre chaque palier de sels, on applique un gradient d'acétonitrile qui va permettre d'éluier les peptides de la résine RP vers le spectromètre de masse [Link et al., 1999].

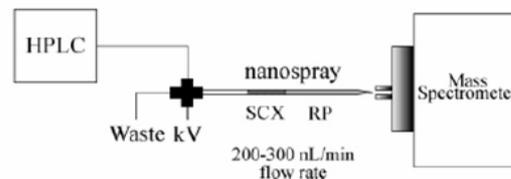


Figure 31 : Exemple du système MudPIT [Fournier et al., 2007].

L'inconvénient majeur de ce type de stratégies réside dans l'interprétation des données de masse générées par l'analyse directe de mélanges complexes de peptides. En effet, les extraits totaux étant directement digérés, le nombre de peptides générés est considérable et bon nombre d'identifications de protéines ne se font qu'avec un peptide. De même, il peut s'avérer difficile d'attribuer avec certitude

un peptide à une protéine notamment dans le cas d'identifications de protéines ayant de fortes homologies de séquences [Nesvizhskii and Aebersold, 2005]. De ce fait la validation des interprétations et des identifications de protéines doivent se faire avec la plus grande prudence en vérifiant la qualité des spectres de fragmentation.

Cependant, ce type de stratégie a montré tout son intérêt pour l'étude d'échantillons très complexes avec l'identification de 1484 protéines contre 300 par une approche de séparation sur gels 2D [Wolters et al., 2001], mais également dans le cas de l'étude de protéines membranaires et la caractérisation de modifications post-traductionnelles [Wu et al., 2003], [Wu and Yates, 2003].

III.3 Conclusion

Nous avons pu voir dans ce chapitre qu'il existe en amont des stratégies d'identification par spectrométrie de masse de nombreuses techniques séparatives dédiées à la séparation de protéines et de peptides. Les plus couramment utilisées sont l'électrophorèse bidimensionnelle (gel 2D) et les systèmes nanoHPLC-MS/MS qui sont souvent utilisés en association.

Néanmoins face à la complexité des échantillons biologiques et les inconvénients de ces techniques, d'autres approches ont été développées telles que la chromatographie multidimensionnelle ou encore des alternatives comme l'approche séparation sur gels 1D ou Off-Gel (figure 32).

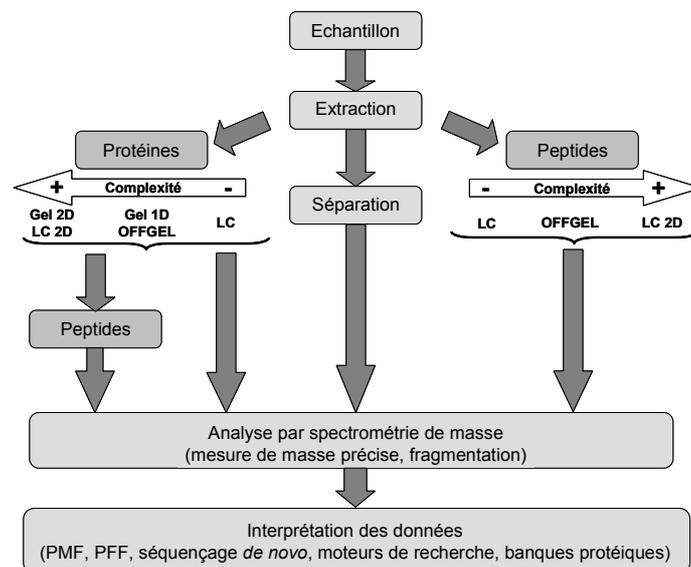


Figure 32 : Schéma de l'ensemble des stratégies utilisées en analyse protéomique pour l'identification de protéines.

Chacune de ces techniques présente des avantages et des inconvénients et le choix d'utiliser l'une d'elle est souvent dicté par la nature des échantillons biologiques à disposition mais également selon le problème posé [Issaq, 2001], [Issaq, 2003], [Issaq et al., 2005]. C'est pourquoi il n'est pas exclu de combiner plusieurs de ces stratégies selon que l'on souhaite séparer, identifier, quantifier ou mettre en évidence la présence de modifications post-traductionnelles ou de mutations.

CHAPITRE 2

LES NOUVEAUX BESOINS : LA QUANTIFICATION ET L'ETUDE DE MODIFICATIONS POST- TRADUCTIONNELLES.

Depuis quelques années, la simple identification de protéines dans un extrait cellulaire n'est plus suffisant, et aujourd'hui les nouveaux besoins résident dans une caractérisation complète des protéines présentes dans un extrait mais également dans la quantification des protéines.

Ainsi, ce chapitre sera consacré dans un premier temps à faire un rapide point sur les méthodes couramment utilisées pour quantifier les protéines et dans un second temps nous verrons comment utiliser la spectrométrie de masse pour la caractérisation fine de N-glycoprotéines.

I. La quantification

La quantification de protéines dans un extrait biologique n'est pas chose aisée en protéomique. Une des premières stratégies ayant été utilisée pour montrer des différences dans le niveau d'expression des protéines fut l'approche gels d'électrophorèse 2D.

Celle-ci consiste à **comparer les profils de gels 2D des protéines** correspondants à différents états (comparaison de différents type de souches, de milieux de culture en absence ou en présence d'un récepteur, d'extraits de patients sains ou malades, ...). Ainsi il est possible de déterminer la sur ou sous expression de certaines protéines ou la présence ou l'absence de certaines protéines (figure 1).

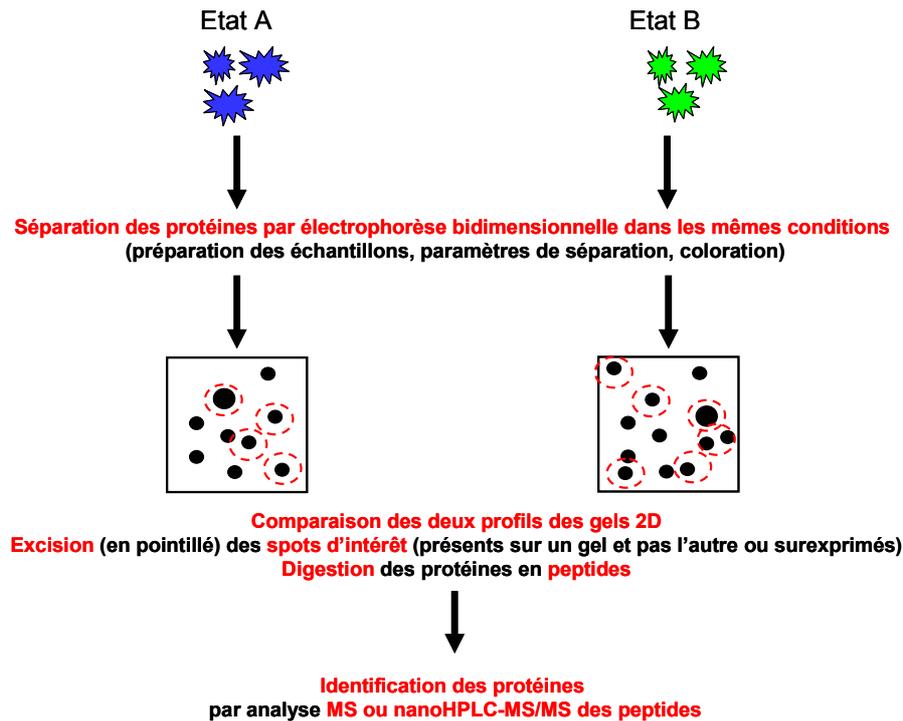


Figure 1 : Principe de la protéomique comparative et quantitative par approche gels d'électrophorèse.

Néanmoins, cette technique étant peu sensible et peu reproductible, les études de quantification sont souvent imprécises. De ce fait, cette approche est généralement utilisée pour des **études comparatives globales**.

Ainsi, cela entraîna le développement de nouvelles techniques pour la quantification basée sur le **marquage des protéines et leur quantification par protéomique shotgun**. En 1999, l'équipe d'Aebersold introduisit pour la première fois une méthode de marquage, la dérivatisation sélective des cystéines par marquage avec des isotopes stables, l'**Isotope-Coded Affinity Tag (ICAT)** [Gygi et al., 1999]. Celle-ci consiste à marquer les résidus de cystéines avec un réactif marqué au deutérium.

La même année, Oda et al, établit un profil quantitatif des protéines de deux états cellulaires différents dans une levure en utilisant le marquage par des **isotopes stables d'azote ^{15}N** [Oda et al., 1999]. Quelques années plus tard en 2001, Conrads et al décrivent une procédure de quantification basée sur la combinaison des deux premières [Conrads et al., 2001]. Depuis, de nombreuses méthodes de quantification alternatives sont proposées tels que (figure 2) :

- Le **marquage à l' ^{18}O des fonction C-terminale des peptides** lors de la digestion protéolytique à la trypsine [Mirgorodskaya et al., 2000], [Fenselau and Yao, 2007], [Heller et al., 2003].
- Le **marquage isobare pour la quantification relative et absolue, iTRAQ**. Celui-ci consiste à marquer les N-terminaux et les chaînes latérales des lysines de tous les peptides par dérivatisation chimique à l'aide de 4 réactifs isobares. La quantification ne se fera pas en MS, puisqu'ils ont tous la même masse mais à l'aide d'ions diagnostiques obtenus après fragmentation [Ross et al., 2004], [Roe and Griffin, 2006].

- Le **marquage métabolique des protéines par isotopes stable dans les milieux de cultures SILAC**. Celui-ci consiste à incorporer des acides aminés marqués avec des isotopes stables in vivo durant la synthèse cellulaire des protéines [Ong et al., 2002].

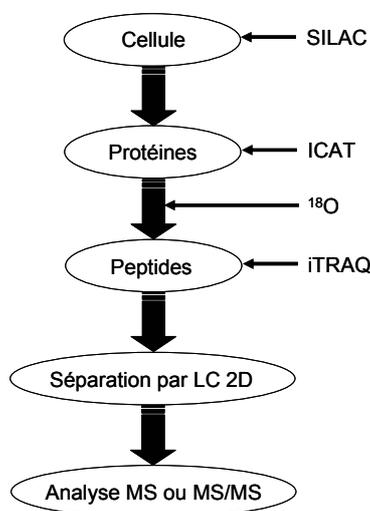


Figure 2 : Les stratégies de marquage les plus couramment utilisées basées sur l'utilisation de la spectrométrie de masse pour quantifier les protéines et les peptides [Roe and Griffin, 2006].

II. L'étude de modifications post-traductionnelles (PTMs) : cas de la glycosylation

II.1. Introduction

Une modification post-traductionnelle d'une protéine se définit comme étant une modification chimique covalente (addition ou dégradation) catalysée par des enzymes et qui se produit après que l'ADN ait été transcrit en ARN elle-même traduite en protéines. Cette réaction est un mécanisme biologique couramment observé et qui est directement impliqué dans le contrôle de l'activité des protéines. Les PTMs peuvent moduler, par exemple, les interactions moléculaires, la localisation des protéines et leur stabilité [Jensen, 2004].

A ce jour, on compte plus de 200 PTMs qui se divisent en deux catégories :

- L'addition covalente de groupements chimiques catalysée par des enzymes.
- La dégradation covalente du squelette peptidique des protéines par l'action de protéases ou par autocatalyse.

L'ensemble de ces PTMs sont très bien décrites par Walsh TC et al. [Walsh et al., 2005] et ne seront pas vus en détails ici.

Cependant, les plus répandues correspondent à des additions covalentes parmi lesquels on trouve la phosphorylation, la glycosylation, l'ubiquitination, la méthylation et l'acétylation [Jensen, 2006]. Chacune de ces PTMs aux fonctions biologiques variables va se produire sur des acides aminés

donnés et va entraîner des différences de masses par rapport à l'acide aminé non-modifié. Ces différences de masses sont mesurables par spectrométrie de masse [Jensen, 2006] (tableau 1).

Modification post-traductionnelle (acides aminés modifiés)	Différence de masse (Da)	Exemples de leur fonctions biologiques
Phosphorylation (Ser, Thr, Tyr)	80	Transduction du signal, régulation de l'activité enzymatique, interactions protéines-protéines ou protéines-ligands
Sulfatation (Tyr)	80	Signalisation et localisation des protéines, interactions protéines-protéines
Glycosylation (Asn, Ser, Thr)	203 à > 1000	Stabilité, solubilité, sécrétion du signal, régulation des interactions, reconnaissance et interaction extracellulaire
Ubiquitination (Lys)	> 1000	Signal pour dégradation des protéines, interactions protéines-protéines
Méthylation (Lys mono-, di- et triméthylation) (Arg mono- et diméthylation) (His monométhylation)	14, 28, 42 14, 28 14	Régulation de l'activité des protéines, interaction protéines-protéines et protéines-acides nucléiques, activité des gènes (histones)
Acétylation (Lys N-terminal)	42	Stabilité et activité des protéines, régule interactions protéines-protéines et protéines-ligands
Oxydation (Met) (Trp)	16 4, 16, 32	Peut réguler l'activité des protéines mais il s'agit souvent d'un artefact chimique
Ponts disulfures (Cys)	- 2	Stabilise l'activité et la structure des protéines, impliqué dans des processus rédox

Tableau 1 : PTMs les plus couramment observées, différence de masse générées et exemple de leur effet sur les protéines [Jensen, 2006].

Cependant, comme illustré dans le tableau 1, plusieurs acides aminés peuvent subir la même modification (cas par exemple de la phosphorylation).

De même, des modifications de nature différentes peuvent engendrer des différences de masses très proches (par exemple la phosphorylation et la sulfatation ou la triméthylation et l'acétylation). Cela peut rendre l'identification et la localisation de la modification difficile et leur étude par analyse protéomique peut rapidement s'avérer compliquée. De plus, il est courant de constater que les protéines sont partiellement modifiées, ainsi les protéines et les peptides portant des PTMs sont souvent présents en faible quantité dans les extraits rendant leur détection difficile.

C'est pourquoi, de nombreuses méthodologies ont été développées qui incluent les optimisations instrumentales, mais également des méthodes d'enrichissement, de marquages chimiques spécifiques ou des combinaisons d'enzymes de digestions notamment pour l'étude de phosphorylations [Mann and Jensen, 2003], [Siuti and Kelleher, 2007], [Witze et al., 2007], [Jensen, 2004].

Au cours de ma thèse, ayant eu l'occasion d'étudier l'hétérogénéité de protéines N-glycosylées, l'utilisation de la spectrométrie de masse pour la caractérisation fine de ce type de PTMs sera décrite en détail dans le chapitre suivant.

II.2. Cas de la N-glycosylation

II.2.1. Introduction

La glycosylation est une des PTMs les plus répandues dans les protéines. On estime actuellement dans la banque Swiss-Prot que plus de 50 % des protéines sont glycosylées [Apweiler et al., 1999].

Pendant de nombreuses années, la glycosylation n'a été perçue par les biologistes comme n'ayant qu'une fonction de solubilisation des protéines. Or, selon sa nature, sa structure fine et sa microhétérogénéité, l'activité et la fonction des protéines seront modifiées pour pouvoir intervenir dans divers processus biologiques.

Parmi ceux-ci, les glycoprotéines sont connues pour être impliquées dans de nombreux phénomènes tels que les réponses immunitaires, le développement cellulaire, les processus de signalisation inter- et intracellulaires qui entrent en jeu dans des processus tels que le cancer, les virus, les infections bactériennes, parasitaires et bien d'autres [Helenius and Aebi, 2001], [Rudd et al., 2001], [Varki, 1993], [Morris et al., 2007], [Lee et al., 2005].

De ce fait, leur caractérisation fine est rapidement devenue une nécessité. Ainsi, en 1978 Morris et al. caractérisent pour la première fois par spectrométrie de masse une glycoprotéine extraite du sang d'un poisson vivant dans l'antarctique [Morris et al., 1978].

Depuis, et suite aux nombreux développements instrumentaux, l'analyse protéomique basée sur l'utilisation de la spectrométrie de masse est devenue un outil de choix pour la caractérisation fine de glycoprotéines.

II.2.2. Nature et structures des glycosylations

La glycosylation consiste en la liaison covalente d'un ou plusieurs glycanes sur un ou plusieurs acides aminés de la chaîne latérale de la protéine. Ces glycanes correspondent à l'association en chaîne de plusieurs résidus monosaccharides en polysaccharides. L'ensemble des résidus les plus observés sont décrit dans le tableau 2.

Monosaccharides	Abréviations	Symboles	Masses moyennes des résidus (Da)
Xylose	Xyl	☆	132,12
Fucose	Fuc	△	146,14
Galactose	Gal	○	162,14
Glucose	Glc	●	162,14
Mannose	Man	◐	162,14
N-acétylglucosamine	GlcNAc	■	203,19
N-acétylgalactosamine	GalNAc	□	203,19
Acide N-acétylneuraminique	NeuAc	◆	291,26

Tableau 2 : Noms, abréviations, symboles et masses des monosaccharides les plus couramment observés dans les structures des glycanes. La nomenclature de symboles a été établie par le « Consortium for functional Glycomics » (<http://www.functionalglycomics.org>).

Il existe deux types de glycosylation qui dépendent de la nature de la liaison covalente : les O-glycosylations et les N-glycosylations.

- **Les O-glycosylations** : le résidu monosaccharide se lie à la protéine par le groupement hydroxyl des résidus sérine Ser, thréonine Thr, hydroxyproline Hyp ou hydroxylysine. La plus répandue est la O-glycosylation de type mucine où le résidu N-acétylglucosamine (GlcNAc) du glycane se lie à la sérine ou la thréonine.
- **Les N-glycosylations** : le glycane est attaché à la protéine par le groupement amide d'une asparagine Asn via le résidu N-acétylglucosamine (GlcNAc). L'asparagine doit se situer dans un motif Asn-X-Ser/Thr (parfois Asn-X-Cys) (X pouvant être n'importe quel acide aminé sauf une proline Pro). Leur chemin de biosynthèse étant commun, tous les N-glycane ont un squelette pentasaccharide identique qui est constitué de 3 mannoses et de 2 N-acétylglucosamines ($\text{Man}_3\text{GlcNAc}_2$) [Gomord and Faye, 2004].

A ce jour il existe de nombreux types de N-glycosylations dont la structure varie considérablement selon qu'elles soient biosynthétisées dans les cellules des mammifères ou des plantes [Gomord and Faye, 2004] (figure 3).

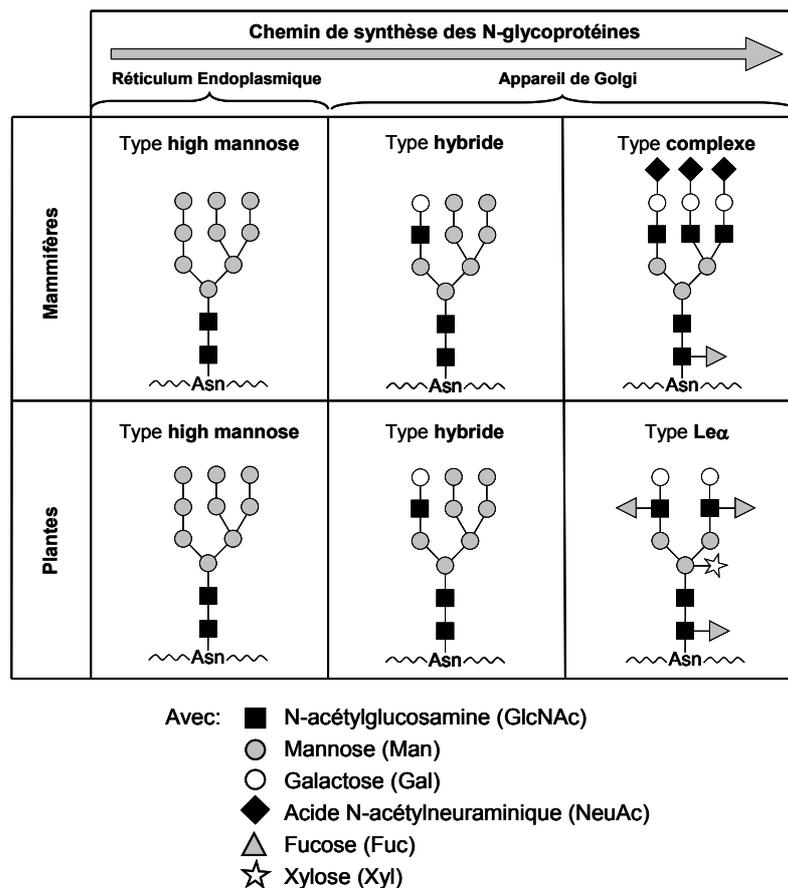


Figure 3 : Exemple de structures de N-glycane couramment trouvés dans les plantes et les mammifères aux différents stades de leur biosynthèse [Gomord and Faye, 2004].

II.2.3. Stratégies pour la caractérisation des N-glycoprotéines

La caractérisation de N-glycoprotéines consiste à déterminer le ou les sites de N-glycosylation et la structure du ou des glycanes sur un ou plusieurs sites. Cependant, leur caractérisation complète est complexe car [An et al., 2003], [Budnik et al., 2006]:

- une protéine glycosylée peut porter **plusieurs sites de N-glycosylation** occupés ou non
- sur un **même site**, la **nature du glycanes** peut **varier**
- les glycanes ont des **structures complexes** et très **diversifiées** (combinaisons des monosaccharides, isomérisation de position des liaisons glycosidiques, l'antennarité, ...)
- les **structures des glycanes** sur une protéine ne sont **pas statiques** et **varient** de manière **qualitative** et **quantitative** en fonction des **processus biologiques**.

De ce fait, les **glycoprotéines** présentent souvent une **hétérogénéité globale** élevée, due à l'**hétérogénéité sur chaque site de glycosylation (micro-hétérogénéité)**, qui les rend en générale difficilement détectables et analysables par des approches protéomiques classiques [An et al., 2003], [Geyer and Geyer, 2006]. De plus, dans un extrait biologique donné, les glycoprotéines et les glycopeptides sont souvent **peu abondants par rapport aux autres protéines et aux peptides** [Budnik et al., 2006].

Afin de contourner ces limitations, de nombreuses améliorations ont été apportées pour la mise au point de méthodes **d'enrichissement des glycoprotéines et des glycopeptides**. La plus utilisée est la **chromatographie d'affinité sur colonnes de lectines**. Les lectines sont des protéines qui vont interagir avec des motifs spécifiques de glycanes sans les modifier. Selon la nature de la lectine, les N-glycoprotéines mais également les N-glycopeptides piégés ne seront pas les mêmes. Un exemple de lectine couramment utilisée est la **concanavalin A (ConA)** qui permet de piéger des protéines et des peptides contenant des glycanes de type high-mannose et de type complexe [Budnik et al., 2006], [Ohyama et al., 1985]. Une autre alternative pour enrichir les extraits en **glycopeptides** est l'**exclusion stérique** et la chromatographie **d'interaction hydrophile (HILIC)**.

De même, il existe de nombreuses stratégies de **fractionnement/purification** pour s'affranchir des problèmes dus à leur hétérogénéité et pouvoir ainsi détecter les différentes glycoformes au niveau protéique, peptidique et sucres libres [Geyer and Geyer, 2006], [Mechref and Novotny, 2002], [Wuhrer et al., 2007], [Wuhrer et al., 2005], [Ruhaak et al., 2008].

Parmi les méthodes de purification les plus utilisées on compte :

- Les gels d'électrophorèse 1D et 2D, la chromatographie liquide échange d'ions (IEX) ou sur phase inverse (RP) et l'électrophorèse capillaire (CE) pour la **séparation des glycoprotéines**.
- La chromatographie liquide sur phase inverse (RP) ou sur phase normale (NP) et l'électrophorèse capillaire (CE) pour **séparer les glycopeptides**.
- La chromatographie liquide haute performance par échangeuse d'anions à pH élevé (HPAEC), sur phase normale (NP), sur phase inverse (RP) et sur colonnes en carbone graphite (GCC) mais également par chromatographie d'affinité sur colonne de lectines et l'électrophorèse capillaire (CE) pour **séparer les glycanes**.

L'ensemble des stratégies d'enrichissement et de purification par fractionnement sont illustrées dans la figure 4.

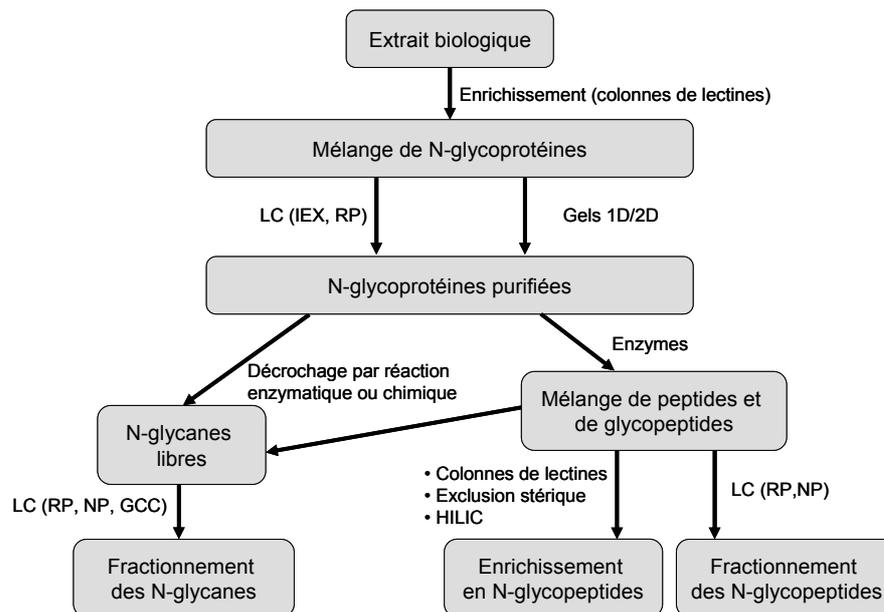


Figure 4 : Schéma des stratégies d'enrichissement et de purification par fractionnement utilisées pour l'analyse de N-glycoprotéines en protéomique.

Pour aller vers une caractérisation fine d'une N-glycoprotéine il est nécessaire de déterminer le ou les sites de N-glycosylations, la ou les structures des N-glycanes sur ce ou ces sites (microhétérogénéité) mais également d'étudier l'hétérogénéité globale des différents glycoformes d'une protéine. Pour cela il faut pouvoir, aux différents stades décrits dans la figure 4, analyser par spectrométrie de masse :

- Les **N-glycoprotéines entières** après enrichissement et purification à partir d'extraits biologiques.
- Les **N-glycopeptides** issus de la digestion des N-glycoprotéines purifiées et qui ont eux-mêmes été enrichis et fractionnés.
- Les **N-glycanes libres** issus de la déglycosylation des N-glycoprotéines ou des N-glycopeptides.

II.2.4. Apport de l'analyse MS pour déterminer le site de glycosylation et étudier l'hétérogénéité globale

a) Analyse MS des N-glycoprotéines entières

L'analyse directe des glycoprotéines peut permettre **d'étudier l'hétérogénéité globale des glycoformes**. En effet, pour une **protéine de masse connue**, il sera possible **d'estimer le nombre de glycoformes** et par différence de masse entre chaque massif il sera possible de remonter à la **nature**

des glycanes. Néanmoins, les glycoprotéines sont extrêmement hétérogènes et leur efficacité d'ionisation est bien plus faible que pour une protéine non glycosylée. De ce fait l'analyse des glycoprotéines entières nécessite de disposer d'échantillons fortement enrichis.

Pour les étudier, deux techniques sont couramment utilisées, l'analyse MALDI-TOF-MS et l'ESI-MS. Chacune d'elle présente des avantages et des inconvénients.

- **L'analyse par MALDI-TOF-MS** : elle permet généralement d'analyser des glycoprotéines n'excédant pas un poids moléculaire de 40 000 Da et contenant un nombre limité de glycanes comme pour la ribonucléase B [Tsarbopoulos et al., 1997].
- **L'analyse par ESI-MS** : avec le développement de nouvelles sources plus sensibles et les nombreuses améliorations instrumentales, l'analyse directe d'une glycoprotéine par ESI-MS est devenue une méthode de choix. La source ESI générant des ions multichargés, il est possible d'analyser des glycoprotéines de PM plus élevés et très hétérogènes. Néanmoins, l'interprétation de ces spectres très chargés n'est pas aisée puisque nous aurons dans un même spectre un mélange d'ions multichargés de toutes les glycoformes. Cependant, la source ESI pouvant être couplée à des systèmes LC, il est possible de séparer les glycoformes en amont avant leur injection sur le spectromètre de masse, évitant ainsi de multiples étapes de fractionnement qui entraînent des pertes d'échantillons.

b) Analyse MS des N-glycopeptides

Lorsque les glycoprotéines ont des PM trop élevés et / ou portent plusieurs sites de glycosylations, leur analyse directe pour étudier l'hétérogénéité globale des glycoformes peut s'avérer fastidieuse. Une alternative réside alors dans l'analyse MS des N-glycopeptides. En effet, en digérant la protéine il sera possible **d'obtenir des peptides ne portant qu'un site de glycosylation.**

Ainsi il sera possible **d'estimer le nombre exact de glycoformes** dans le cas d'une **protéine ne contenant qu'un site de glycosylation** mais également d'étudier **la microhétérogénéité sur chacun des sites** dans le cas **d'une protéine portant plusieurs sites** de glycosylation.

Néanmoins, pour ce type d'approche il est nécessaire de connaître la protéine potentiellement glycosylée pour pouvoir **choisir les enzymes de digestions** utilisées afin de générer des **glycopeptides de taille détectable dans les gammes de m/z** des analyseurs utilisés.

Généralement, étant en présence de **mélanges complexes de peptides et de glycopeptides**, l'analyse MS des glycopeptides est associée à une **étape de chromatographie liquide en amont.**

Dans le cas d'une protéine de séquence connue contenant un site de glycosylation il est facile de retrouver dans une analyse LC-MS quel ion correspond à un glycopeptide. En revanche, dans le cas d'une protéine contenant plusieurs sites de glycosylations, **l'interprétation des analyses LC-MS** peut rapidement s'avérer **fastidieuse.** De ce fait des études ont été menées pour pouvoir **rapidement détecter sélectivement les N-glycopeptides.**

L'approche la plus couramment utilisée consiste à induire une **fragmentation en source des N-glycopeptides** et a été montrée pour la première fois par Carr et al. [Carr et al., 1993]. En effet, en optimisant les réglages du spectromètre de masse, il est possible d'induire une fragmentation des N-

glycopeptides. La liaison glycosidique étant plus labile que la liaison peptidique, cette fragmentation va permettre de générer des **ions oxoniums** qui correspondent à la **perte de sucres**. Ces **ions diagnostiques** seront détectés dans les spectres MS contenant les glycopeptides. Les plus couramment observés sont les ions de m/z 204 (HexNAc)⁺, 366 (HexHexNAc)⁺, 292 (NeuAc)⁺ (figure 5).

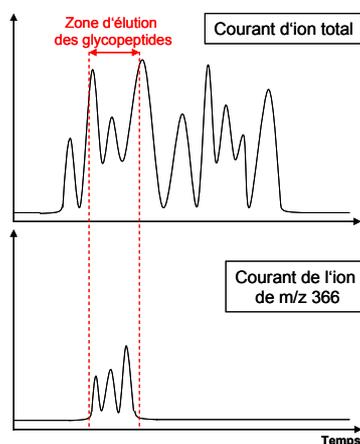


Figure 5 : LC-MS d'un mélange complexe de peptides et de glycopeptides avec une fragmentation induite en source pour déterminer la zone d'éluion des glycopeptides et étudier leur hétérogénéité.

Aujourd'hui, ce type d'ions peut être généré pratiquement par tous les spectromètres de masse à source ESI utilisés pour l'analyse protéomique [Jiang et al., 2004]. Ces ions peuvent également être produits dans la cellule de collision d'un Q-TOF pendant une LC-MS [Jebanathirajah et al., 2003], [Bateman et al., 2002]. De même, malgré les limitations des trappes ioniques pour détecter les ions dans les bas m/z , ces ions peuvent être détectés en modifiant plusieurs paramètres tels que la tension et la température du capillaire ou encore le choix des gammes de m/z scannées [Sullivan et al., 2004].

Une fois les glycopeptides détectés il est alors possible de **déterminer quels glycanes** sont présents **sur quels sites** et en **quelles proportions**. L'analyse des glycopeptides permettra donc de **déterminer le nombre de sites glycosylés** mais également de déterminer le **nombre de glycanes sur chaque site** et donc **d'étudier l'hétérogénéité globale de la glycoprotéine** dans la mesure où la nature exacte des glycanes est connue au préalable (clé « Analyse MSⁿ des N-glycanes », p. 64).

c) Analyse MS des protéines et des peptides déglycosylés

L'analyse MS des protéines et des peptides déglycosylés permet de **déterminer le nombre de sites glycosylés** sur une protéine ou **d'identifier rapidement des N-glycopeptides dans un chromatogramme obtenu par LC-MS**. A ce jour il existe de nombreuses enzymes dédiées à cet effet et chacune d'elle coupe à des endroits spécifiques du glycanes.

En effet, certaines enzymes de type N-glycosidases comme la **PNGase F** et la **PNGase A**, clivent la liaison entre le GlcNAc et l'asparagine pour libérer le **glycane intacte**. Dans la partie peptidique, le résidu asparagine Asn est converti en résidu acide aspartique Asp entraînant ainsi une **différence de**

masse de 1 Da détectable par spectrométrie de masse. Ce type d'enzyme est généralement utilisée dans l'objectif de caractériser la structure des glycanes puisqu'elle laisse ceux-ci intacts (figure 6a).

Les enzymes de type **endoglycosidase** comme **l'Endo H** ou **l'Endo D** vont agir entre les deux GlcNAc du squelette commun. Ces enzymes ne laissent donc **pas le glycan intacte** mais par contre l'Asn contiendra toujours un résidu GlcNAc qui entraîne une **différence de masse de 203 Da** plus facilement détectable par spectrométrie de masse qu'une différence de 1 Da (figure 6b).

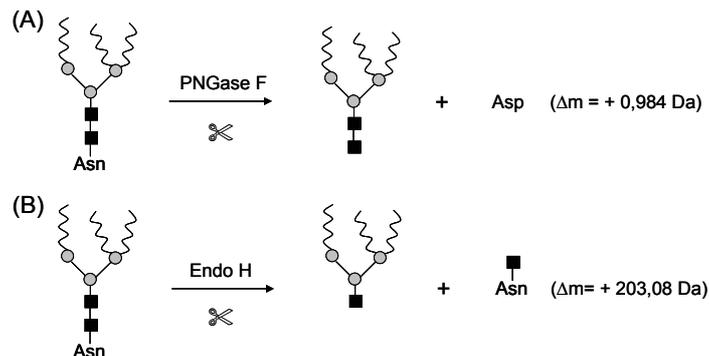


Figure 6 : Schéma des digestions enzymatiques des protéines et des peptides glycosylés permettant de décrocher les N-glycanes avec (A) la PNGase F et (B) l'Endo H.

Par exemple, dans le cas de l'analyse MALDI-TOF-MS d'une glycoprotéine, en effectuant un **comparatif des spectres MS mesurés avant et après** déglycosylation, il est possible d'estimer le **nombre de sites glycosylés** et dans le cas d'un seul site, estimer le nombre de glycoformes et la nature des glycoformes par une analyse plus poussée des N-glycanes libres (figure 7).

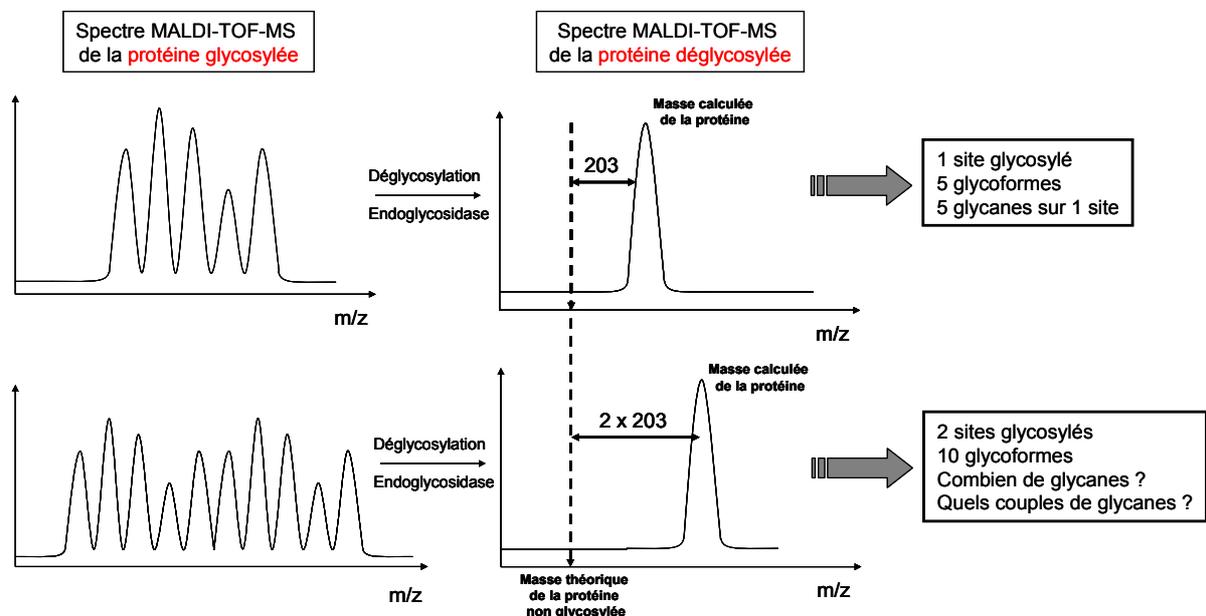


Figure 7 : Schéma de la stratégie utilisant la déglycosylation de la protéine pour déterminer le nombre de sites glycosylés et éventuellement le nombre de glycoformes à partir d'un spectre MALDI-TOF-MS d'une glycoprotéine.

II.2.5. Apport de la MSⁿ pour la caractérisation structurale

L'analyse par spectrométrie de masse en tandem permet de déterminer l'enchaînement des monosaccharides mais également de donner des informations sur l'isomérisation de position des liaisons glycosidiques conduisant à l'élucidation de la **structure fine des glycanes**.

a) Nomenclature de fragmentation des polysaccharides

La fragmentation des glycanes se produit au niveau des liaisons interglycosidiques (Bi, Ci, Yi, Zi) ou intracycliques (Ai, Xi) [Domon, 1988]. Les fragments Xi, Yi et Zi contiennent l'extrémité réductrice, les fragments Ai, Bi et Ci contiennent l'extrémité non-réductrice. Les fragments de type Bi, Ci, Yi et Zi permettent de déterminer l'enchaînement des monosaccharides mais ne permettent pas de différencier un mannose d'un glucose par exemple. Par contre, les fragments de type Ai et Bi permettent de différencier les isomères (figure 8).

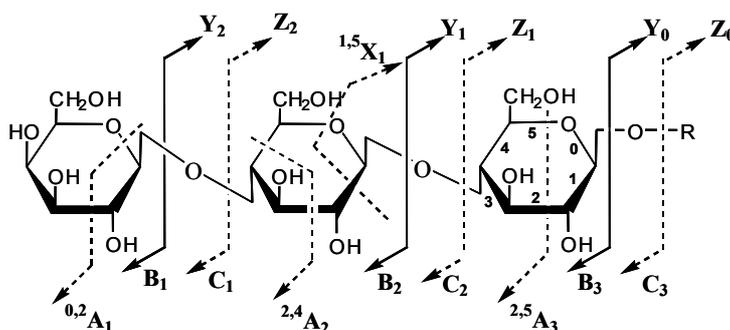


Figure 8 : Nomenclature de fragmentation des polysaccharides (Domon et Costello, 1988). Pour les fragments de type Ai et Xi, les numéros en exposants correspondent aux liaisons intracycliques cassées.

Selon que la fragmentation se produise à haute ou basse énergie, les fragments observés vont varier. A **haute énergie**, sur des spectromètres de masse de type secteurs magnétiques ou TOF-TOF, l'énergie fournie est suffisamment intense pour que **tous les chemins de fragmentations** soient empruntés sans aucune sélectivité. Il sera donc possible de déterminer l'enchaînement des monosaccharides mais également de différencier les isomères tels que le mannose et le glucose par exemple. A **basse énergie**, sur des instruments de type Q-Q-Q ou trappe ionique, l'énergie fournie étant plus faible, on observera préférentiellement les fragments thermodynamiquement stables (**Bi, Ci, Yi, Zi**). Ces instruments ne permettront pas de différencier les isomères aussi facilement.

b) Analyse MSⁿ des N-glycanes

L'analyse MSⁿ des N-glycanes est l'étape la plus importante pour la caractérisation de glycoprotéines. En effet, dans le cas de l'analyse de glycoprotéine inconnue, il est primordial de connaître la **structure fine des polysaccharides** pour pouvoir ensuite estimer le nombre de

glycoformes et étudier leur hétérogénéité. Néanmoins, plusieurs paramètres sont à prendre en compte lors de leur caractérisation structurale fine.

En effet, une première étape importante réside dans le **choix de la méthode enzymatique ou chimique utilisée pour décrocher les glycanes** de la protéine glycosylée. Chacune de ces méthodes de décrochage ayant des avantages et des inconvénients (spécifiques ou non d'un type de glycanes, glycanes laissés intacts ou non), il n'est pas peu courant de combiner plusieurs de ces enzymes pour décrocher l'ensemble des glycanes d'une protéine glycosylée [Harvey, 2001].

De manière générale les **fragments majoritairement observés correspondent à la coupure de la liaison interglycosidique (Bi, Ci, Yi, Zi)**. Ces fragments renseignent sur l'enchaînement des sucres dans le glycanes. Néanmoins dans certains cas, il peut être **difficile de déterminer si ces ions contiennent la partie réductrice du glycanes ou non**. Toutefois il existe à ce jour de nombreux réactifs permettant de **dérivée chimiquement la partie réductrice** du sucre et qui entraînent un incrément de masse qui permettra de localiser l'extrémité réductrice. Parmi les plus utilisés on compte par exemple le 2-aminobenzamide (2-AB) [Harvey, 2001], [Morelle et al., 2005]. De plus ces agents favorisent la formation et l'ionisation des ions fragments contenant la partie réductrice (Yi et Zi).

Néanmoins, ce type de fragments ne permet pas de différencier les isoméries de positions des liaisons interglycosidiques qui sont primordiales pour une caractérisation complète.

De ce fait, afin de permettre de **déterminer les isoméries de position**, de nombreuses stratégies ont été développées pour **favoriser la formation des ions fragments de type Ai et Xi**. Elles utilisent des **additions de métaux** tels que le lithium ou **l'addition d'alcalino-terreux** [Zaia, 2004]. De même, en **perméthylant les polysaccharides**, il sera possible d'obtenir également des informations sur les isoméries de position. En effet, les fragments obtenus seront différenciables qu'il s'agisse d'un simple ou d'un double clivage [Weiskopf et al., 1998], [Morelle et al., 2004a], [Morelle et al., 2004b].

Enfin il est également possible de favoriser l'ionisation et la fragmentation de ces ions en convertissant **le glycanes en aminoglycanes** [Muzikar et al., 2004]. Le choix d'une stratégie ou d'une autre est souvent dictée selon le type de spectromètre de masse utilisé (MALDI-TOF/TOF, ESI-IT, ESI-Q/TOF, ...).

c) Analyse MSⁿ des N-glycopeptides

La fragmentation des N-glycopeptides permet de déterminer la **composition de chaque glycanes sur chaque site de glycosylation** mais également de **déterminer le site de glycosylation**.

En effet, de manière générale, les spectres de **fragmentation CID des N-glycopeptides** contiennent trois types d'ions fragments majoritaires :

- Des ions fragments correspondant à des **pertes des sucres successifs sur la chaîne peptidique** : ces ions permettent de déterminer les enchaînements de monosaccharides et ainsi de remonter à la composition de chaque glycanes.
- Des ions **fragments internes aux sucres de type Bi et Yi** : ces ions donnent des informations supplémentaires sur la composition du glycanes mais permettent surtout d'identifier rapidement si un spectre MS/MS correspond à celui d'un glycopeptide ou non, puisqu'on retrouvera dans

ce spectre les ions oxoniums utilisés comme ions diagnostiques. La présence de ces ions ne nécessite donc pas d'effectuer une première LC-MS pour détecter la zone d'éluion des glycopeptides.

- L'ion correspondant au **peptide non-glycosylé**.

Dans certains cas il est possible d'observer des ions fragments correspondant à la fragmentation du peptide non glycosylé permettant d'identifier la séquence en acides aminés du peptide (figure 9) [Wuhrer et al., 2007].

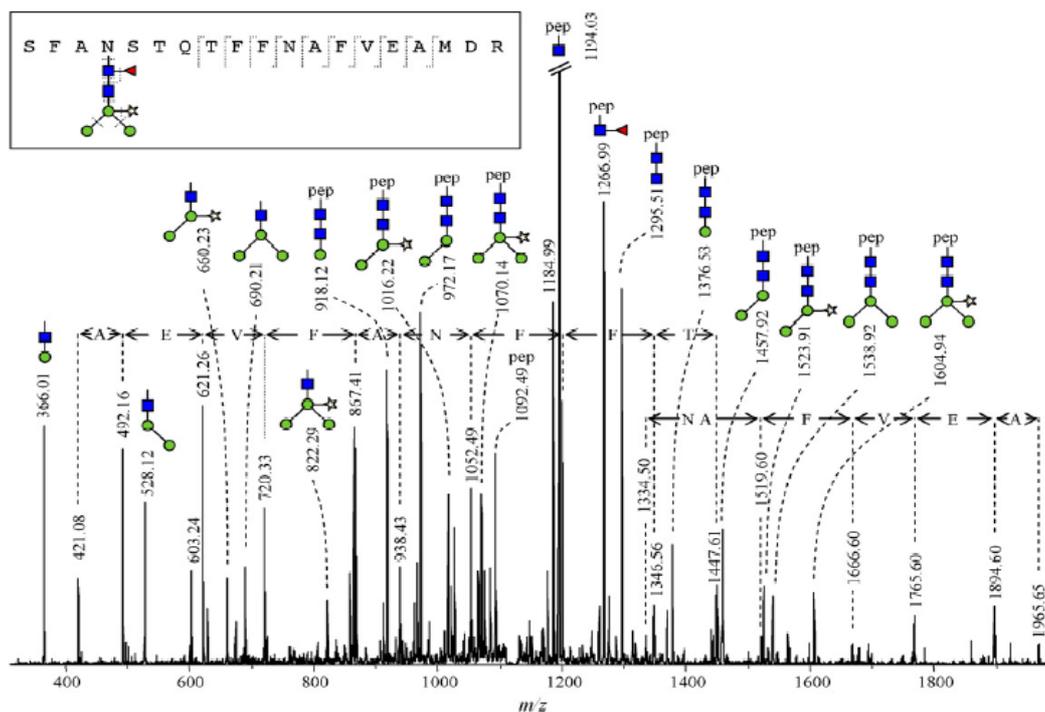


Figure 9 : Exemple d'un spectre MS/MS obtenu par CID de l'ion $(M+3H)^{3+}$ de m/z 1119 correspondant au glycopeptide tryptique Ser²⁹⁵-Arg³¹³ de la peroxidase de raifort (horseradish peroxidase HRP) obtenu sur une trappe Bruker Daltonics HCTultra [Wuhrer et al., 2007].

Néanmoins, la liaison glycosidique étant beaucoup plus labile que la liaison peptidique, les fragments les plus intenses correspondent à des pertes successives de sucres sur le glycopeptide, et la présence d'ions fragments correspondant au peptide sont peu intenses voire inexistantes. De ce fait, il est difficile de pouvoir déterminer la séquence du peptide à partir du spectre MS/MS. Toutefois, il n'est pas rare de constater que l'ion fragment le plus intense correspond au peptide avec un GlcNAc lié à l'asparagine. En utilisant des spectromètres de masse comme la trappe à ions, il est alors possible de sélectionner cet ion pour le fragmenter une seconde fois (MS^3) et ainsi déterminer la séquence du peptide [Demelbauer et al., 2004].

Plus récemment avec le développement de la source ETD, une nouvelle voie s'est ouverte pour la caractérisation de N-glycopeptides par MS^n . En effet la fragmentation par ETD étant aspécifique (Introduction Bibliographique, Chapitre 1, La fragmentation en mode ETD, p.29) il est possible

d'obtenir dans un même spectre MS/MS des ions fragments du peptide, des glycanes et du peptide glycosylé (figure 10).

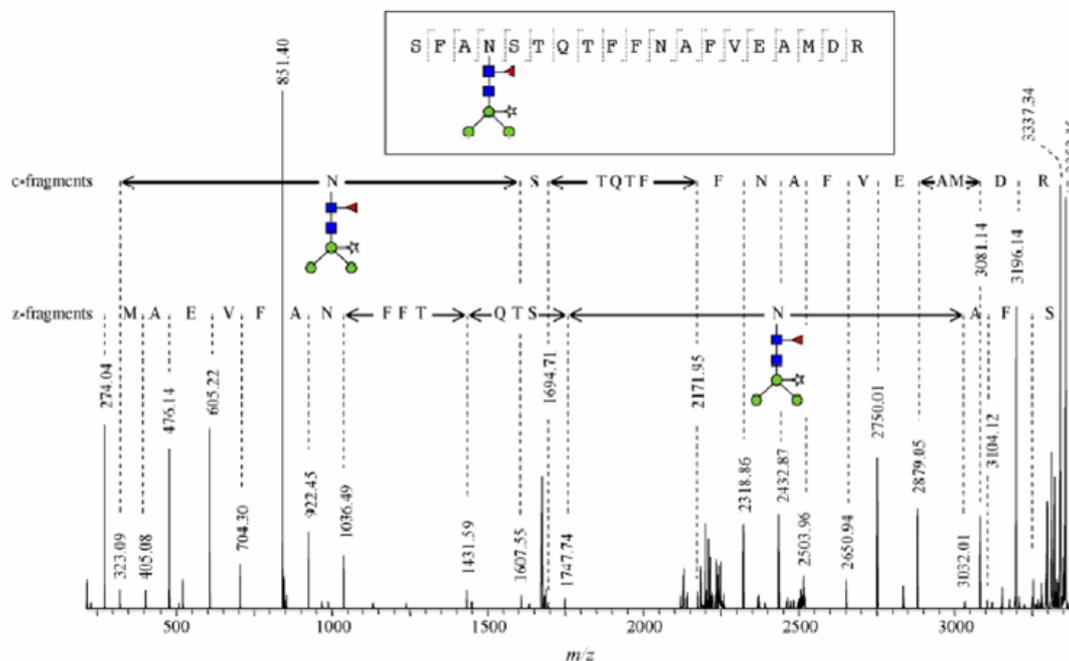


Figure 10 : Exemple d'un spectre MS/MS obtenu par ETD de l'ion $(M+3H)^{3+}$ de m/z 1119 correspondant au glycopeptide tryptique Ser²⁹⁵-Arg³¹³ de la peroxidase de raifort (horseradish peroxidase HRP) obtenu sur une trappe HCTultra (Bruker Daltonics) équipée avec une source ETD [Wuhrer et al., 2007].

De même en combinant des étapes multiples de fragmentation ETD /CID il sera également possible de remonter à la structure du glycanes et la séquence du peptide [Catalina et al., 2007].

L'analyse MS/MS de glycopeptides présente donc de nombreux avantages. Cependant, plusieurs paramètres sont à prendre en compte lors de l'analyse LC-MS/MS de glycopeptides. Dans un premier temps, il s'agit de bien choisir l'enzyme de digestion pour générer des glycopeptides de taille pas trop élevée pour qu'ils soient détectables dans les gammes de m/z des spectromètres de masse. Ensuite, il s'agit d'optimiser les paramètres de transmission de ces ions pour améliorer leur détection. Enfin il n'est pas peu courant de devoir favoriser leur sélection pour les fragmenter. C'est pourquoi l'analyse LC-MS/MS des glycopeptides en automatique peut rapidement s'avérer compliquée. Enfin leur interprétation est fastidieuse.

III. Conclusion

Nous avons pu voir qu'il existe de nombreuses stratégies pour la caractérisation des N-glycoprotéines. Chacune d'elle présente des avantages et des inconvénients, mais leur complémentarité est primordiale pour une caractérisation complète qui passe par l'identification de la protéine N-glycosylée, la détermination du site de N-glycosylation, la structure du N-glycane et l'étude de l'hétérogénéité globale des glycoformes.

Néanmoins, de nombreuses optimisations ont été apportées pour pouvoir analyser ces PTMs, qu'il s'agisse des modes d'enrichissements, des modes de marquage, des réactions chimiques ou enzymatiques pour décrocher les N-glycanes, mais également des améliorations et des optimisations dans les systèmes de chromatographie et les spectromètres de masse.

Cependant, même si de nombreux efforts ont été apportés pour leur caractérisation, chaque étude demande de nombreuses optimisations et de nouvelles méthodologies cela étant du au type d'échantillons et à la nature des protéines glycosylées (connue ou inconnue, 1 site ou plusieurs sites, de nombreux glycoformes ou quelques glycoformes, taille de la protéine ...).

Enfin, de manière générale les données générées lors de l'analyse MS ou MS/MS de glycanes et de glycopeptides se font manuellement cela étant du à leur complexité.

Cependant depuis quelques années, de nombreux efforts ont été apportés pour faciliter l'interprétation des données générées lors de la fragmentation et surtout dans le cas des glycanes. L'ensemble de ces logiciels est très bien décrit par von der Lieth [von der Lieth et al., 2004]. Parmi ceux-ci on trouve par exemple:

- Des logiciels qui prédisent les sites potentiels de glycosylation.
- Des outils pour l'aide à l'interprétation des spectres de fragmentation des glycanes.
- Des outils pour représenter les carbohydrates en 3D.
- Des banques de données de carbohydrates.

CONCLUSION

Avec les nombreux développements instrumentaux, l'analyse protéomique qui à l'heure actuelle consiste à identifier, quantifier et étudier les PTMs, a pris une place considérable dans l'étude de processus biologiques depuis les années 1990.

Aujourd'hui la stratégie d'identification en analyse protéomique la plus couramment utilisée est

- (1) séparation des protéines sur gel d'électrophorèse 2D
- (2) digestion à la trypsine des protéines en peptides
- (3) analyses MALDI-TOF-MS ou nanoHPLC-MS/MS des peptides
- (4) interprétation des données via des logiciels bioinformatiques et des recherches dans des banques protéiques.

Cependant, d'autres alternatives sont proposées qui présentent des avantages selon la nature des échantillons et le type d'étude menée. Ainsi, il n'est pas rare de constater qu'une stratégie n'est plus suffisante et que pour obtenir des études les plus complètes, il est nécessaire de combiner plusieurs stratégies de séparation, d'analyse par spectrométrie de masse et d'identification.

En effet, à tout stade d'une étude, plusieurs points sont à prendre en compte qui dépendent essentiellement de la problématique et de la nature des échantillons (figure 11):

- Identifier l'ensemble des protéines dans un extrait, quantifier des protéines dans un milieu particulier ou identifier des PTMs et des mutations sur des protéines ?
- Quel type d'échantillon (complexe ou non, protéines membranaires, protéines minoritaires)
- Séparer ou non les protéines ou les peptides et comment les séparer ?
- Quel spectromètre de masse utiliser ?

- Mesure de masse précise ou fragmentation ?
- Comment interpréter les données de MS ?

Même si à ce jour les stratégies en analyse protéomique pour l'identification, la quantification et la caractérisation de certaines PTMs sont bien établies, l'analyse protéomique reste complexe et continue d'évoluer d'un point de vue instrumental et méthodologique.

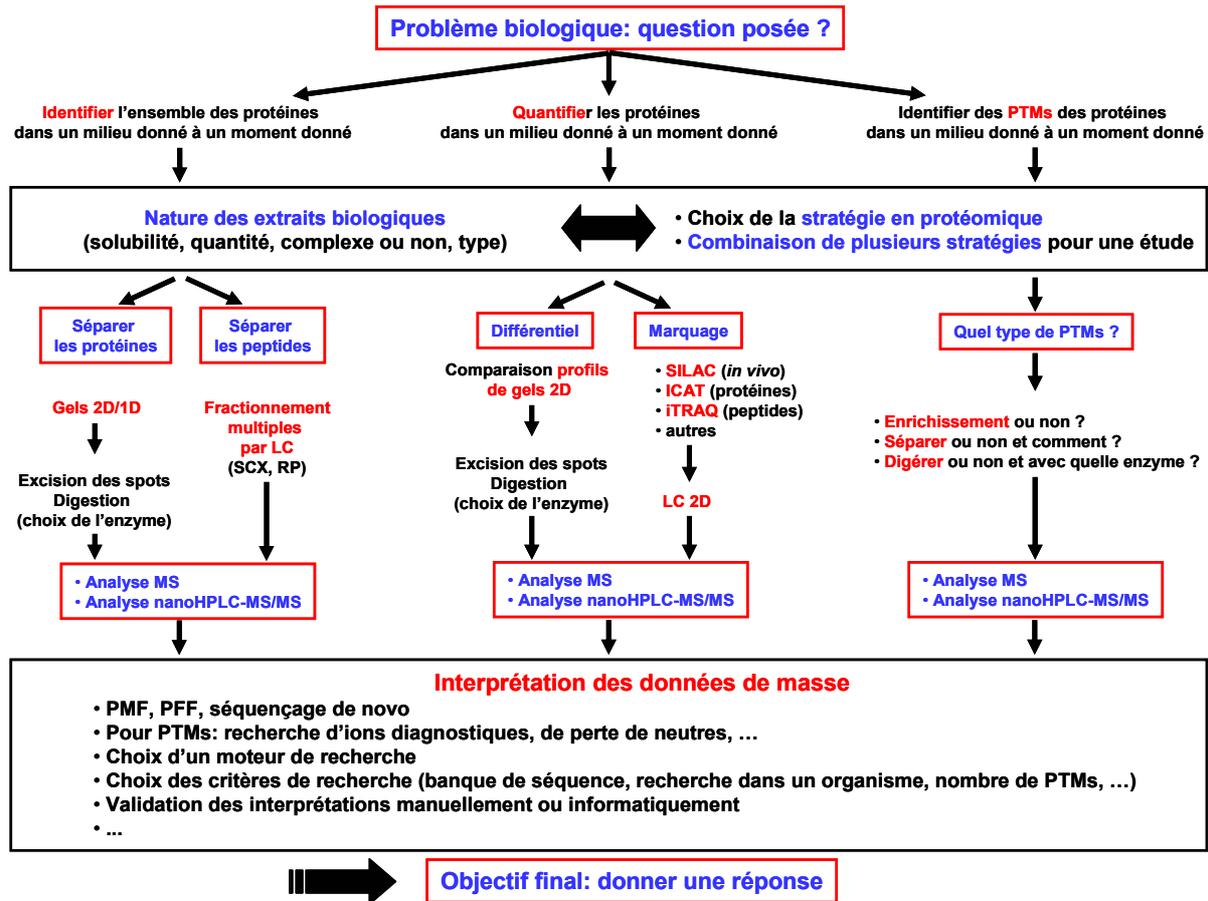


Figure 11 : Schéma des stratégies en analyses protéomique de la question posée à l'objectif final illustrant la complexité de l'analyse protéomique.

BIBLIOGRAPHIE

- Aebersold R. and Goodlett D.R.** (2001). Mass spectrometry in proteomics. *Chem Rev* **101**, 269-295.
- Aebersold R. and Mann M.** (2003). Mass spectrometry-based proteomics. *Nature* **422**, 198-207.
- An der Lan B. and Chrambach A.** (1985). Gel electrophoresis of proteins. A practical approach. IRL press, Oxford.
- An H.J., Peavy T.R., Hedrick J.L. and Lebrilla C.B.** (2003). Determination of N-glycosylation sites and site heterogeneity in glycoproteins. *Anal Chem* **75**, 5628-5637.
- Anderson N.G. and Anderson N.L.** (1996). Twenty years of two-dimensional electrophoresis: past, present and future. *Electrophoresis* **17**, 443-453.
- Anderson N.L. and Anderson N.G.** (1998). Proteome and proteomics: new technologies, new concepts, and new words. *Electrophoresis* **19**, 1853-1861.
- Apweiler R., Bairoch A. and Wu C.H.** (2004a). Protein sequence databases. *Curr Opin Chem Biol* **8**, 76-80.
- Apweiler R., Bairoch A., Wu C.H., Barker W.C., Boeckmann B., Ferro S., Gasteiger E., Huang H., Lopez R., Magrane M., Martin M.J., Natale D.A., O'Donovan C., Redaschi N. and Yeh L.S.** (2004b). UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res* **32**, D115-119.
- Apweiler R., Hermjakob H. and Sharon N.** (1999). On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim Biophys Acta* **1473**, 4-8.
- Bateman R.H., Carruthers R., Hoyes J.B., Jones C., Langridge J.I., Millar A. and Vissers J.P.** (2002). A novel precursor ion discovery method on a hybrid quadrupole orthogonal acceleration time-of-flight (Q-TOF) mass spectrometer for studying protein phosphorylation. *J Am Soc Mass Spectrom* **13**, 792-803.
- Benson D.A., Karsch-Mizrachi I., Lipman D.J., Ostell J. and Wheeler D.L.** (2003). GenBank. *Nucleic Acids Res* **31**, 23-27.
- Biemann K.** (1988). Contributions of mass spectrometry to peptide and protein structure. *Biomed Environ Mass Spectrom* **16**, 99-111.
- Biemann K.** (1990). Appendix 5. Nomenclature for peptide fragment ions (positive ions). *Methods Enzymol* **193**, 886-887.
- Biemann K.** (1992). Mass spectrometry of peptides and proteins. *Annu Rev Biochem* **61**, 977-1010.
- Bjellqvist B., Ek K., Righetti P.G., Gianazza E., Gorg A., Westermeier R. and Postel W.** (1982). Isoelectric focusing in immobilized pH gradients: principle, methodology and some applications. *J Biochem Biophys Methods* **6**, 317-339.
- Blades A.T., Ikonomou M.G. and Kebarle P.** (1991). Mechanism of Electrospray Mass Spectrometry, Electrospray as an Electrolysis Cell. *Anal Chem* **63**, 2109-2114.
- Blueggel M., Chamrad D. and Meyer H.E.** (2004). Bioinformatics in proteomics. *Curr Pharm Biotechnol* **5**, 79-88.

Bradshaw R.A., Burlingame A.L., Carr S. and Aebersold R. (2005). Protein identification: the good, the bad, and the ugly. *Mol Cell Proteomics* **4**, 1221-1222.

Braun R.J., Kinkl N., Beer M. and Ueffing M. (2007). Two-dimensional electrophoresis of membrane proteins. *Anal Bioanal Chem* **389**, 1033-1045.

Budnik B.A., Lee R.S. and Steen J.A. (2006). Global methods for protein glycosylation analysis by mass spectrometry. *Biochim Biophys Acta* **1764**, 1870-1880.

Campana J.E. (1980). Elementary theory of the quadrupole mass filter. *Int J Mass Spectrom Ion Processes* **33**, 101-117.

Caprioli R.M., Farmer T.B. and Gile J. (1997). Molecular imaging of biological samples: localization of peptides and proteins using MALDI-TOF MS. *Anal Chem* **69**, 4751-4760.

Carapito C. (2006). Vers une meilleure utilisation des données de spectrométrie de masse en analyse protéomique. Strasbourg: Thèse de l'Université Louis Pasteur.

Carr S.A., Huddleston M.J. and Bean M.F. (1993). Selective identification and differentiation of N- and O-linked oligosaccharides in glycoproteins by liquid chromatography-mass spectrometry. *Protein Sci* **2**, 183-196.

Catalina M.I., Koeleman C.A., Deelder A.M. and Wuhrer M. (2007). Electron transfer dissociation of N-glycopeptides: loss of the entire N-glycosylated asparagine side chain. *Rapid Commun Mass Spectrom* **21**, 1053-1061.

Cech N.B. and Enke C.G. (2001). Practical implications of some recent studies in electrospray ionization fundamentals. *Mass Spectrom Rev* **20**, 362-387.

Chernushevich I.V., Loboda A.V. and Thomson B.A. (2001). An introduction to quadrupole-time-of-flight mass spectrometry. *J Mass Spectrom* **36**, 849-865.

Cole R.B. (2000). Some tenets pertaining to electrospray ionization mass spectrometry. *J Mass Spectrom* **35**, 763-772.

Conrads T.P., Alving K., Veenstra T.D., Belov M.E., Anderson G.A., Anderson D.J., Lipton M.S., Pasa-Tolic L., Udseth H.R., Chrisler W.B., Thrall B.D. and Smith R.D. (2001). Quantitative analysis of bacterial and mammalian proteomes using a combination of cysteine affinity tags and ¹⁵N-metabolic labeling. *Anal Chem* **73**, 2132-2139.

Cooper H.J., Hakansson K. and Marshall A.G. (2005). The role of electron capture dissociation in biomolecular analysis. *Mass Spectrom Rev* **24**, 201-222.

Cravatt B.F., Simon G.M. and Yates J.R., 3rd. (2007). The biological impact of mass-spectrometry-based proteomics. *Nature* **450**, 991-1000.

De Hoffmann E., Charette J. and Stroobant V. (1994). Spectrométrie de masse. Paris: MASSON.

Demelbauer U.M., Zehl M., Plematl A., Allmaier G. and Rizzi A. (2004). Determination of glycopeptide structures by multistage mass spectrometry with low-energy collision-induced dissociation: comparison of electrospray ionization quadrupole ion trap and matrix-assisted laser desorption/ionization quadrupole ion trap reflectron time-of-flight approaches. *Rapid Commun Mass Spectrom* **18**, 1575-1582.

Dole M., Mack L.L., Himes R.L., Mobley R.C., Ferguson L.D. and Alice M.B. (1968). Molecular beams of macroions. *J. Chem. Phys.* **49**, 2240-2249.

Domon B. and Aebersold R. (2006). Mass spectrometry and protein analysis. *Science* **312**, 212-217.

Domon B. and Costello C.E. (1988). A Systematic Nomenclature for Carbohydrate Fragmentation in FAB-MS/MS spectra of Glycoconjugates. *Glycoconj J* **5**, 397-409.

Dongré A.R., Jones J.L., Somogyi A. and Wysocki V.H. (1996). Influence of peptide composition, gas-phase basicity, and chemical modification on fragmentation efficiency: Evidence for the mobile proton model. *J Am Chem Soc* **118**, 8365-8374.

Edman P. (1949). A method for the determination of amino acid sequence in peptides. *Arch Biochem* **22**, 475.

Edman P. and Begg G. (1967). A protein sequenator. *Eur J Biochem* **1**, 80-91.

Emmett M.R. and Caprioli R.M. (1994). Micro-electrospray: ultra-high-sensitivity analysis of peptides and proteins. *J. Am. Soc. Mass Spectrom.* **5**, 605-613.

Eng J.K., McCormack A.L. and Yates J.R.I. (1994). An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* **5**, 976-989.

Fenselau C. and Yao X. (2007). Proteolytic labeling with 18O for comparative proteomics studies: preparation of 18O-labeled peptides and the 18O/16O peptide mixture. *Methods Mol Biol* **359**, 135-142.

Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F., Kerlavage A.R., Bult C.J., Tomb J.F., Dougherty B.A., Merrick J.M. and et al. (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**, 496-512.

Fournier M.L., Gilmore J.M., Martin-Brown S.A. and Washburn M.P. (2007). Multidimensional separations-based shotgun proteomics. *Chem Rev* **107**, 3654-3686.

Fraterman S., Zeiger U., Khurana T.S., Rubinstein N.A. and Wilm M. (2007). Combination of peptide OFFGEL fractionation and label-free quantitation facilitated proteomics profiling of extraocular muscle. *Proteomics* **7**, 3404-3416.

Gevaert K. and Vandekerckhove J. (2000). Protein identification methods in proteomics. *Electrophoresis* **21**, 1145-1154.

Geyer H. and Geyer R. (2006). Strategies for analysis of glycoprotein glycosylation. *Biochim Biophys Acta* **1764**, 1853-1869.

Gomord V. and Faye L. (2004). Posttranslational modification of therapeutic proteins in plants. *Curr Opin Plant Biol* **7**, 171-181.

Gorg A., Postel W., Domscheit A. and Gunther S. (1988). Two-dimensional electrophoresis with immobilized pH gradients of leaf proteins from barley (*Hordeum vulgare*): method, reproducibility and genetic aspects. *Electrophoresis* **9**, 681-692.

Gorg A., Weiss W. and Dunn M.J. (2004). Current two-dimensional electrophoresis technology for proteomics. *Proteomics* **4**, 3665-3685.

Gygi S.P., Corthals G.L., Zhang Y., Rochon Y. and Aebersold R. (2000). Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *Proc Natl Acad Sci U S A* **97**, 9390-9395.

Gygi S.P., Rist B., Gerber S.A., Turecek F., Gelb M.H. and Aebersold R. (1999). Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* **17**, 994-999.

Harvey D.J. (2001). Identification of protein-bound carbohydrates by mass spectrometry. *Proteomics* **1**, 311-328.

Helenius A. and Aebi M. (2001). Intracellular functions of N-linked glycans. *Science* **291**, 2364-2369.

Heller M., Mattou H., Menzel C. and Yao X. (2003). Trypsin catalyzed 16O-to-18O exchange for comparative proteomics: tandem mass spectrometry comparison using MALDI-TOF, ESI-QTOF, and ESI-ion trap mass spectrometers. *J Am Soc Mass Spectrom* **14**, 704-718.

Henzel W.J., Billeci T.M., Stults J.T., Wong S.C., Grimley C. and Watanabe C. (1993). Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. *Proc Natl Acad Sci U S A* **90**, 5011-5015.

Horth P., Miller C.A., Preckel T. and Wenz C. (2006). Efficient fractionation and improved protein identification by peptide OFFGEL electrophoresis. *Mol Cell Proteomics* **5**, 1968-1974.

Hörth P., Preckel T., Wenz C. and Miller C.A. (2006). Application Note 5989-5206EN : Efficient peptide fractionation and improved protein identification with the Agilent 3100 OFFGEL Fractionator. *Agilent Technologies*.

Hunt D.F., Henderson R.A., Shabanowitz J., Sakaguchi K., Michel H., Sevilir N., Cox A.L., Appella E. and Engelhard V.H. (1992). Characterization of peptides bound to the class I MHC molecule HLA-A2.1 by mass spectrometry. *Science* **255**, 1261-1263.

Ikonomou M.G., Blades A.T. and Kebarle P. (1991). Electrospray-Ion spray: A comparison of mechanisms and performance. *Anal. Chem.* **63**, 1989-1998.

Iribarne J.V. and Thomson B.A. (1976). On the evaporation of small ions from charged droplets. *The Journal of Chemical Physics* **64**, 2287-2294.

Issaq H.J. (2001). The role of separation science in proteomics research. *Electrophoresis* **22**, 3629-3638.

Issaq H.J. (2003). Application of separation technologies to proteomics research. *Adv Protein Chem* **65**, 249-269.

Issaq H.J., Chan K.C., Janini G.M., Conrads T.P. and Veenstra T.D. (2005). Multidimensional separation of peptides for effective proteomic analysis. *J Chromatogr B Analyt Technol Biomed Life Sci* **817**, 35-47.

James P., Quadroni M., Carafoli E. and Gonnet G. (1993). Protein identification by mass profile fingerprinting. *Biochem Biophys Res Commun* **195**, 58-64.

Jebanathirajah J., Steen H. and Roepstorff P. (2003). Using optimized collision energies and high resolution, high accuracy fragment ion selection to improve glycopeptide detection by precursor ion scanning. *J Am Soc Mass Spectrom* **14**, 777-784.

Jensen O.N. (2004). Modification-specific proteomics: characterization of post-translational modifications by mass spectrometry. *Curr Opin Chem Biol* **8**, 33-41.

Jensen O.N. (2006). Interpreting the protein language using proteomics. *Nat Rev Mol Cell Biol* **7**, 391-403.

Jiang H., Desaire H., Butnev V.Y. and Bousfield G.R. (2004). Glycoprotein profiling by electrospray mass spectrometry. *J Am Soc Mass Spectrom* **15**, 750-758.

Kaltschmidt E. and Wittmann H.G. (1970). Ribosomal proteins. VII. Two-dimensional polyacrylamide gel electrophoresis for fingerprinting of ribosomal proteins. *Anal Biochem* **36**, 401-412.

Karas M., Bachmann D., Bahr U. and Hillenkamp F. (1987). Matrix-assisted ultraviolet laser desorption of non-volatile compounds. *Int. J. Mass Spectrom. Ion Processes* **78**, 53-68.

Karas M. and Hillenkamp F. (1988). Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem* **60**, 2299-2301.

Kay R.G., Gregory B., Grace P.B. and Pleasance S. (2007). The application of ultra-performance liquid chromatography/tandem mass spectrometry to the detection and quantitation of apolipoproteins in human serum *Rapid Commun Mass Spectrom* **21**, 2585-2593.

Kebarle P. (2000a). A brief overview of the present status of the mechanisms involved in electrospray mass spectrometry. *J Mass Spectrom* **35**, 804-817.

Kebarle P. and Peschke M. (2000b). On the mechanisms by which the charged droplets produced by electrospray lead to gas phase ions. *Analytical Chimica Acta* **406**, 11-35.

Kenrick K.G. and Margolis J. (1970). Isoelectric focusing and gradient gel electrophoresis: a two-dimensional technique. *Anal Biochem* **33**, 204-207.

Knochenmuss R. and Zenobi R. (2003). MALDI ionization: the role of in-plume processes. *Chem Rev* **103**, 441-452.

Kulikova T., Aldebert P., Althorpe N., Baker W., Bates K., Browne P., van den Broek A., Cochrane G., Duggan K., Eberhardt R., Faruque N., Garcia-Pastor M., Harte N., Kanz C., Leinonen R., Lin Q., Lombard V., Lopez R., Mancuso R., McHale M., Nardone F., Silventoinen V., Stoehr P., Stoesser G., Tuli M.A., Tzouvara K., Vaughan R., Wu D., Zhu W. and Apweiler R. (2004). The EMBL Nucleotide Sequence Database. *Nucleic Acids Res* **32**, D27-30.

Laemmli U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685.

Lane C.S. (2005). Mass spectrometry-based proteomics in the life sciences. *Cell Mol Life Sci* **62**, 848-869.

Lee R.T., Lauc G. and Lee Y.C. (2005). Glycoproteomics: protein modifications for versatile functions. *EMBO Rep* **6**, 1018-1022.

Link A.J., Eng J., Schieltz D.M., Carmack E., Mize G.J., Morris D.R., Garvik B.M. and Yates J.R., 3rd. (1999). Direct analysis of protein complexes using mass spectrometry. *Nat Biotechnol* **17**, 676-682.

Liu H., Finch J.W., Lavalley M.J., Collamati R.A., Benevides C.C. and Gebler J.C. (2007). Effects of column length, particle size, gradient length and flow rate on peak capacity of nano-scale liquid chromatography for peptide separations. *J Chromatogr A* **1147**, 30-36.

Lopez L.J. (2007). Two-dimensional electrophoresis in proteome expression analysis. *Journal of Chromatography B* **849**, 190-202.

Ma B., Zhang K., Hendrie C., Liang C., Li M., Doherty-Kirby A. and Lajoie G. (2003). PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid Commun Mass Spectrom* **17**, 2337-2342.

Makarov A. (2000). Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. *Anal Chem* **72**, 1156-1162.

Makarov A., Denisov E., Kholomeev A., Balschun W., Lange O., Strupat K. and Horning S. (2006). Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. *Anal Chem* **78**, 2113-2120.

Mamyrin B.A., Karataev V.I., Shmikk D.V. and Zagulin V.A. (1973). The mass-reflectron, a new nonmagnetic time-of-flight mass spectrometer with high resolution. *Zh. Eksp. Theor. Fiz.* **64**, 82-89.

Mann M., Hojrup P. and Roepstorff P. (1993). Use of mass spectrometric molecular weight information to identify proteins in sequence databases. *Biol Mass Spectrom* **22**, 338-345.

Mann M. and Jensen O.N. (2003). Proteomic analysis of post-translational modifications. *Nat Biotechnol* **21**, 255-261.

Mann M. and Wilm M. (1994). Error-tolerant identification of peptides in sequence databases by peptide sequence tags. *Anal Chem* **66**, 4390-4399.

March R.E. (1997). An introduction to quadrupole ion trap mass spectrometry. *J Mass Spectrom* **32**, 351-369.

March R.E. (1998). Quadrupole ion trap mass spectrometry: theory, simulation, recent developments and applications. *Rapid Commun Mass Spectrom* **12**, 1543-1554.

March R.E. (2000). Quadrupole ion trap mass spectrometry: a view at the turn of the century. *International Journal of Mass Spectrometry* **200**, 285-312.

Mechref Y. and Novotny M.V. (2002). Structural investigations of glycoconjugates at high sensitivity. *Chem Rev* **102**, 321-369.

Michel P.E., Reymond F., Arnaud I.L., Josserand J., Girault H.H. and Rossier J.S. (2003a). Protein fractionation in a multicompartiment device using Off-Gel isoelectric focusing. *Electrophoresis* **24**, 3-11.

Michel P.E., Reymond F., Arnaud I.L., Josserand J., Girault H.H. and Rossier J.S. (2003b). Protein fractionation in a multicompartiment device using Off-Gel isoelectric focusing. *Electrophoresis* **24**, 3-11.

Miguet L. (2006). Nouvelles méthodologies dans l'analyse protéomique par spectrométrie de masse. Application à la recherche de biomarqueurs dans le cadre des leucémies. Strasbourg: Thèse de l'Université Louis Pasteur.

Mirgorodskaya O.A., Kozmin Y.P., Titov M.I., Korner R., Sonksen C.P. and Roepstorff P. (2000). Quantitation of peptides and proteins by matrix-assisted laser desorption/ionization mass spectrometry using (18)O-labeled internal standards. *Rapid Commun Mass Spectrom* **14**, 1226-1232.

Miyazaki S., Sugawara H., Gojobori T. and Tateno Y. (2003). DNA Data Bank of Japan (DDBJ) in XML. *Nucleic Acids Res* **31**, 13-16.

Morelle W., Faid V. and Michalski J.C. (2004a). Structural analysis of permethylated oligosaccharides using electrospray ionization quadrupole time-of-flight tandem mass spectrometry and deuteration. *Rapid Commun Mass Spectrom* **18**, 2451-2464.

Morelle W., Slomianny M.C., Diemer H., Schaeffer C., van Dorsselaer A. and Michalski J.C. (2004b). Fragmentation characteristics of permethylated oligosaccharides using a matrix-assisted laser desorption/ionization two-stage time-of-flight (TOF/TOF) tandem mass spectrometer. *Rapid Commun Mass Spectrom* **18**, 2637-2649.

Morelle W., Slomianny M.C., Diemer H., Schaeffer C., van Dorsselaer A. and Michalski J.C. (2005). Structural characterization of 2-aminobenzamide-derivatized oligosaccharides using a matrix-assisted laser desorption/ionization two-stage time-of-flight tandem mass spectrometer. *Rapid Commun Mass Spectrom* **19**, 2075-2084.

Morris H.R., Chalabia S., Panico M., Sutton-Smith M., Clarke G.F., Goldberg D. and Della A. (2007). Glycoproteomics: Past, present and future *Int J Mass Spectrom* **259**, 16-31.

Morris H.R., Thompson M.R., Osuga D.T., Ahmed A.I., Chan S.M., Vandenheede J.R. and Feeney R.E. (1978). Antifreeze glycoproteins from the blood of an antarctic fish. The structure of the proline-containing glycopeptides. *J Biol Chem* **253**, 5155-5162.

Muzikar J., Mechref Y., Huang Y. and Novotny M.V. (2004). Enhanced post-source decay and cross-ring fragmentation of oligosaccharides facilitated by conversion to amino derivatives. *Rapid Commun Mass Spectrom* **18**, 1513-1518.

Nesvizhskii A.I. and Aebersold R. (2005). Interpretation of shotgun proteomic data: the protein inference problem. *Mol Cell Proteomics* **4**, 1419-1440.

Nesvizhskii A.I., Vitek O. and Aebersold R. (2007). Analysis and validation of proteomic data generated by tandem mass spectrometry. *Nat Methods* **4**, 787-797.

Neuhoff V., Arold N., Taube D. and Ehrhardt W. (1988). Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250. *Electrophoresis* **9**, 255-262.

Neverova I. and Van Eyk J.E. (2005). Role of chromatographic techniques in proteomic analysis. *J Chromatogr B Analyt Technol Biomed Life Sci* **815**, 51-63.

Oda Y., Huang K., Cross F.R., Cowburn D. and Chait B.T. (1999). Accurate quantitation of protein expression and site-specific phosphorylation. *Proc Natl Acad Sci U S A* **96**, 6591-6596.

Ohyama Y., Kasai K., Nomoto H. and Inoue Y. (1985). Frontal affinity chromatography of ovalbumin glycoasparagines on a concanavalin A-sepharose column. A quantitative study of the binding specificity of the lectin. *J Biol Chem* **260**, 6882-6887.

Ong S.E., Blagoev B., Kratchmarova I., Kristensen D.B., Steen H., Pandey A. and Mann M. (2002). Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* **1**, 376-386.

Ong S.E. and Pandey A. (2001). An evaluation of the use of two-dimensional gel electrophoresis in proteomics. *Biomol Eng* **18**, 195-205.

Oosterkamp A.J., Gelpi E. and Abian J. (1998). Quantitative peptide bioanalysis using column-switching nano liquid chromatography/mass spectrometry. *J Mass Spectrom* **33**, 976-983.

Ouyang Z., Gao L., Fico M., Chappell W.J., Noll R.J. and Cooks R.G. (2007). Quadrupole ion traps and trap arrays: geometry, material, scale, performance. *Eur J Mass Spectrom (Chichester, Eng)* **13**, 13-18.

Ouyang Z., Wu G., Song Y., Li H., Plass W.R. and Cooks R.G. (2004). Rectilinear ion trap: concepts, calculations, and analytical performance of a new mass analyzer. *Anal Chem* **76**, 4595-4605.

Paizs B. and Suhai S. (2005). Fragmentation pathways of protonated peptides. *Mass Spectrom Rev* **24**, 508-548.

Pandey A. and Mann M. (2000). Proteomics to study genes and genomes. *Nature* **405**, 837-846.

Pappin D.J., Hojrup P. and Bleasby A.J. (1993). Rapid identification of proteins by peptide-mass fingerprinting. *Curr Biol* **3**, 327-332.

Patterson S.D. and Aebersold R.H. (2003). Proteomics: the first decade and beyond. *Nat Genet* **33 Suppl**, 311-323.

Paul W. (1990). Elektromagnetische Käfige für geladene und neutrale Teilchen (Nobel-Vortrag). *Angew Chem* **102**, 780-789.

Paul W. and Steinwedel H. (1953). Ein neues Massenspektrometer ohne Magnetfeld. *Zeitschrift für Naturforschung* **8a**, 448.

Peng J. and Gygi S.P. (2001). Proteomics: the move to mixtures. *J Mass Spectrom* **36**, 1083-1091.

Perkins D.N., Pappin D.J., Creasy D.M. and Cottrell J.S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551-3567.

Quarmby S.T. and Yost R.A. (1999). Fundamental studies of ion injection and trapping of electrosprayed ions on a quadrupole ion trap. *Int J Mass Spectrom* **190/191**, 81-102.

Rabilloud T. (2000). Proteome research: Two-dimensional gel electrophoresis and identification methods. Berlin Heidelberg.

Richert S., Luche S., Chevallet M., Van Dorsselaer A., Leize-Wagner E. and Rabilloud T. (2004). About the mechanism of interference of silver staining with peptide mass spectrometry. *Proteomics* **4**, 909-916.

Roe M.R. and Griffin T.J. (2006). Gel-free mass spectrometry-based high throughput proteomics: tools for studying biological response of proteins and proteomes. *Proteomics* **6**, 4678-4687.

Roepstorff P. and Fohlman J. (1984). Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed Mass Spectrom* **11**, 601.

Roepstorff P. and Kristiansen K. (1974). The use of Edman degradation in peptide mixture analysis by mass spectrometry. *Biomed Mass Spectrom* **1**, 231-236.

Ross P.L., Huang Y.N., Marchese J.N., Williamson B., Parker K., Hattan S., Khainovski N., Pillai S., Dey S., Daniels S., Purkayastha S., Juhasz P., Martin S., Bartlet-Jones M., He F., Jacobson A. and Pappin D.J. (2004). Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics* **3**, 1154-1169.

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

Ruhaak L.R., Huhn C., Waterreus W.J., de Boer A.R., Neuss C., Hokke C.H., Deelder A.M. and Wührer M. (2008). Hydrophilic interaction chromatography-based high-throughput sample preparation method for N-glycan analysis from total human plasma glycoproteins. *Anal Chem* **80**, 6119-6126.

Santoni V., Molloy M. and Rabilloud T. (2000). Membrane proteins and proteomics: un amour impossible? *Electrophoresis* **21**, 1054-1070.

Shevchenko A., Sunyaev S., Loboda A., Shevchenko A., Bork P., Ens W. and Standing K.G. (2001). Charting the proteomes of organisms with unsequenced genomes by MALDI-quadrupole time-of-flight mass spectrometry and BLAST homology searching. *Anal Chem* **73**, 1917-1926.

Shevchenko A., Wilm M., Vorm O. and Mann M. (1996). Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal Chem* **68**, 850-858.

Siuti N. and Kelleher N.L. (2007). Decoding protein modifications using top-down mass spectrometry. *Nat Methods* **4**, 817-821.

Stafford G., Jr. (2002). Ion trap mass spectrometry: a personal perspective. *J Am Soc Mass Spectrom* **13**, 589-596.

Stafford G.C., Kelley P.E., Syka J.E.P., Reynolds W.E. and Todd J.F.J. (1984). Recent improvements in and analytical applications of advanced ion trap technologies. *Int J Mass Spectrom Ion Proc* **60**, 85-98.

Sullivan B., Addona T.A. and Carr S.A. (2004). Selective detection of glycopeptides on ion trap mass spectrometers. *Anal Chem* **76**, 3112-3118.

Syka J.E., Coon J.J., Schroeder M.J., Shabanowitz J. and Hunt D.F. (2004). Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci U S A* **101**, 9528-9533.

Tanaka K., Waki H., Ido Y., Akita S., Yoshida Y. and Yoshida T. (1988). Protein and polymer analyses up to m/z 100,000 by laser ionization time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom* **2**, 151-153.

Tsarbopoulos A., Bahr U., Pramanik B.N. and Karas M. (1997). Glycoprotein analysis by delayed extraction and post-source decay MALDI-TOF-MS. *Int J Mass Spectrom and Ion Proc* **169-170**, 251-261.

Tyers M. and Mann M. (2003). From genomics to proteomics. *Nature* **422**, 193-197.

Varki A. (1993). Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology* **3**, 97-130.

Vestal M.C., Juhasz P. and Martin S.A. (1995). Delayed extraction matrix-assisted laser desorption time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom* **9**, 1044-1050.

von der Lieth C.W., Bohne-Lang A., Lohmann K.K. and Frank M. (2004). Bioinformatics for glycomics: status, methods, requirements and perspectives. *Brief Bioinform* **5**, 164-178.

Walsh C.T., Garneau-Tsodikova S. and Gatto G.J., Jr. (2005). Protein posttranslational modifications: the chemistry of proteome diversifications. *Angew Chem Int Ed Engl* **44**, 7342-7372.

Weiskopf A.S., Vouros P. and Harvey D.J. (1998). Electrospray ionization-ion trap mass spectrometry for structural analysis of complex N-linked glycoprotein oligosaccharides. *Anal Chem* **70**, 4441-4447.

Wiley W.C. and McLaren I.H. (1955). Time-of-flight mass spectrometry with improved resolution. *Rev Sci Instrum* **26**, 1150-1157.

Wilkins M.R., Appel R.D., Van Eyk J.E., Chung M.C., Gorg A., Hecker M., Huber L.A., Langen H., Link A.J., Paik Y.K., Patterson S.D., Pennington S.R., Rabilloud T., Simpson R.J., Weiss W. and Dunn M.J. (2006). Guidelines for the next 10 years of proteomics. *Proteomics* **6**, 4-8.

Wilkins M.R., Pasquali C., Appel R.D., Ou K., Golaz O., Sanchez J.C., Yan J.X., Gooley A.A., Hughes G., Humphery-Smith I., Williams K.L. and Hochstrasser D.F. (1996). From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Biotechnology (N Y)* **14**, 61-65.

Wilm M. and Mann M. (1994). Electrospray and Taylor-Cone theory, Dole's beam of macromolecules at last? *International Journal of Mass Spectrometry and Ion Processes* **136**, 167-180.

Wilm M. and Mann M. (1996). Analytical properties of the nanoelectrospray ion source. *Anal Chem* **68**, 1-8.

Witze E.S., Old W.M., Resing K.A. and Ahn N.G. (2007). Mapping protein post-translational modifications with mass spectrometry. *Nat Methods* **4**, 798-806.

Wolters D.A., Washburn M.P. and Yates J.R., 3rd. (2001). An automated multidimensional protein identification technology for shotgun proteomics. *Anal Chem* **73**, 5683-5690.

Wood T.D., Chen L.H., White C.B., Babbitt P.C., Kenyon G.L. and McLafferty F.W. (1995). Sequence verification of human creatine kinase (43 kDa) isozymes by high-resolution tandem mass spectrometry. *Proc Natl Acad Sci U S A* **92**, 11451-11455.

Wu C.C., MacCoss M.J., Howell K.E. and Yates J.R., 3rd. (2003). A method for the comprehensive proteomic analysis of membrane proteins *Nat Biotechnol* **21**, 532-538.

Wu C.C. and Yates J.R., 3rd. (2003). The application of mass spectrometry to membrane proteomics *Nat Biotechnol* **21**, 262-267.

Wuhrer M., Catalina M.I., Deelder A.M. and Hokke C.H. (2007). Glycoproteomics based on tandem mass spectrometry of glycopeptides. *J Chromatogr B Analyt Technol Biomed Life Sci* **849**, 115-128.

Wuhrer M., Deelder A.M. and Hokke C.H. (2005). Protein glycosylation analysis by liquid chromatography-mass spectrometry. *J Chromatogr B Analyt Technol Biomed Life Sci* **825**, 124-133.

Wysocki V.H., Tsaprailis G., Smith L.L. and Breci L.A. (2000). Mobile and localized protons: a framework for understanding peptide dissociation. *J Mass Spectrom* **35**, 1399-1406.

Yamashita M. and Fenn J. (1984). Electrospray ion source. Another variation on the Free-Jet theme. *J Phys Chem* **88**, 4451-4460.

Yates J.R., 3rd, Speicher S., Griffin P.R. and Hunkapiller T. (1993). Peptide mass maps: a highly informative approach to protein identification. *Anal Biochem* **214**, 397-408.

Yin H. and Killeen K. (2007). The fundamental aspects and applications of Agilent HPLC-Chip. *J Sep Sci* **30**, 1427-1434.

Yin H., Killeen K., Brennen R., Sobek D., Werlich M. and van de Goor T. (2005). Microfluidic chip for peptide analysis with an integrated HPLC column, sample enrichment column, and nanoelectrospray tip. *Anal Chem* **77**, 527-533.

Zaia J. (2004). Mass spectrometry of oligosaccharides. *Mass Spectrom Rev* **23**, 161-227.

Zenobi R. and Knochenmuss R. (1998). Ion formation in MALDI mass spectrometry. *Mass Spectrom Rev* **17**, 337-366.

Zubarev R.A., Kelleher N.L. and McLafferty F.W. (1998). Electron capture dissociation of multiply charged protein cations. A nonergotic process. *J Am Chem Soc* **120**, 3265-3266.

PARTIE RESULTATS

CHAPITRE 1

Caractérisation fine des glycoformes d'une hormone de fécondité, la « Follicle Stimulating Hormone » (FSH)

CHAPITRE 2

Caractérisation fine des N-glycosylations : stratégies basées sur l'analyse nanoLC-MS/MS de glycopeptides

CHAPITRE 3

Etude du protéome du cheveu humain

CHAPITRE 1

CARACTERISATION FINE DES GLYCOFORMES D'UNE HORMONE DE FECONDITE, LA « FOLLICLE STIMULATING HORMONE » (FSH)

Cette étude a fait l'objet d'une collaboration avec Merck-Serono (Italie) qui s'est employé à nous fournir les lots de FSH recombinante humaine (r-hFSH). Les échantillons confiés par Merck-Serono ne correspondent en aucun cas à des lots commercialisés.

La **FSH** est une glycoprotéine impliquée dans la **croissance des follicules ovariens** dont la nature et la distribution de ces glycoformes varient durant le cycle menstruel. Des traitements à base d'injection de r-hFSH ont été élaborés et sont aujourd'hui utilisés pour combattre l'infertilité.

Néanmoins, ce n'est que très récemment que **l'importance de la micro-hétérogénéité de la glycosylation durant le cycle menstruel** a été mise en évidence grâce à la co-cristallisation de la FSH avec son récepteur. De ce fait, il est apparu qu'il fallait disposer de méthodes analytiques permettant de **déterminer exactement quelles étaient les glycoformes présentes sur chaque site de glycosylation** pour pouvoir en perspective améliorer les traitements contre l'infertilité.

Dans ce but, deux stratégies ont été développées et ont débouché sur deux méthodes dont nous avons démontré la reproductibilité et la robustesse:

- **L'analyse de la protéine entière** par ESI-MS afin d'étudier son hétérogénéité globale.
- **L'analyse par LC-MS des glycopeptides** issus de la digestion de la protéine afin d'étudier la micro-hétérogénéité sur chaque site de glycosylation de la FSH.

Les optimisations méthodologiques ont été effectuées sur un échantillon de r-hFSH de référence de composition connue en glycoformes puis, par la suite, ces deux méthodologies ont été appliquées à des lots de r-hFSH purifiés selon différentes méthodes afin d'enrichir ceux-ci en glycoformes selon le degré d'acides sialiques ou selon l'antennarité.

I. Introduction

La **Follicle Stimulating Hormone** (FSH) fait partie de la famille des gonadotrophines, des hormones pituitaires, qui incluent la LH (Luteinizing Hormone), la CG (Chorionic Gonadotropin) et la TSH (Thyroid-Stimulating Hormone). Elle est produite et sécrétée dans les glandes pituitaires sous forme d'une glycoprotéine très hétérogène et est véhiculée par le sang. Elle agit en se fixant à des récepteurs spécifiques couplés à des protéines G localisés exclusivement dans les cellules granulosas [Fan and Hendrickson, 2005].

Elle joue un **rôle central dans la régulation de la reproduction des mammifères** [Ulloa-Aguirre et al., 1999] en étant impliquée dans des processus tels que le développement des follicules et l'ovulation. Elle agit chez la femme en **stimulant la croissance des follicules ovariens** et chez l'homme, en stimulant les cellules de Sertoli des testicules, qui produisent les spermatozoïdes. Sa sécrétion est donc pulsative et cyclique chez la femme, alors qu'elle est continue chez l'homme.

Une **baisse ou un déséquilibre** des taux de FSH peut conduire à une **absence d'ovulation**, et donc à une **infertilité** féminine.

I.1. Structure de la FSH

Plusieurs études utilisant la cristallographie mais également la spectrométrie de masse ont permis de caractériser la FSH [Fox et al., 2001], [Gervais et al., 2003].

Elle est constituée de **deux sous-unités** (alpha et beta). La **sous-unité alpha** se constitue de 92 acides aminés dont deux asparagines sont glycosylées (**Asn⁵²** et **Asn⁷⁸**). La **sous-unité beta** est constituée de 111 acides aminés dont deux asparagines sont glycosylées (**Asn⁷** et **Asn²⁴**) [Loumaye et al., 1998] (figure 1).

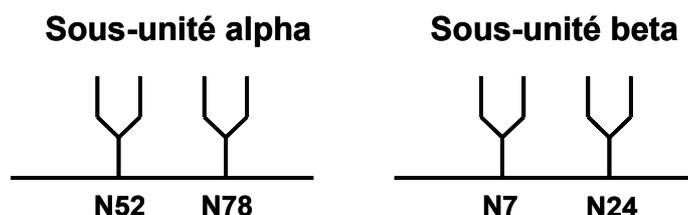


Figure 1 : Représentation schématique des deux sous-unités de la FSH.

Les N-glycanes sont de **type complexe** et **peuvent être bi-, tri- ou tetra-antennés** avec un **nombre variable d'acides N-acétylneuraminiques** (ou acides sialiques) à leur extrémité (figure 2). De ce fait, il existe de nombreuses combinaisons possibles de glycoformes pour chaque sous-unité de la FSH.

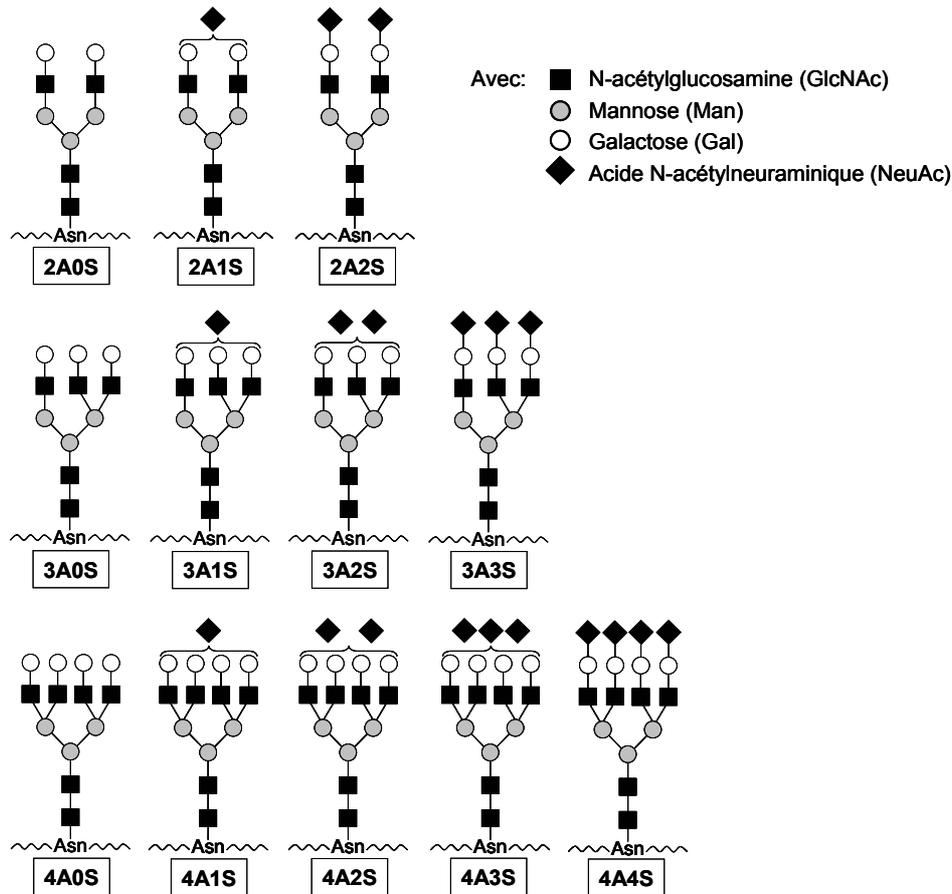


Figure 2 : Structures des N-glycanes majoritairement observés sur la sous-unité alpha de la FSH et nomenclature pour chacune d'elles (avec nAxS : n antennes et x acides N-acétylneuraminiques appelés aussi acides sialiques).

I.2. Rôle de la FSH et de ses glycosylations dans le cycle menstruel de la femme

Plusieurs études ont été effectuées pour montrer le rôle de la FSH glycosylée mais également le rôle de la nature des glycosylations sur chaque site.

Dans un premier temps, des études ont montré que **durant le cycle menstruel**, le **taux** de FSH ainsi que sa **composition en glycoformes varient**.

Le cycle menstruel de la femme se déroule suivant une série de 3 phases déclenchées par des hormones sécrétées par l'hypophyse (la FSH et la LH) et par les ovaires (l'œstrogène et la progestérone) (figure 3).

La première phase est la phase folliculaire : elle débute lorsque l'hypophyse libère la FSH qui va, une fois qu'elle atteint un taux seuil, stimuler la maturation des follicules. Chaque follicule contient un œuf. Lorsque le taux de FSH commence à diminuer, seul un de ces follicules continue à se développer. Il entraîne la production d'œstrogènes.

La deuxième phase est la phase ovulatoire : Lorsque le taux d'œstrogène a atteint une valeur maximale, il agit en retour sur l'hypophyse entraînant une augmentation considérable du taux de LH. Cela a pour effet d'entraîner la libération de l'œuf, l'ovulation est déclenchée.

La troisième phase est la phase lutéale : Une fois le follicule rompu, il se referme et se transforme en ce qu'on appelle le corps jaune, qui sécrète à son tour de l'œstrogène et de la progestérone pour préparer l'endomètre à la nidation de l'œuf et arrêter la libération de FSH et LH.

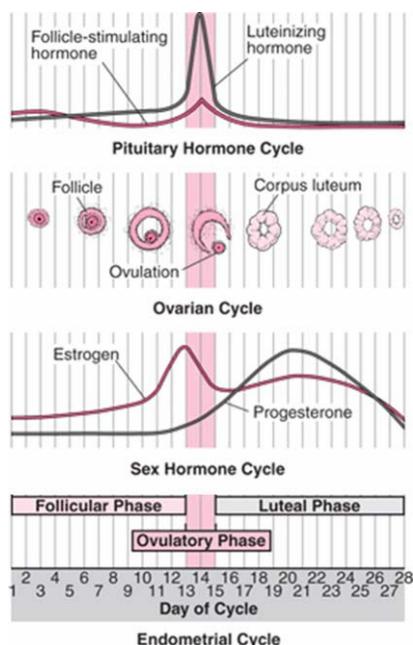


Figure 3 : Variation du taux des hormones (FSH, LH, Oestrogènes et Progestérone) au cours du cycle menstruel d'après le « Merck Manual Home Edition : Biology of the female reproductive system, Menstrual cycle » (<http://www.merck.com/mmhe/sec22/ch241/ch241e.html>).

La composition des glycanes de la FSH évolue durant ces phases. En effet, la FSH devient moins acide au milieu du cycle par rapport aux autres stades du cycle. Elle est à un point isoélectrique (pI) de 4,5 pendant la phase folliculaire et lutéale, alors que pendant la phase pré-ovulatoire elle est à un pI de 6,23-7,13. Les glycoformes de la FSH perdraient donc des acides sialiques au niveau des sites de glycosylation au cours du cycle [Zambrano et al., 1995], [Anobile et al., 1998], [Ulloa-Aguirre et al., 1999].

Enfin, il a été montré que les **sous-unités alpha et beta** de la FSH jouent un **rôle crucial** dans la **transduction du signal** et la **fixation à son récepteur**. En effet, des études ont montré qu'une **déglycosylation de l'Asn⁷⁸** de la sous-unité alpha entraînait une **augmentation de 72% de l'affinité de la FSH pour son récepteur**. De même, en **déglycosylant l'Asn⁵²** de la sous-unité alpha, il a pu être montré qu'il y avait un **effet sur la transduction du signal**, puisque **l'effet de l'hormone était diminué à 26%** [Bishop et al., 1994].

En résumé, l'ensemble de ces études a permis de montrer que **selon la nature des N-glycanes** en position **Asn⁵²** (antennarité et degré de sialylation), le **signal biologique** délivré par l'hormone est **différent**. Six jours après ovulation, elle envoie un signal (voie de signalisation 1) permettant la nidation. Puis elle envoie un autre signal (voie de signalisation 2) pour stopper toutes nidations

supplémentaires tout en maintenant la première, ceci afin d'éviter des fécondations multiples (figure 4).

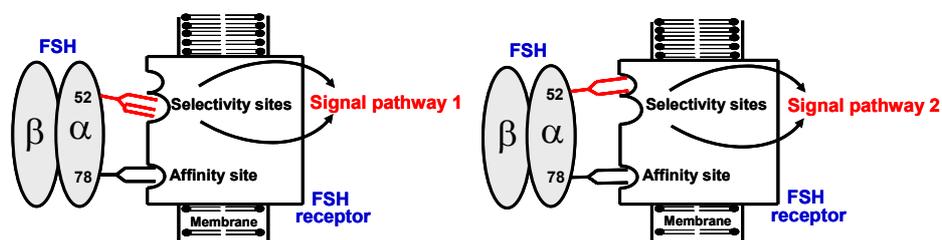


Figure 4 : Représentation schématique de l'effet de la nature des N-glycanes en position Asn⁵² de la sous-unité alpha de la FSH. Selon leur nature la voie de signalisation sera différente.

I.3. Contexte de l'étude et choix de la stratégie

Nous avons pu voir que durant le cycle menstruel la nature des glycoformes varie pour ne pas avoir les mêmes effets sur la fertilité. Une baisse ou un déséquilibre des taux de FSH peut conduire à une absence d'ovulation, et donc à une infertilité féminine. Pour combattre celle-ci, des traitements à base d'injection de FSH recombinante humaine (r-hFSH) ont été élaborés [Loumaye et al., 1998].

Cependant, ils ne prennent pas en compte la distribution fine des glycoformes. Or pour présenter un traitement plus efficace et mieux adapté, il est indispensable de disposer de plusieurs lots de r-hFSH dont la composition en glycoformes soit similaire aux taux et à la nature de la FSH naturelle lors du cycle d'une femme fertile. La chaîne beta de la FSH ne jouant qu'un rôle de stabilisation, nous ne nous sommes intéressés qu'à l'étude de la sous-unité alpha.

Dans ce but, nous avons travaillé à la mise au point de méthodes d'analyses rapides et reproductibles pour étudier l'hétérogénéité des glycoformes de la chaîne alpha de la r-hFSH. Celle-ci ayant deux sites de glycosylation occupés (Asn⁵² et Asn⁷⁸) deux études complémentaires ont été menées :

- **Analyse directe par spectrométrie de masse de la FSH entière** pour étudier l'hétérogénéité globale en termes de couples de glycanes donnant chacun une glycoforme.
- **Analyse LC-MS des glycopeptides de digestion de la FSH** pour étudier la micro-hétérogénéité sur chaque site glycosylé (Asn⁵² et Asn⁷⁸).

L'optimisation de ces deux méthodes a été effectuée sur un échantillon de référence de r-hFSH de composition connue en glycoformes. Le taux et la nature de la FSH variant durant le cycle menstruel, l'objectif ultime est d'appliquer ces méthodes à différents lots de r-hFSH purifiés pour connaître leur composition en glycoformes et de mesurer les effets de ces lots durant le cycle de femmes infertiles.

Nous présenterons dans un premier temps les résultats obtenus sur un lot de r-hFSH de référence de composition connue qui nous ont permis de mettre au point une méthode de quantification relative des glycoformes sur la protéine entière et sur chaque site de glycosylation.

Dans un second temps, nous présenterons les premiers résultats obtenus sur des lots de r-hFSH dont la composition en glycoformes varie.

II. Résultats : optimisations méthodologiques sur un lot de r-hFSH de référence

II.1. Etude de la r-hFSH entière de référence par spectrométrie de masse

II.1.1. Méthodologie

L'objectif de ce travail est d'étudier l'hétérogénéité globale des glycoformes de manière précise et reproductible. Pour cela plusieurs paramètres sont à prendre en compte tels que la préparation d'échantillon, le choix du spectromètre de masse en fonction de sa résolution et de sa sensibilité, l'optimisation des paramètres de celui-ci et la reproductibilité des analyses.

- **Préparation des échantillons :** Afin d'éviter toute manipulation des échantillons pouvant entraîner des pertes et fausser les résultats de quantification relative, les sous-unités alpha et beta de la r-hFSH ne sont pas traitées pour réduire et alkyliser les ponts disulfures. Les extraits de r-hFSH sont aliquotés en fractions contenant chacune 8 nmol, évaporés à sec et repris dans 200 μ L d'un mélange H₂O/ACN (50/50) + 1% d'acide formique. La concentration finale est de 40 pmol/ μ L pour chaque aliquot. Afin d'éviter toute dégradation des échantillons, les aliquots sont stockés à -80°C et les étapes d'évaporation et de resolubilisation ne sont effectuées qu'au dernier moment avant analyse.
- **Choix du spectromètre de masse :** L'analyse MS par ionisation MALDI de N-glycoprotéines dépendant de nombreux paramètres (par exemple : choix de la matrice, qualité du dépôt, l'utilisateur), ce spectromètre de masse n'est pas le plus adapté pour des analyses reproductibles. De ce fait, notre choix s'est porté sur l'analyse ESI-MS en infusion. Deux instruments ont été testés : un ESI-TOF, le micrOTOF de Bruker Daltonics et un ESI-IT, le HCTultra de Bruker Daltonics. Après plusieurs essais, notre choix s'est porté sur le micrOTOF. Certes ce spectromètre de masse nécessite une calibration journalière, mais sa résolution étant supérieure à celle d'une ESI-IT, il offre une précision de mesure de masse plus élevée (5ppm contre 100ppm).
- **Analyse ESI-MS de la r-hFSH entière sur le micrOTOF**
Les échantillons sont infusés à une concentration de 40 pmol/ μ L à un débit de 3 μ L/min en utilisant un pousse-seringue (Cole Parmer). L'ionisation des glycoformes de la FSH générant des ions dans les gammes de m/z 1600-2000, les tensions au niveau des lentilles et des octopôles sont ajustées de manière à favoriser la transmission des ions de m/z 1800. Avant chaque série d'analyse, l'instrument est calibré. La calibration s'effectue par infusion d'un échantillon standard de myoglobine à une concentration de 200 fmol/ μ L avec la même méthode d'analyse et le même débit d'infusion. La calibration est validée lorsque l'erreur sur

la masse moyenne de la protéine n'excède pas 3 ppm. De même la sensibilité est vérifiée et acceptée lorsque l'ion multichargé de m/z 942,76 a une intensité supérieure à 10000, que sa largeur à mi-hauteur se situe entre 0,4 et 0,6 m/z, que sa résolution est supérieure à 1800 et que le rapport signal sur bruit (S/N) est supérieur à 130.

➤ **Validation des résultats de quantification des glycoformes de la r-hFSH**

Afin de valider les résultats de quantification, il est nécessaire d'effectuer des études de reproductibilité et de répétabilité. Pour cela nous avons procédé de la manière suivante :

- Pour toutes les analyses nous utilisons la même méthode ESI-MS dans les mêmes conditions (solvant de solubilisation, débit d'infusion, même concentrations injectées, ...).
- Les séries d'échantillons de r-hFSH sont injectées plusieurs fois un même jour mais également des jours différents.
- Avant chaque série d'analyses, nous injectons un échantillon standard de myoglobine à une concentration de 200 fmol/μL avec la même méthode d'analyse et le même débit d'infusion. Ceci nous permet de vérifier les performances du système.
- Une fois les séries d'analyses terminées, nous injectons une seconde fois l'échantillon standard de myoglobine toujours à la même concentration et dans les mêmes conditions. Ceci nous permet de vérifier la stabilité du système et qu'il n'y a pas eu de variations dans la calibration de masse au cours du temps.
- Enfin, entre chaque analyse de r-hFSH, nous rinçons trois fois le système ESI-MS avec le solvant de solubilisation pour s'assurer qu'il n'y a plus de traces de la r-hFSH qui risqueraient de fausser les résultats de quantification.

II.1.2. Résultats : Analyse ESI-MS de la r-hFSH entière

a) *Masses théoriques des glycoformes de la sous-unité alpha*

Des études préliminaires ayant permis de montrer que les N-glycanes majoritaires sont de type biantennés (B), triantennés (T) et tetraantennés (Te) avec des degrés de sialylation variant de 0S à totalement sialylés (6S), les masses moyennes théoriques de chacun des couples possibles de N-glycanes sur la sous-unité alpha donnant des glycoformes sont calculées de la manière suivante :

Par exemple, pour la glycoforme correspondant au couple B/B 4S le calcul de la masse moyenne est

$$\begin{array}{cccccc}
 \text{B/B 4S} = & \underbrace{10195,742}_{\substack{\text{masse de la} \\ \text{chaîne } \alpha \text{ de} \\ \text{la FSH}}} & + & 2 \times & \underbrace{(4 \times 203,195)}_{\substack{\text{masse de} \\ 4 \text{ GlcNAc}}} & + & \underbrace{3 \times 162,142}_{\substack{\text{masse de} \\ 3 \text{ Man}}} & + & \underbrace{2 \times 162,142}_{\substack{\text{masse de} \\ 2 \text{ Gal}}} & + & \underbrace{2 \times 291,258}_{\substack{\text{masse de} \\ 2 \text{ NeuAc}}} & = & 14607,754 \text{ Da}
 \end{array}$$

L'ensemble des masses de chaque glycoforme ainsi que leur rapport m/z pour un état de charge 8+ sont présentés dans le tableau 1.

Glycoformes	Masses théoriques (Da)	Rapport m/z des ions (8+)
B/B 4S	14607,754	1826,98
B/B 3S	14316,496	1790,57
B/B 2S	14025,238	1754,16
B/B 1S	13733,980	1717,76
B/B 0S	13442,722	1681,35
B/T 5S	15264,349	1909,05
B/T 4S	14973,091	1872,64
B/T 3S	14681,833	1836,24
B/T 2S	14390,575	1799,83
B/T 1S	14099,317	1763,42
B/T 0S	13808,059	1727,02
T/T 6S ou B/Te 6S	15920,944	1991,13
T/T 5S ou B/Te 5S	15629,686	1954,72
T/T 4S ou B/Te 4S	15338,428	1918,31
T/T 3S ou B/Te 3S	15047,170	1881,90
T/T 2S ou B/Te 2S	14755,912	1845,50
T/T 1S ou B/Te 1S	14464,654	1809,09
T/T 0S ou B/Te 0S	14173,396	1772,68

Tableau 1 : Masses théoriques et valeur du rapport m/z pour un état de charge 8+ de chacune des glycoformes de la sous-unité alpha de la r-hFSH.

b) Analyses par ESI-MS

La figure 5 représente le spectre de masse obtenu en infusant la r-hFSH à une concentration de 40 pmol/μL sur un spectromètre de masse ESI-TOF, le microTOF de Bruker Daltonics. Ce spectre de masse présente une forte hétérogénéité. Cela s'explique par le fait qu'il s'agit d'une protéine composée de 2 sous-unités (alpha et beta), chaque sous-unité contenant deux sites de glycosylations occupés et chaque site pouvant être occupé par différents N-glycanes. Son interprétation manuelle n'est donc pas aisée.

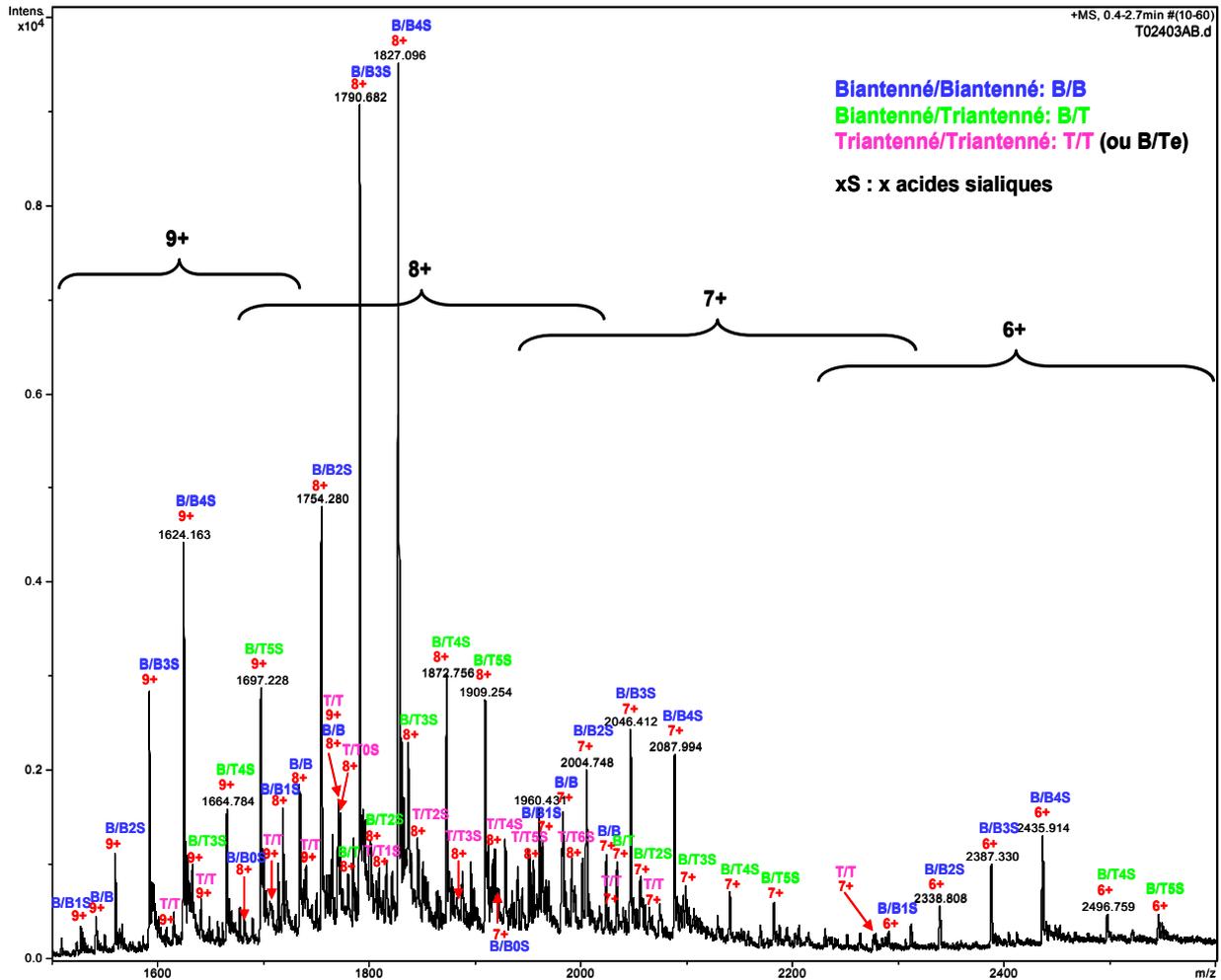


Figure 5 : Spectre ESI-MS de la r-hFSH entière obtenue sur un instrument microTOF de Bruker Daltonics. Les ions correspondants aux espèces majoritaires sont annotés et l'état de charge de chaque ion est également affiché. L'ensemble des ions n'ont été attribués qu'à la sous-unité alpha.

L'ionisation par source ESI de la r-hFSH entière génère des ions dans les gammes de m/z allant de 1500 à 2600 qui correspondent à 4 états de charges (9+, 8+, 7+ et 6+). Le fait que nous n'observons pas d'états de charges plus élevés peut être dû aux ponts disulfures puisque la r-hFSH n'a pas été traitée pour réduire et alkyliser les ponts disulfures. En effet, la chaîne alpha contient 10 cystéines qui forment 5 ponts disulfures. La chaîne beta, elle, contient 6 ponts disulfures.

Le microTOF nous a permis de mesurer les masses des différents glycoformes avec une précision de $\pm 0,2 m/z$, ce qui équivaut à un écart de ± 1 à 2 Da sur la masse moyenne de la protéine glycosylée. De ce fait l'attribution d'un ion à une glycoforme s'est faite en comparant les masses théoriques aux masses mesurées.

La majorité des ions a été attribuée à des glycoformes de la sous-unité alpha. Aucun ion n'a été attribué à la sous-unité beta. Des études préliminaires ont montré que cette sous-unité s'ionise très mal par rapport à la sous-unité alpha [Gervais et al., 2003]. De ce fait, son signal semble supprimé par celui généré par les glycoformes de la sous-unité alpha. Néanmoins, l'objectif de cette étude étant

d'étudier l'hétérogénéité des glycoformes de la sous-unité alpha, ceci est un avantage, évitant un spectre de masse encore plus complexe.

Pour chaque état de charge, les trois types de couples majoritairement observés formant les glycoformes sont (1) biantenné/biantenné (B/B), (2) biantenné/triantenné (B/T) et (3) triantenné/triantenné (T/T) avec des degrés de sialylation qui varient allant de 0 acides sialiques (0S) à totalement sialylés (6S). Concernant le couple T/T, il est impossible de le différencier du couple biantenné/tetraantenné (B/Te). En effet, ces deux couples ont une composition en monosaccharides identique (10 GlcNAc, 6 Man, 6 Gal, 6 NeuAc). De ce fait la quantification relative de ces couples peut être faussée. Néanmoins les N-glycanes tetraantennés sont très faiblement représentés par rapport aux autres. En effet, aucun couple triantenné /tetraantenné (T/Te) ou tetraantenné/tetraantenné (Te/Te) n'est observé.

Quant aux ions minoritaires, ils ont pu être attribués à différentes glycoformes correspondant à des pertes successives de galactoses et de N-glucosamines. Les paramètres du spectromètre de masse ayant été optimisés pour éviter toute fragmentation de N-glycanes en source, ces différentes glycoformes proviennent des étapes de production de cette hormone recombinante. En effet, la r-hFSH est produite dans des cellules d'ovaires d'hamster chinois (CHO), et ces glycoformes sont les produits obtenus durant les différents étapes de la biosynthèse des N-glycanes [Gomord and Faye, 2004], [Loumaye et al., 1998]. L'ensemble représente plus de 80 différentes glycoformes possibles.

La quantification relatives des glycoformes n'ayant été effectuées que sur les espèces majoritaires, l'ensemble de toutes les glycoformes ne seront pas présentées ici, ni traitées dans la suite des expériences.

c) Quantification relative des glycoformes de la sous-unité alpha de la r-hFSH entière

Pour chacun des états de charges, les rapports d'intensité entre chaque glycoforme majeure étant proches, nous avons choisi de les quantifier directement sur le spectre de masse et sur un état de charge particulier, celui pour lequel les ions sont les plus intenses, l'état de charge 8+. La quantification se fait de la manière suivante :

- 1) Les spectres de masses sont combinés de la 0,5 à la 0,9 minute du temps d'infusion (on somme environ 6 spectres de masse).
- 2) Nous effectuons un zoom correspondant à la zone des ions multichargés 8+ (1600 à 2000 m/z).
- 3) Nous mesurons manuellement l'intensité de chaque ion correspondant à une glycoforme.
- 4) Le pourcentage relatif de chaque glycoforme est calculé selon la formule suivante:

$$\% \text{ de chaque glycoforme} = \frac{\text{Intensité de l'ion } 8+ \text{ correspondant à une glycoforme}}{\Sigma \text{ de toutes les intensités des ions } 8+ \text{ correspondant aux glycoformes majeures}} \times 100$$

Comme nous avons pu le voir dans la partie introductive (voir chapitre « I.2. Rôle de la FSH et de ces glycosylations au cours du cycle menstruel de la femme »), deux paramètres semblent jouer un rôle important durant le cycle menstruel, l'antennarité et le degré de sialylation.

De ce fait la quantification relative s'est faite à ces deux niveaux. Les résultats obtenus pour une analyse ESI-MS sont présentés dans la figure 6.

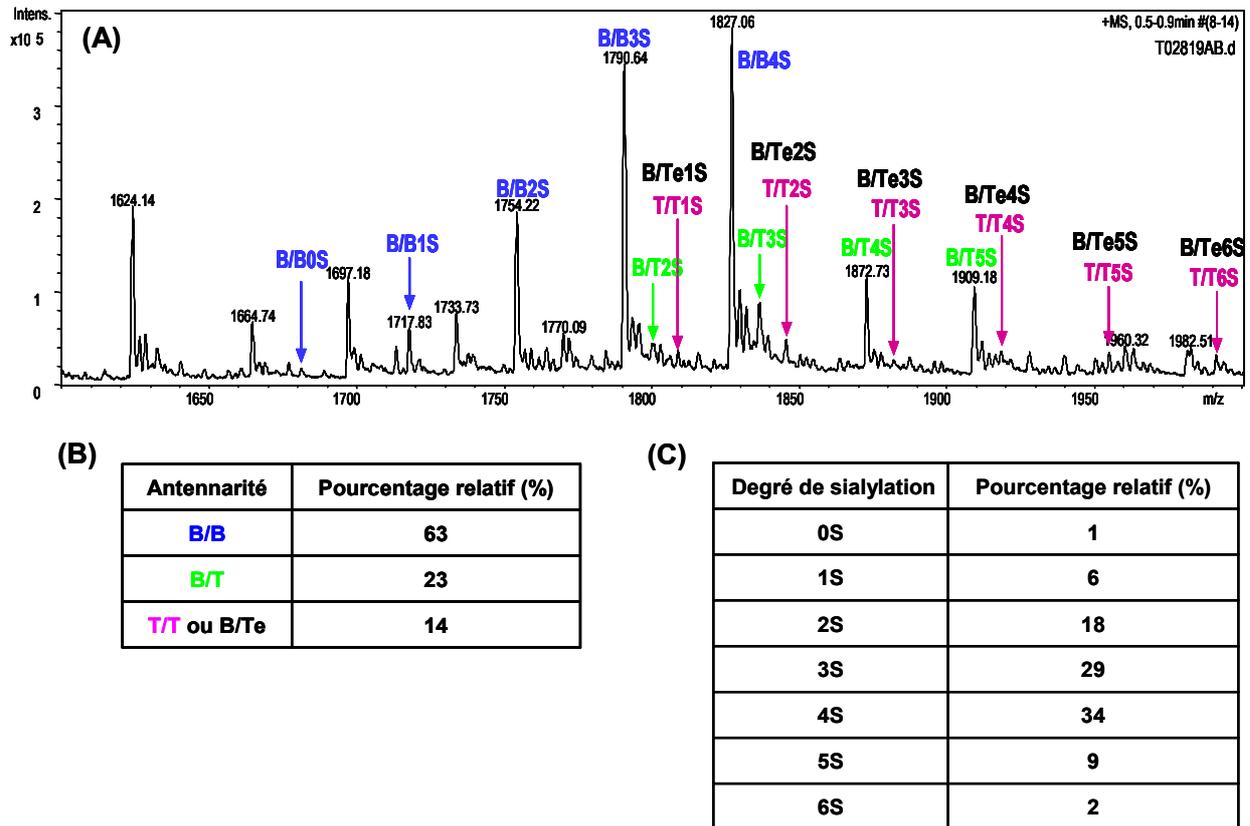


Figure 6 : Résultats de quantification relative obtenus pour une analyse ESI-MS.

(A) Zoom dans la zone correspondant aux ions multichargés 8+ des spectres combinés entre 0,5 et 0,9 minutes.

(B) Quantification relative des glycoformes selon l'antennarité, tout degré de sialylation confondu.

(C) Quantification relative des glycoformes selon le degré de sialylation, toute antennarité confondu.

Les résultats montrent que les glycoformes de la sous-unité alpha les plus abondants sont ceux ayant chaque site de glycosylation occupé par un N-glycane biantenné. Concernant le degré de sialylation, les glycoformes contenant 2, 3 et 4 acides sialiques sont les espèces majoritaires.

Afin de vérifier la qualité des résultats et d'évaluer la stabilité des lots de r-hFSH, le même échantillon de r-hFSH de référence est analysé plusieurs fois sur plusieurs jours dans les mêmes conditions. Les résultats obtenus pour différentes analyses sont présentés dans la figure 7.

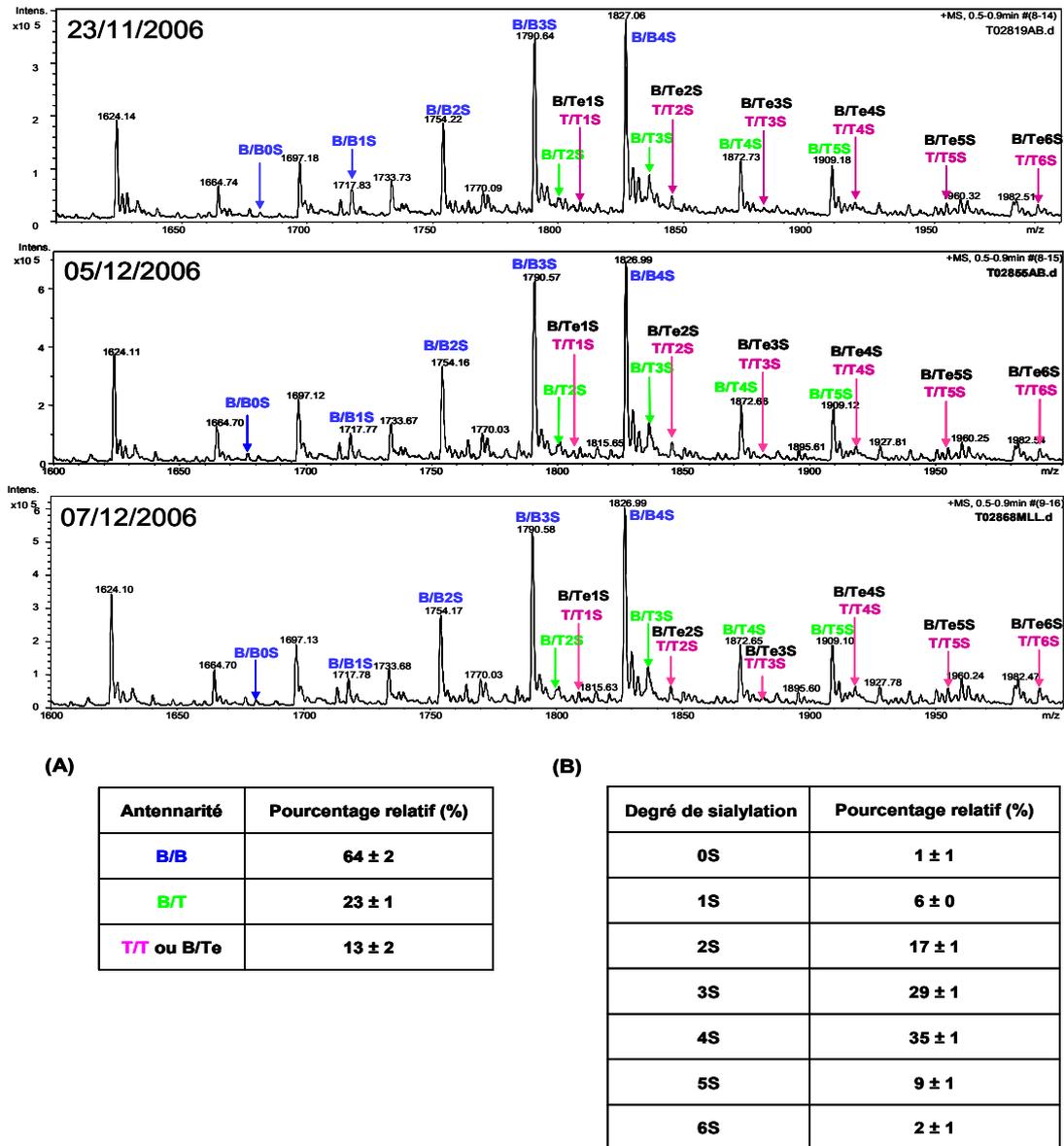


Figure 7 : Comparaison des profils de 3 spectres de masses de la r-hFSH mesurés à différentes dates (zoom dans la zone des ions 8+) et valeurs moyennes des résultats de quantification relatives selon (A) l'antennarité tout degré de sialylation confondu et (B) le degré de sialylation toute antennarité confondu.

Dans un premier temps on constate que les profils des spectres de masse sont reproductibles. De même, les résultats de quantification relative sont reproductibles.

En effet, les variations d'un échantillon à un autre n'est que de $\pm 2\%$ pour l'antennarité et de $\pm 1\%$ pour le degré de sialylation. Ces résultats sont tout à fait acceptables dans la mesure où il s'agit d'une quantification relative sur les espèces majoritaires de r-hFSH.

En effet, l'objectif était d'étudier l'hétérogénéité globale de la FSH. Ces expériences nous ont permis de mettre au point une méthode d'analyse par ESI-MS rapide pour identifier l'ensemble des glycoformes de r-hFSH entière. Nous connaissons donc maintenant les couples de N-glycanes majoritaires de la sous-unité alpha ainsi que leur abondance relative.

Néanmoins dans le cas du couple T/T il est impossible de le différencier du couple B/Te. De plus nous avons de nombreuses glycoformes qui correspondent à des formes immatures qui proviennent du processus de biosynthèse de ces glycoformes.

De ce fait l'interprétation n'est pas aisée et une quantification absolue n'est pas possible sans ajout d'étalons internes et sans une meilleure résolution et sensibilité permettant de différencier précisément les différentes glycoformes. Toutefois cette méthode permet d'étudier l'hétérogénéité globale des glycoformes et de les quantifier de manière relative.

II.2. Etudes des glycopeptides de digestion de la r-hFSH de référence par LC-MS

II.2.1. Méthodologie

L'objectif de cette étude est d'étudier la micro-hétérogénéité sur chaque site de glycosylation de la sous-unité alpha, qui sont l'Asn⁵² et l'Asn⁷⁸, de manière précise et reproductible. La stratégie choisie pour l'étude des glycoformes des glycopeptides est un traitement enzymatique, suivi d'un couplage LC-MS.

Néanmoins, plusieurs paramètres sont à prendre en compte tels que la préparation d'échantillon, l'optimisation des paramètres de digestion, le choix du système de chromatographie liquide et du spectromètre de masse, l'optimisation des paramètres de celui-ci et la reproductibilité des analyses.

- **Digestion à la trypsine de la r-hFSH :** Dans un premier temps différents types d'enzymes et de protocoles de digestion ont été testés. Finalement notre choix s'est porté sur la trypsine qui permet non seulement de digérer la protéine totalement mais qui permet aussi de générer des peptides de taille raisonnable. 8 nmol de r-hFSH sont évaporés à sec puis repris dans 40 µL de Trifluoroethanol (TFE) et 40 µL de 10mM DL-dithiothréitol (DTT) solubilisé dans 50mM de NH₄HCO₃ afin de réduire les ponts disulfures. Le temps d'incubation est de 1 heure à 65°C. Les fonctions thiol sont ensuite alkylés par ajout de 40µL d'iodoacetamide (IAM) 55mM préparé dans 25mM de NH₄HCO₃. L'alkylation se fait en 1 heure dans le noir. Avant digestion, le pourcentage de TFE est réduit à 5% en diluant l'échantillon dans 680µL de NH₄HCO₃ 25mM. Finalement, on ajoute 10µg de trypsine solubilisée dans 200µL de NH₄HCO₃ 25mM. Après une nuit à 37°C, la réaction de digestion est stoppée en ajoutant 1µL d'acide formique. La concentration finale des échantillons est de 8 pmol/µL. Afin de vérifier si la digestion est complète, les échantillons sont analysés en MALDI-TOF. La matrice utilisée est de l'acide α -cyano-4-hydroxycinnamique. Le mode de dépôt utilisée est le dépôt en sandwich.
- **Choix du système LC-MS :** Plusieurs systèmes LC-MS ainsi que différents types de colonnes ou encore différents gradients d'élution ont été testés. Notre choix s'est porté sur la microHPLC 1100 d'Agilent Technologies couplée à une trappe ionique, l'HCTultra de Bruker

Daltonics. Ce système est compatible avec les concentrations d'échantillons nécessaires à la détection des N-glycopeptides mais montre également une bonne reproductibilité. Il constitue donc un bon compromis entre un système LC et un système nanoLC en termes de quantité de matériel et de sensibilité. L'objectif étant d'avoir un profil global des N-glycopeptides, les conditions chromatographiques ont été optimisées pour séparer l'ensemble des glycopeptides des autres peptides.

- **Analyse LC-MS du digeste de r-hFSH:** 112 pmol du digeste de r-hFSH sont injectées, dessalées sur une pré-colonne Zorbax 300SB-C18, 5mm x 300µm, 3,5µm (Agilent Technologies) puis élués en sens inverse vers une colonne Zorbax 300SB-C18, 150 mm x 300 µm, 3,5µm (Agilent Technologies). Ils sont ensuite séparés par un gradient allant de 10% de solvant B à 26% de solvant B en 19 minutes puis de 26% de B à 70% de B en 5 minutes à une débit de 4µL/min, avec A = H₂O + 0,1% acide formique et B= ACN + 0,1% acide formique. La tension appliquée sur la contre électrode est fixée à 4000V. Afin d'éviter toute fragmentation des glycopeptides en source, les paramètres d'ionisation ont été optimisés. De même les glycopeptides ionisés ayant des rapports m/z élevés, les tensions ont été optimisés pour favoriser la transmission des ions de m/z 1100.
- **Validation de la méthode de quantification**
Afin de valider les résultats de quantification, il est nécessaire d'effectuer des études de reproductibilité et de répétabilité. Pour cela nous avons procédé de la manière suivante :
 - Pour toutes les analyses LC-MS, nous utilisons la même méthode LC-MS dans les mêmes conditions (volume injecté, gradient chromatographique, même type de colonnes, ...).
 - Les différents mélanges de peptides de la r-hFSH sont analysés par séries de trois plusieurs fois le même jour mais également à des jours différents.
 - Avant chaque série d'analyses, nous injectons 1 pmol d'un échantillon standard de BSA digérée à la trypsine en utilisant la même méthode LC-MS. Ceci nous permet de vérifier les performances du système.
 - Une fois les séries d'analyses terminées, nous analysons une seconde fois 1 pmol de l'échantillon standard de BSA digérée à la trypsine toujours dans les mêmes conditions. Ceci nous permet de vérifier la stabilité du système et qu'il n'y a pas eu de variations dans la calibration et dans la résolution chromatographique au cours du temps.
 - Enfin, entre chaque analyse LC-MS des mélanges de peptides de la r-hFSH, nous effectuons un blanc de colonne pour s'assurer qu'il n'y a plus de traces de glycopeptides de la r-hFSH, ceux-ci pouvant fausser les résultats de quantification.

II.2.2. Résultats : Analyse LC-MS du digeste de r-hFSH de référence

a) *Masses théoriques des N-glycopeptides de la sous-unité alpha*

La digestion à la trypsine permet de générer deux peptides de taille convenable contenant chacun un des sites de glycosylation. Nous les appellerons les glycopeptides N52 et N78. Leurs caractéristiques sont disponibles dans le tableau 2.

Peptides issus de la digestion à la trypsine	Position des peptides dans la séquence de la r-hFSH entière	Séquences (Asn glycosylée soulignée)	Masses monoisotopiques des peptides non-glycosylés après réduction et alkylation des cystéines (Da)
N52	52-63	<u>N</u> VTSESTCCVAK	1354,586
N78	76-91	VE <u>N</u> HTACHCSTCY <u>Y</u> HK	2065,819

Tableau 2 : Caractéristiques des peptides issus de la digestion à la trypsine de la chaîne alpha de la r-hFSH.

Le calcul des masses monoisotopiques des différents glycopeptides se fait de la même manière que pour la r-hFSH entière. L'ensemble des masses théoriques des espèces majoritaires déduites des résultats obtenus lors de l'analyse de la r-hFSH entière sont représentés dans le tableau 3.

		Masses (Da)	m/z (ion 2+)	m/z (ion 3+)	m/z (ion 4+)			Masses (Da)	m/z (ion 2+)	m/z (ion 3+)	m/z (ion 4+)
N52	2A0S	2979,0	1490,5	994,0	745,8	N78	2A0S	3690,8	1846,4	1231,3	923,7
	2A1S	3270,2	1636,1	1091,1	818,6		2A1S	3982,0	1992,0	1328,3	996,5
	2A2S	3561,5	1781,8	1188,2	891,4		2A2S	4273,3	2137,7	1425,4	1069,3
	3A0S	3344,3	1673,2	1115,8	837,1		3A0S	4056,1	2029,1	1353,0	1015,0
	3A1S	3635,6	1818,8	1212,9	909,9		3A1S	4347,3	2174,7	1450,1	1087,8
	3A2S	3926,8	1964,4	1309,9	982,7		3A2S	4638,6	2320,3	1547,2	1160,7
	3A3S	4218,1	2110,1	1407,0	1055,5		3A3S	4929,8	2465,9	1644,3	1233,5
	4A0S	3709,6	1855,8	1237,5	928,4		4A0S	4421,4	2211,7	1474,8	1106,4
	4A1S	4000,9	2001,5	1334,6	1001,2		4A1S	4712,7	2357,4	1571,9	1179,2
	4A2S	4292,2	2147,1	1431,7	1074,1		4A2S	5003,9	2503,0	1669,0	1252,0
	4A3S	4583,4	2292,7	1528,8	1146,9		4A3S	5295,2	2648,6	1766,1	1324,8
	4A4S	4874,7	2438,4	1625,9	1219,7		4A4S	5586,4	2794,2	1863,1	1397,6

Tableau 3 : Liste des masses théoriques des peptides N-glycosylés N52 et N78 selon la nature du N-glycane (2A0S, 2A1S, 2A2A, ...) et valeur des rapports m/z pour différents états de charges (2+, 3+ et 4+).

b) Analyse LC-MS

La figure 8 représente le chromatogramme obtenu après analyse LC-MS du mélange de peptides de digestion de la r-hFSH. En extrayant les courants d'ions correspondants aux différents N-glycopeptides dont les masses théoriques ont été calculées au préalable, on constate qu'ils sont tous élués en début de gradient dans un intervalle de 8 minutes (entre la 10^{ème} et la 18^{ème} minute). Cet intervalle correspond à un pourcentage d'acétonitrile de 16% à 20%. Durant cet intervalle de temps aucun peptide non-glycosylé n'est élué ce qui est un avantage pour une quantification relative puisqu'il n'y a pas d'effets de suppression de signal.

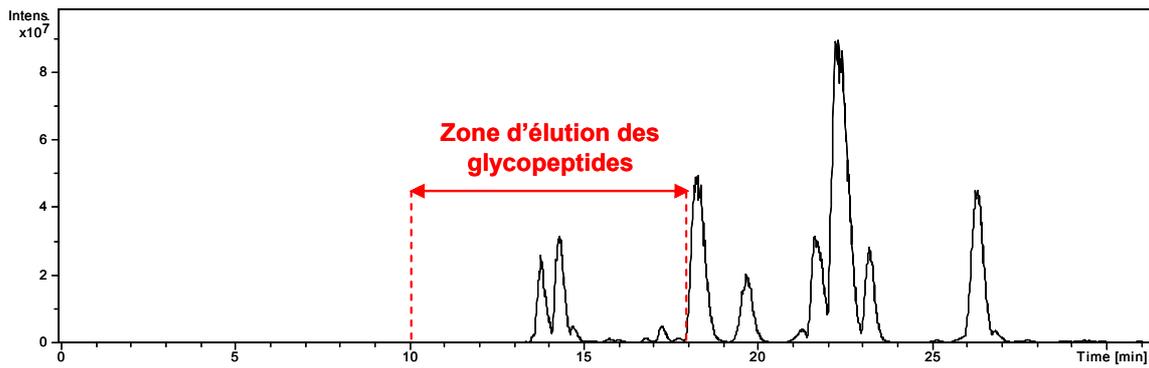


Figure 8 : Chromatogramme LC-MS du mélange des peptides issus de la digestion de r-hFSH montrant la zone d'éluion des glycopeptides de la sous-unité alpha.

En combinant les spectres de masse dans cette zone d'éluion, on obtient un profil MS global de tous les glycopeptides de la sous-unité alpha. Comme dans le cas de l'analyse ESI-MS de la r-hFSH entière aucun glycopeptide correspondant à la sous-unité beta n'est observé. On constate que les glycopeptides N52 sont préférentiellement triplement chargés. Les glycopeptides N78, eux, sont préférentiellement quatre fois chargés. Ceci est dû à la nature des acides aminés de ces deux peptides.

La figure 9 représente le spectre combiné dans la zone d'éluion des glycopeptides et l'attribution des différents ions aux peptides correspondant à la sous-unité alpha de la r-hFSH.

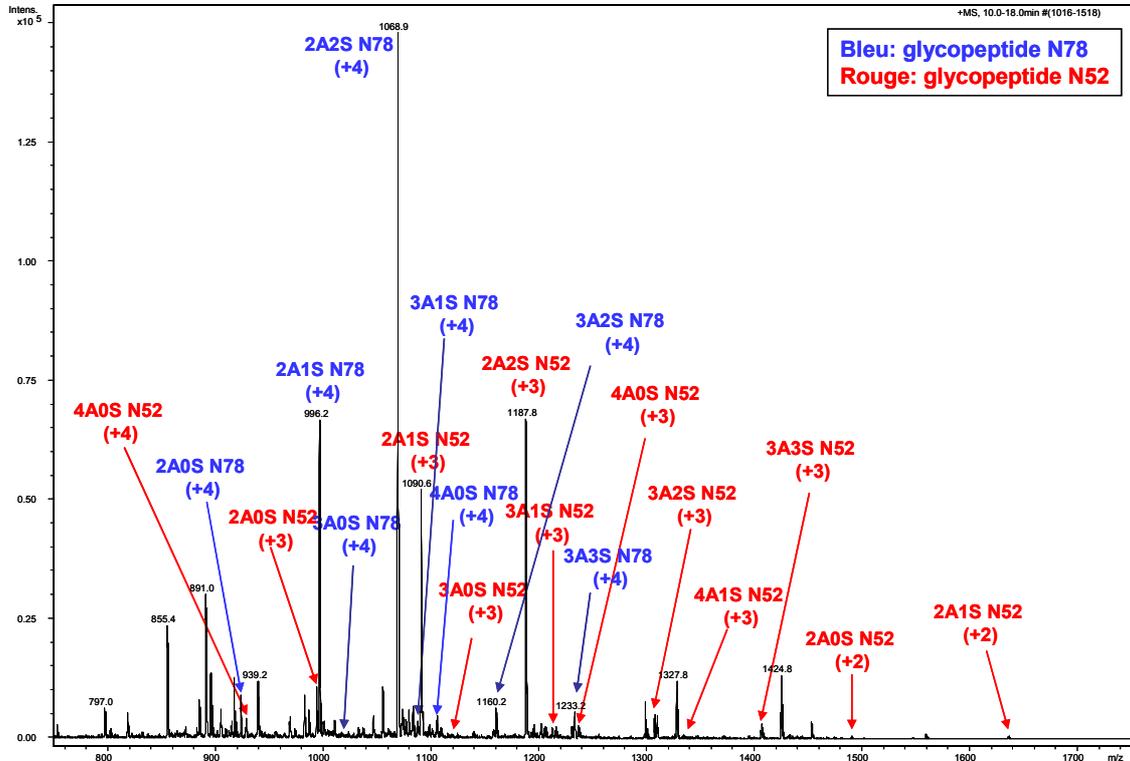


Figure 9 : Profil MS des glycopeptides de la sous-unité alpha. Somme des spectres de masse obtenus dans la zone d'éluion des glycopeptides.

c) Quantification des glycopeptides de la sous-unité alpha de la r-hFSH

La quantification est faite, pour chacun des types de glycopeptides, sur les états de charge les plus abondants. Pour les glycopeptides N52 on utilisera les ions multichargés (3+) et pour les glycopeptides N78, on utilisera les ions multichargés (4+). Pour la quantification, on admet que le pouvoir ionisant de tous les glycopeptides est identique. La quantification est réalisée de la manière suivante :

- 1) On combine les spectres de masse de la 10^{ème} à la 18^{ème} minute pour avoir dans un même spectre tous les ions correspondants aux différents glycopeptides (N52 et N78 confondus).
- 2) On mesure l'intensité de tous les ions. On considère que toutes les glycoformes contenant une des asparagines glycosylées (N52 ou N78) représentent 100%.
- 3) Le pourcentage relatif de chaque glycopeptide est mesuré de la manière suivante :

$$\text{Abondance d'un glycopeptide N52 (\%)} = \frac{\text{Intensité de l'ion 3+ correspondant à un glycopeptide N52}}{\Sigma \text{ de toutes les intensités des ions 3+ correspondant à tous les glycopeptides N52}} \times 100$$

$$\text{Abondance d'un glycopeptide N78 (\%)} = \frac{\text{Intensité de l'ion 4+ correspondant à un glycopeptide N78}}{\Sigma \text{ de toutes les intensités des ions 4+ correspondant à tous les glycopeptides N78}} \times 100$$

De la même façon que pour l'étude de la protéine entière, la quantification relative se fait au niveau de l'antennarité et du degré de sialylation.

Pour une analyse donnée, les résultats obtenus pour la quantification des N-glycopeptides de la chaîne alpha de la r-hFSH sont présentés dans les tableaux 4.

(A)	Pourcentage relatif (%)			(B)	Pourcentage relatif (%)				
	2A	3A	4A		0S	1S	2S	3S	4S
N52	89	8	3	N52	9	35	54	2	0
N78	91	7	2	N78	6	26	66	3	0

Tableaux 4 : Résultats de quantification relative pour chacun des glycopeptides N52 et N78 selon (A) l'antennarité et (B) le degré de sialylation. Ces résultats sont ceux obtenus pour une analyse.

Que ce soit en position Asn⁵² ou en position Asn⁷⁸, le N-glycane de type biantenné est l'espèce majoritaire. De même, dans le cas du degré de sialylation, ce sont les N-glycanes contenant un ou deux acides sialiques (1S et 2S) qui sont majoritaires. Ces résultats sont en accord avec ceux obtenus sur l'analyse ESI-MS de la r-hFSH entière où nous avons pu déterminer que les glycoformes majoritaires sont les couples biantenné/biantenné avec 2, 3 et 4 acides sialiques. Les N-glycanes tetraantennés ne représentent que 3 % soit deux à trois fois moins que les triantennés. Aucun N-glycane contenant 4 acides sialiques n'est observé sur les deux glycopeptides N52 et N78.

Afin de valider les résultats, le même échantillon de digeste de r-hFSH de référence est analysé plusieurs fois le même jour mais également différents jours dans les mêmes conditions.

Afin d'illustrer les résultats de quantification relative obtenus pour différentes analyses, j'ai choisi de vous présenter les profils MS de trois analyses effectuées le même jour (figure 10) ainsi que les valeurs moyennes des résultats de quantification obtenus à partir de trois analyses effectuées le même jour (figure 11) et effectuées trois jours différents (figure 12).

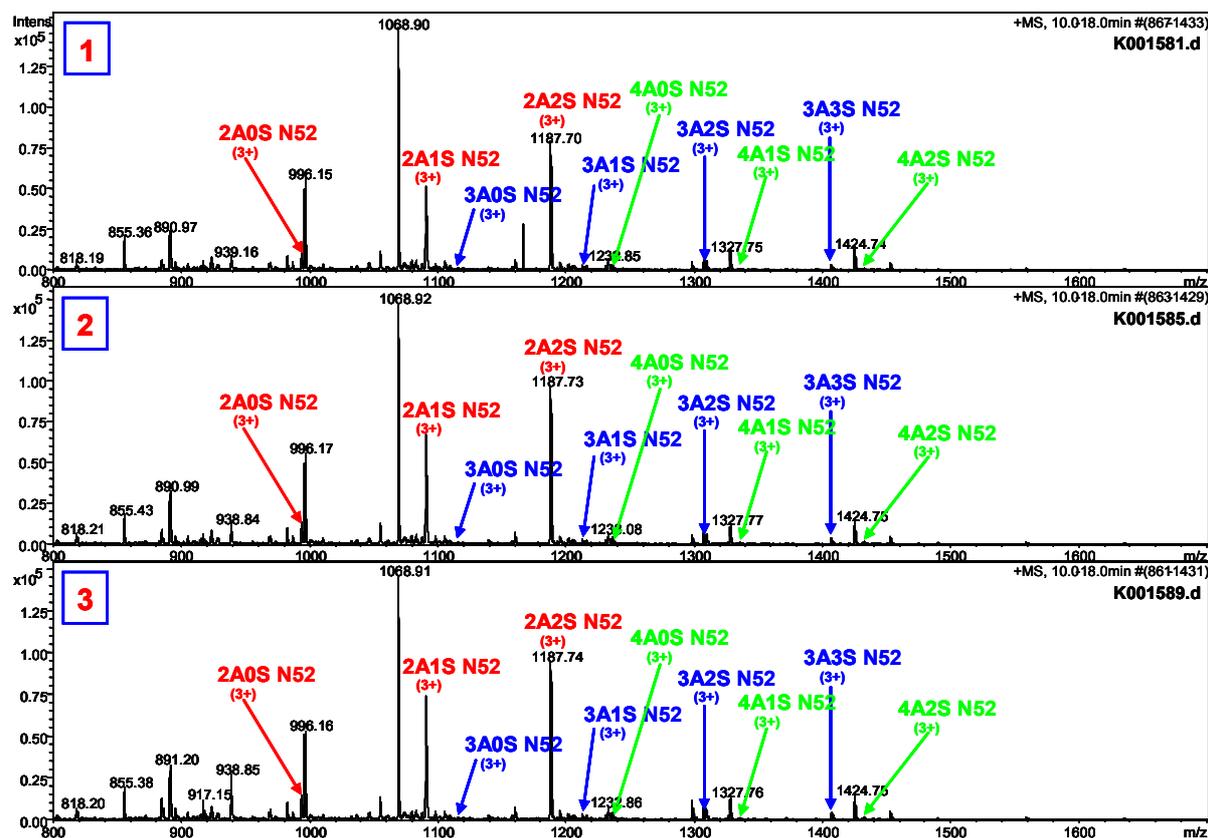


Figure 10 : Superposition des profils de trois spectres de masse correspondant aux spectres combinés dans les zones d'éluion des glycopeptides (10 à 18 minutes) de trois analyses LC-MS effectuées le même jour. Les échelles d'intensité et de m/z sont identiques. Les ions annotés correspondent aux glycopeptides N52.

En comparant les trois profils des spectres de masse issus de trois analyses LC-MS effectuées le même jour, on constate qu'ils sont identiques. En effet, on retrouve dans les trois spectres de masse les mêmes ions avec des intensités identiques pour chacun d'eux et de ce fait la même distribution.

De plus en termes de précision de mesure de masse, la variation d'une valeur à une autre est très faible. En effet, prenons par exemple l'ion de m/z 1187,70 qui correspond au glycopeptide N52 contenant le N-glycane biantenné avec deux acides sialiques, sa valeur ne varie que de $\pm 0,04$ m/z d'une analyse à une autre, ce qui équivaut pour un ion triplement chargé à une variation de $\pm 0,12$ Da sur la masse du peptide.

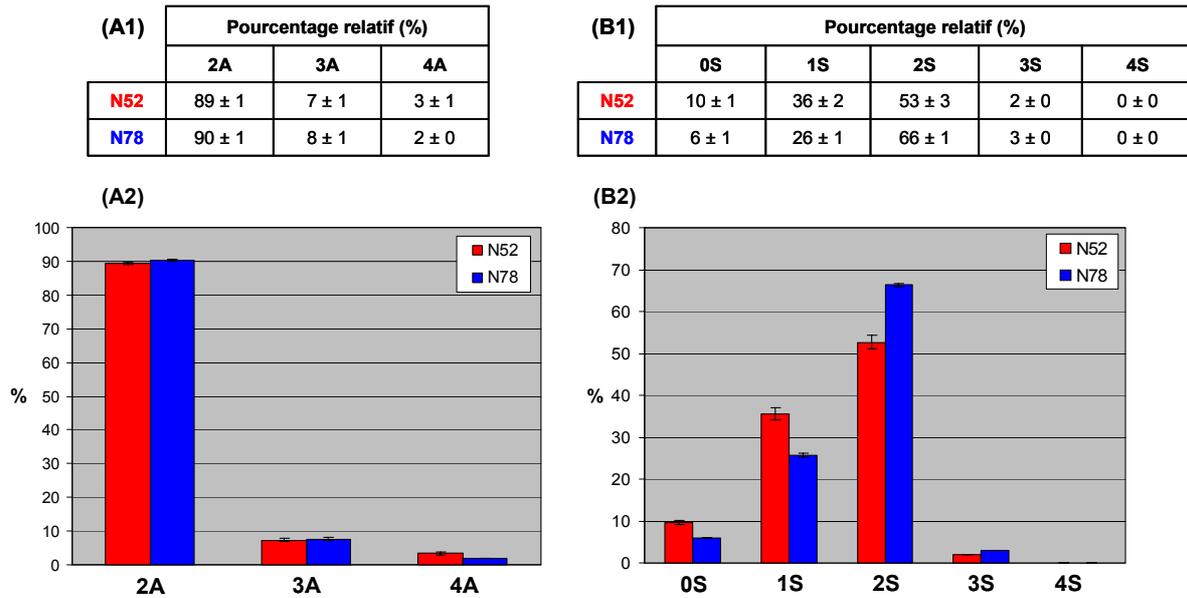


Figure 11 : Valeurs moyennes des résultats de quantification des glycopeptides N52 et N78 de la sous-unité alpha de la r-hFSH pour trois analyses effectuées le même jour selon (A1) l'antennarité et (B1) le degré de sialylation ainsi que la représentation graphique de ces valeurs moyennes selon (A2) l'antennarité et (B2) le degré de sialylation.

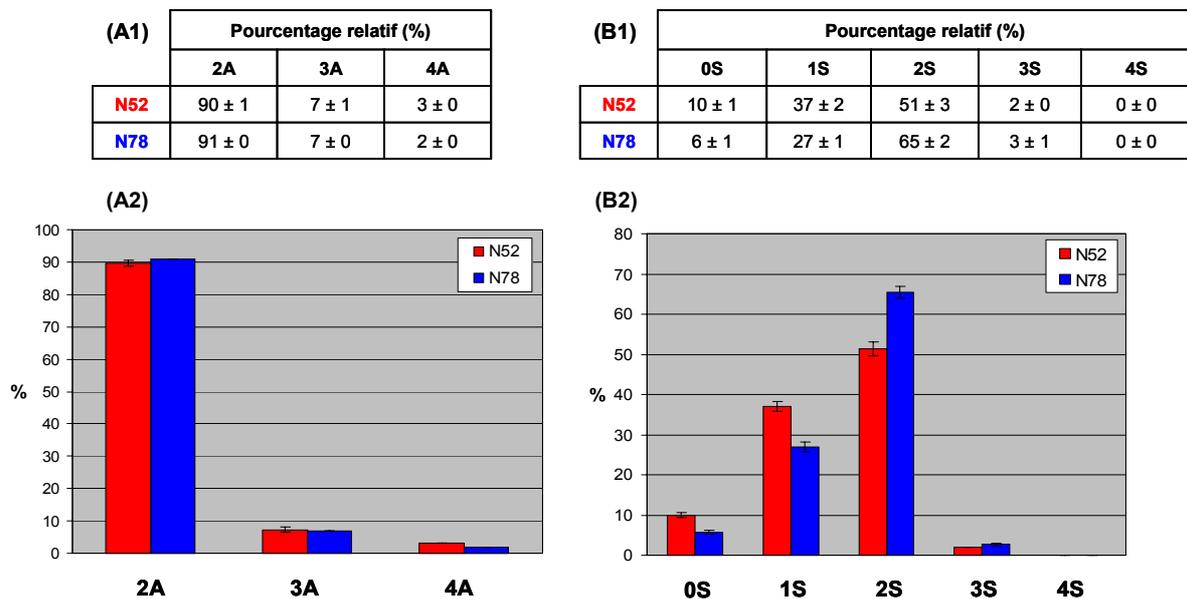


Figure 12 : Valeurs moyennes des résultats de quantification des glycopeptides N52 et N78 de la sous-unité alpha de la r-hFSH pour trois analyses effectuées trois jours différents selon (A1) l'antennarité et (B1) le degré de sialylation ainsi que la représentation graphique de ces valeurs moyennes selon (A2) l'antennarité et (B2) le degré de sialylation.

L'ensemble des résultats de quantification nous permet de constater que nous allons toujours dans le même sens en termes d'abondance relative de chaque espèce. En effet, pour toutes les analyses, les espèces majoritaires sont les biantennés avec deux acides sialiques. En terme d'antennarité qu'il

s'agisse du glycopeptides N52 ou N78, les glycoformes majeures sont les biantennés, de même pour le degré de sialylation, qu'il s'agisse des glycopeptides N52 ou N78, les glycoformes majeures ont deux acides sialiques avec néanmoins pour les glycopeptides N78 une abondance plus élevée (65 %) que pour les N52 (51 %).

Les variations des pourcentages obtenus entre trois analyses effectuées le même jour ne sont que de $\pm 1\%$ pour l'antennarité et de $\pm 3\%$ pour le degré de sialylation.

Les variations des pourcentages obtenus entre trois analyses effectuées trois jours différents ne varient que de $\pm 1\%$ pour l'antennarité et de $\pm 3\%$ pour le degré de sialylation.

Enfin en comparant les valeurs moyennes obtenus pour les trois analyses effectuées le même jour et pour les analyses effectuées trois jours différents, on constate que celle-ci sont proches à $\pm 2\%$.

De ce fait, ces résultats sont tout à fait acceptables du fait qu'il s'agit d'une quantification relative sur les espèces majoritaires de r-hFSH et que nous voulons étudier la micro-hétérogénéité sur chaque site de glycosylation.

Nous pouvons donc conclure, que la méthode de quantification relative sur chacun des sites de glycosylation Asn⁵² et Asn⁷⁸ est reproductible.

III. Application des méthodes de quantification à des lots de composition en glycoformes variable

Après avoir vérifié que les méthodes donnaient des résultats reproductibles et répétables, ces deux méthodes ont été appliquées à des échantillons de composition en glycoformes variable.

Nous avons travaillé sur 4 expériences différentes avec en tout 11 fractions différentes à quantifier. Pour des raisons de confidentialité, nous ne présenterons que 2 expériences, une concernant des échantillons enrichis en espèces basiques et acides et l'autre concernant des échantillons enrichis selon l'antennarité.

III.1. Expérience 1 : quantification relative de lots enrichis en glycoformes acides et basiques

Dans cette expérience un lot de r-hFSH de référence a été purifié en deux fractions selon leur acidité :

- la fraction acide : enrichie en glycoformes acides.
- la fraction basique : enrichie en glycoformes basiques.

Pour des raisons de confidentialité, les méthodes de fractionnement utilisées pour purifier les lots de r-hFSH ne seront pas décrites ici.

Chacune de ces deux fractions ont été analysées selon les deux méthodologies optimisées et décrites précédemment :

- analyse directe de la fraction entière par ESI-MS
- analyses des mélanges de peptides issus de la digestion à la trypsine de la fraction entière par LC-MS

III.1.1. Quantification par analyses directes ESI-MS des fractions acides et basiques

La figure 13 représente les spectres de masse des fractions entières obtenus après analyse par ESI-MS. Pour pouvoir évaluer la qualité du fractionnement, ces deux spectres de masse sont comparés à celui obtenu pour un lot de r-hFSH de référence. De manière générale, dans la zone des ions multichargés 8+, on retrouve dans les trois spectres les mêmes glycoformes mais avec des intensités qui varient.

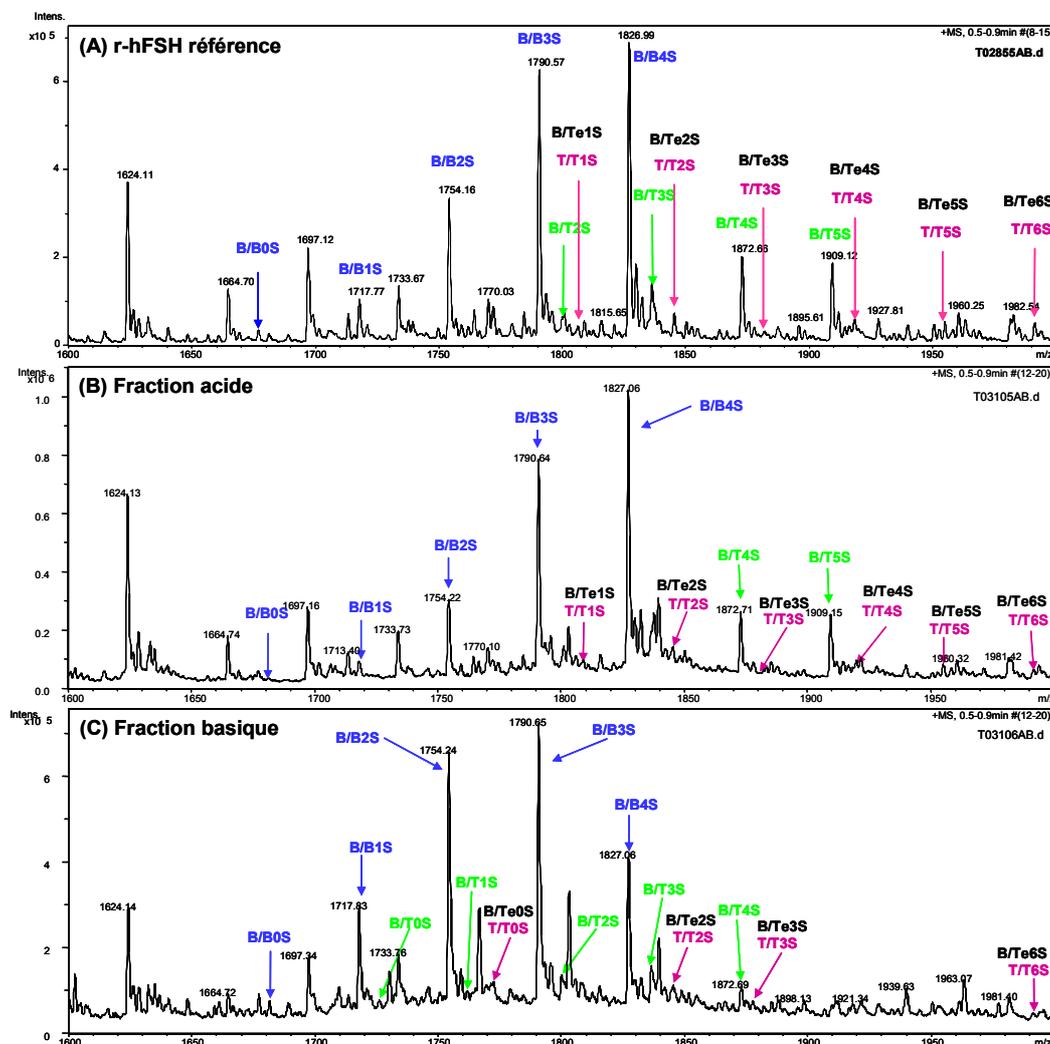


Figure 13 : Comparaison des spectres de masse obtenus par analyse directe par ESI-MS de
(A) la r-hFSH de référence, (B) la fraction acide et (C) la fraction basique.
Les profils correspondent aux gammes de m/z des ions multichargés 8+ (1600 à 2000 m/z).

En effet, le spectre de masse de la fraction acide montre que l'espèce B/B4S a légèrement augmenté par rapport à l'espèce B/B3S. De même, on constate que les espèces B/T4S et B/T5S sont présentes dans des proportions similaires à celles des espèces B/B2S, ce qui n'était pas le cas dans le lot de r-hFSH de référence. Enfin, on constate que la glycoforme B/B0S a fortement diminué.

Le spectre de masse de la fraction basique, quant à lui, montre que les espèces majoritaires sont les glycoformes B/B2S et B/B3S. De plus on constate une augmentation du taux des glycoformes B/B0S et B/B1S ainsi qu'une diminution des espèces B/T4S et B/T5S.

Concernant les glycoformes T/T ou B/Te, on ne constate aucune variation de celles-ci ni dans la fraction acide, ni dans la fraction basique.

La figure 14 présente les résultats de quantification relative des glycoformes en fonction de l'antennarité et du degré de sialylation pour chacune des fractions acide et basique.

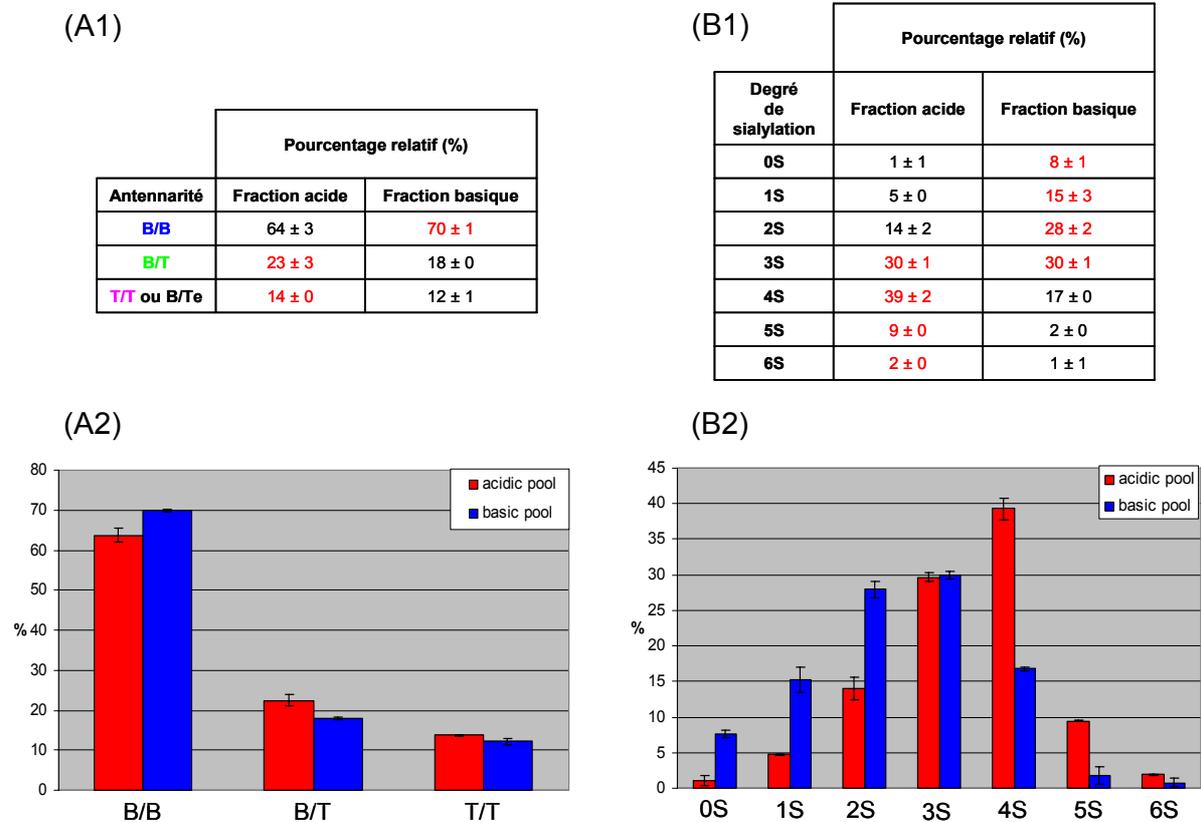


Figure 14 : Résultats de quantification relative des glycoformes en fonction de l'antennarité et du degré de sialylation pour chacune des fractions avec en (A1) les résultats de quantification selon l'antennarité et en (A2) la représentation graphique, en (B1) les résultats de quantification selon le degré de sialylation et en (B2) la représentation graphique.

En comparant ces résultats, on constate que le pourcentage relatif des espèces B/B a augmenté dans la fraction basique, tandis que celui des espèces B/T a légèrement augmenté dans la fraction acide. Néanmoins en comparant ces résultats à ceux obtenus pour la r-hFSH de référence, il n'y a pas de variation importante.

Toutefois, concernant les résultats de quantification relative selon le degré de sialylation, on constate une forte augmentation du pourcentage relatif des espèces 0S, 1S et 2S dans la fraction

basique, et inversement, on constate une forte augmentation des espèces 4S et 5S dans la fraction acide. Concernant les espèces 3S et 6S, on ne constate pas de variation.

III.1.2. Quantification par analyse LC-MS des mélanges de peptides issus de la digestion des fractions acides et basiques

La figure 15 représente les spectres de masse obtenus pour chacune des fractions en combinant l'ensemble des spectres de masse dans la zone d'éluion des glycopeptides.

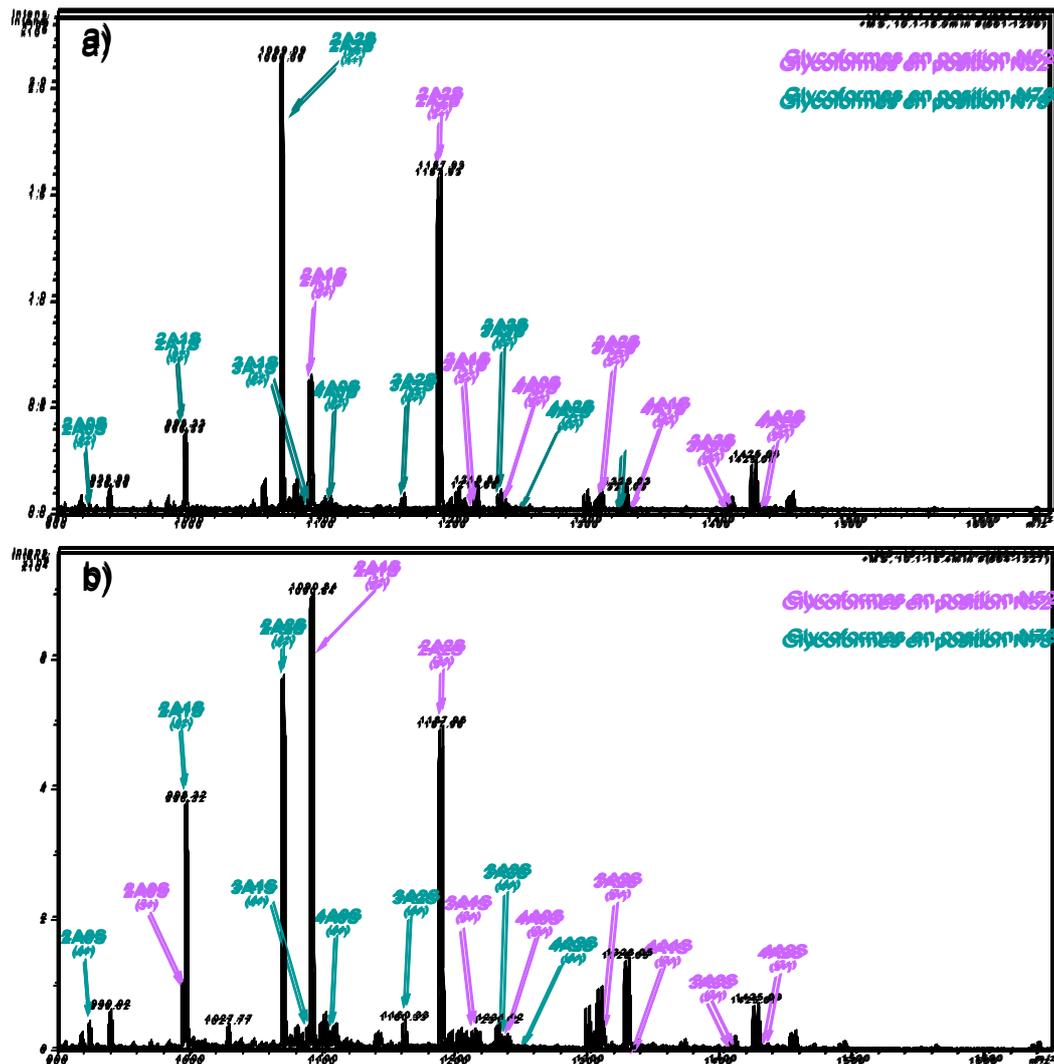
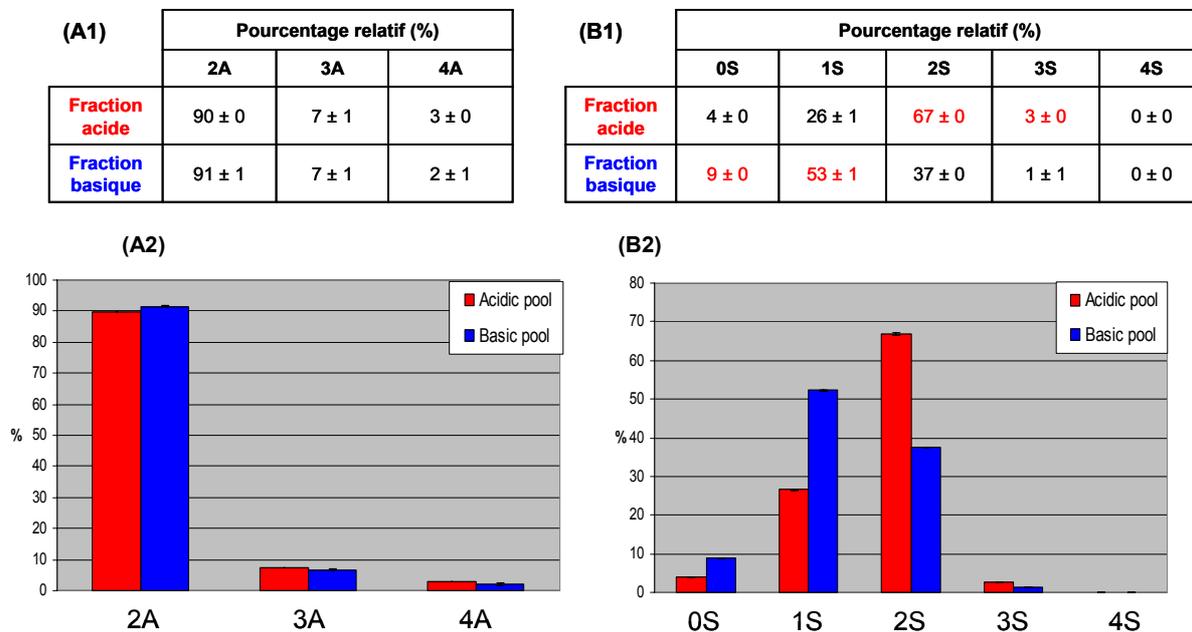


Figure 15 : Comparaison des profils des spectres de masse combinés dans la zone d'éluion des glycopeptides issus des analyses LC-MS de (a) la fraction acide et (b) la fraction basique.

En comparant les profils des spectres de masse combinés, on identifie dans les deux les mêmes ions mais avec des abondances variables. En effet, on constate une forte augmentation des glycopeptides N52 et N78 contenant les N-glycanes 2A0S et 2A1S dans la fraction basique.

La figure 16 illustre les valeurs moyennes des résultats de quantification pour les glycopeptides N52 et les glycopeptides N78.

Quantification relative des glycopeptides N52



Quantification relative des glycopeptides N78

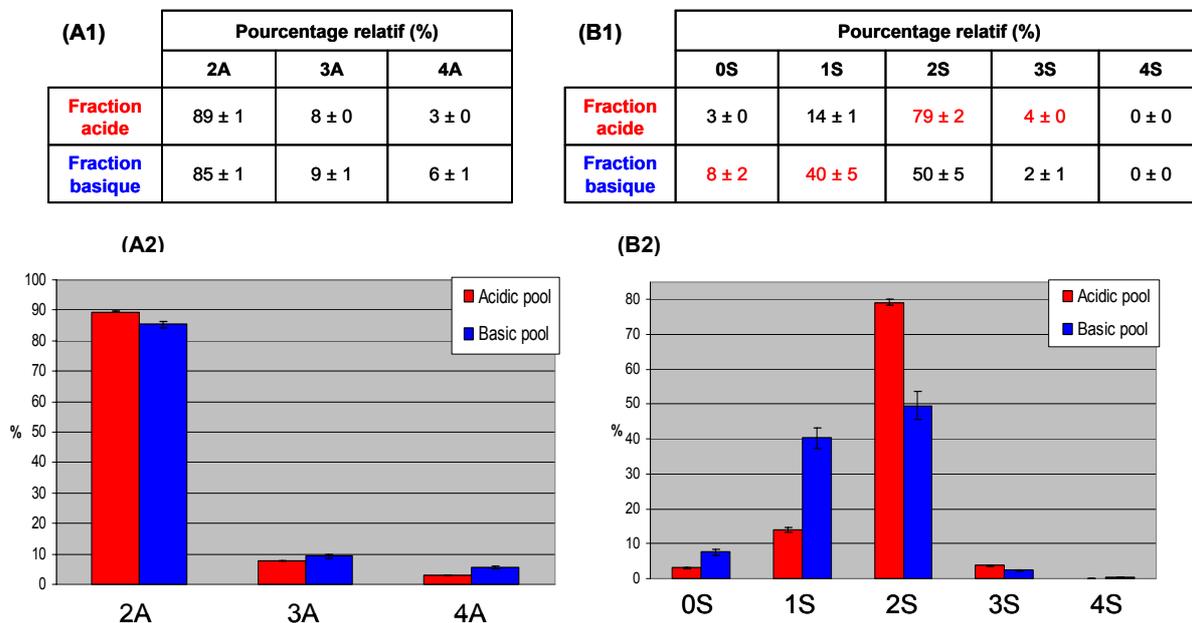


Figure 16 : Résultats de quantification relative des glycopeptides N52 et N78 en fonction de l'antennarité et du degré de sialylation pour chacune des fractions avec en (A1) les résultats de quantification selon l'antennarité et en (A2) la représentation graphique, en (B1) les résultats de quantification selon le degré de sialylation et en (B2) la représentation graphique.

Qu'il s'agisse des glycopeptides N52 ou N78, on constate une forte augmentation des espèces contenant deux et trois acides sialiques dans la fraction acide ainsi qu'une augmentation de l'abondance des espèces contenant un ou aucun acide sialique dans la fraction basique.

Concernant les résultats de quantification selon l'antennarité, on constate pour les glycopeptides N52 qu'il n'y a aucune différence entre la fraction acide et la fraction basique. Ces valeurs sont d'ailleurs identiques à celles obtenues pour la r-hFSH de référence (voir figures 11 et 12). Par contre concernant les glycopeptides N78, on constate une augmentation de l'abondance des espèces tetraantennée puisqu'on passe de 3 % à 6 %.

Ces résultats sont concluants et cohérents puisque les lots de r-hFSH ont été purifiés selon leur acidité donc leur nombre d'acides sialiques.

III.2. Expérience 2 : quantification relative de lots enrichis en glycoformes selon l'antennarité

Dans cette expérience un lot de r-hFSH de référence a été purifié en deux fractions selon l'antennarité :

- la fraction antennarité élevée : enrichie en glycoformes d'antennarité élevée.
- la fraction antennarité faible : enrichie en glycoformes d'antennarité faible.

Pour des raisons de confidentialité, les méthodes de fractionnement utilisées pour purifier les lots de r-hFSH ne seront pas décrites ici.

Chacune de ces deux fractions ont été analysées selon les deux méthodologies optimisées et décrites précédemment :

- analyse directe de la fraction entière par ESI-MS
- analyses des mélanges de peptides issus de la digestion à la trypsine de la fraction entière par LC-MS

III.2.1. Analyses directes par ESI-MS des fractions purifiées selon l'antennarité

La figure 17 représente les spectres de masse des fractions entières obtenus après analyse par ESI-MS. Pour pouvoir évaluer la qualité du fractionnement, ces deux spectres de masse sont comparés à celui obtenu pour un lot de r-hFSH de référence. De manière générale, dans la zone des ions multichargés 8+, on retrouve dans les trois spectres les mêmes glycoformes mais avec des intensités qui varient.

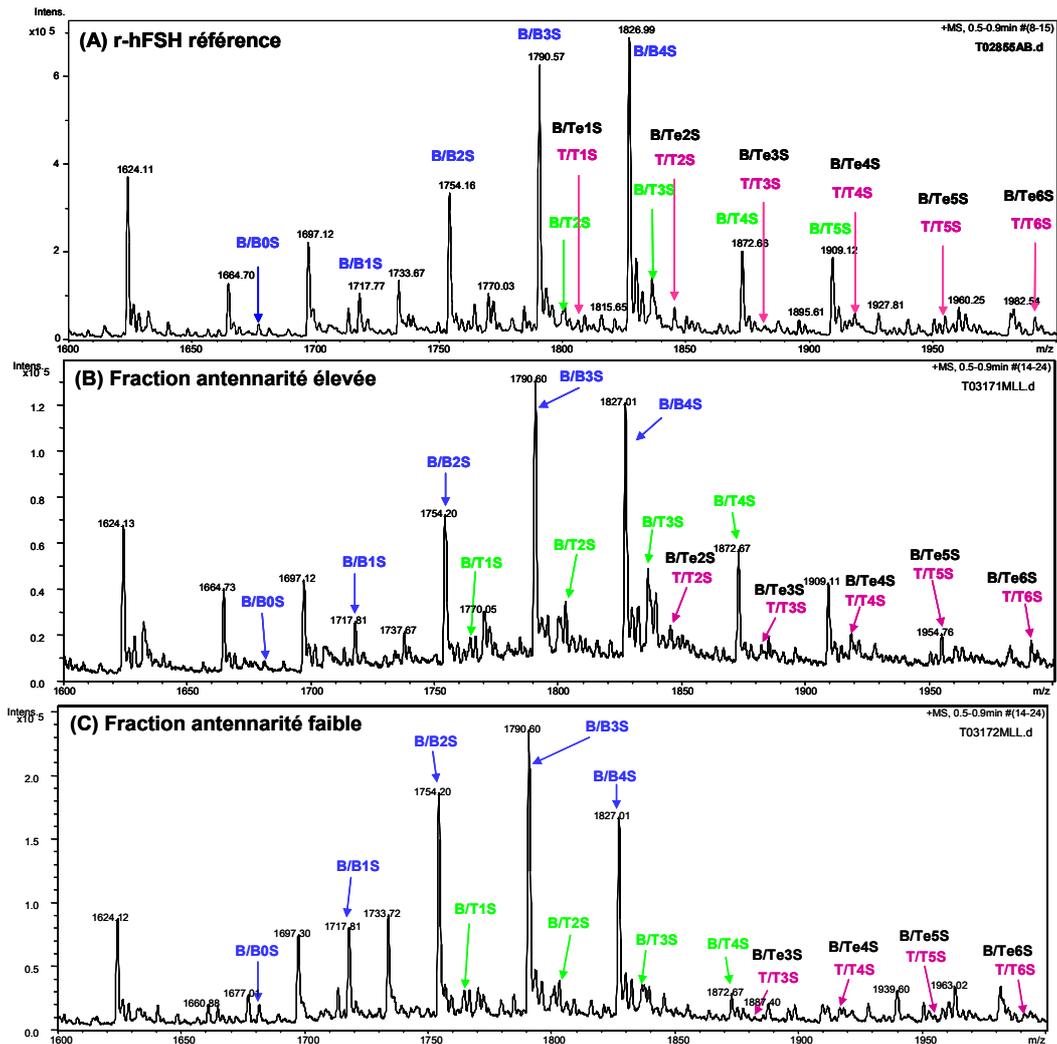


Figure 17 : Comparaison des spectres de masse obtenus par analyse directe par ESI-MS de (A) la r-hFSH de référence, (B) la fraction antennarité élevée et (C) la fraction antennarité faible. Les profils correspondent aux gammes de m/z des ions multichargés 8⁺ (1600 à 2000 m/z).

En effet, le spectre de masse de la fraction antennarité élevée montre que de manière générale, les espèces B/T et T/T ou B/Te sont plus abondantes. De même, à l'inverse, dans la fraction antennarité faible, ce sont les espèces B/B qui sont plus abondantes et les espèces T/T ou B/Te sont beaucoup moins abondantes que dans le spectre de la r-hFSH de référence. Toujours dans la fraction antennarité faible on constate que l'espèce B/T 5S a une abondance proche de zéro.

En termes d'acides sialiques, on ne constate pas de grosse variation si ce n'est que dans les deux fractions, c'est la glycoforme B/B 3S qui est la plus abondante en comparaison avec le spectre de référence.

La figure 18 présente les résultats de quantification relative des glycoformes en fonction de l'antennarité et du degré de sialylation pour chacune des deux fractions.

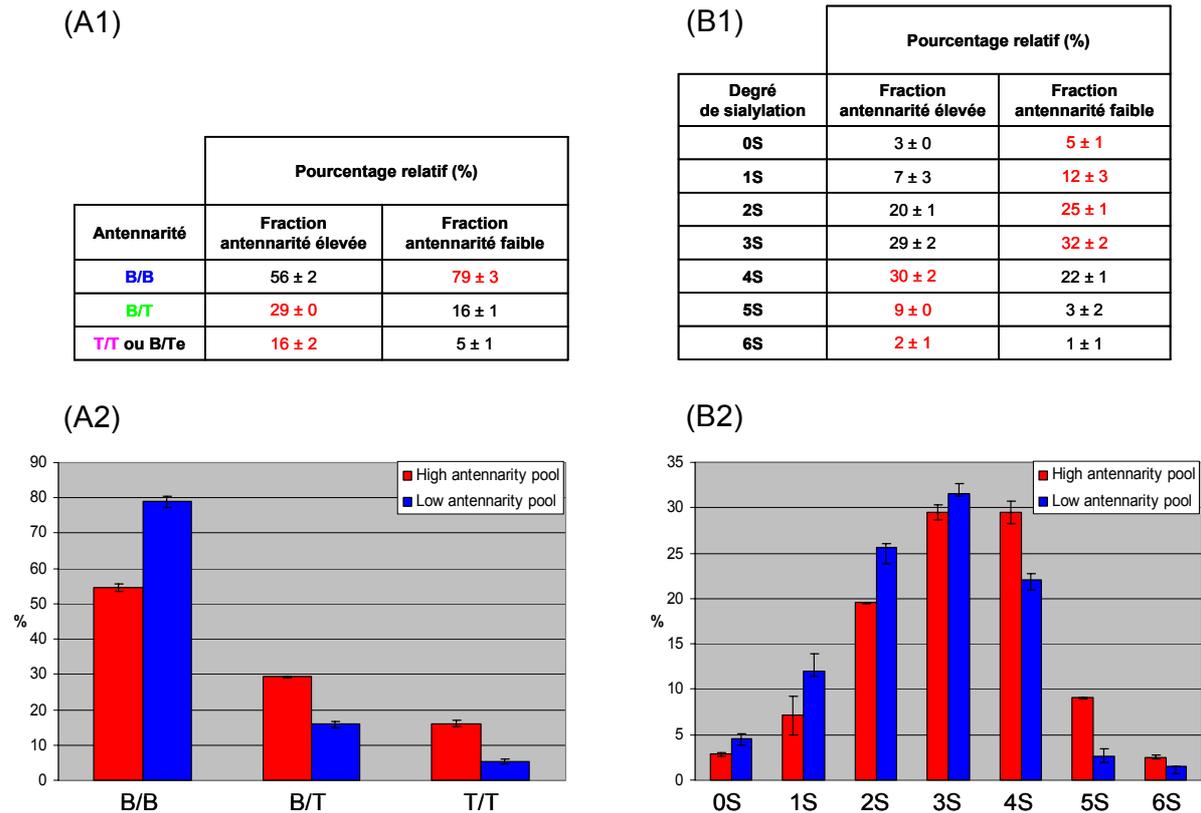


Figure 18 : Résultats de quantification relative des glycoformes en fonction de l'antennarité et du degré de sialylation pour chacune des fractions avec en (A1) les résultats de quantification selon l'antennarité et en (A2) la représentation graphique, en (B1) les résultats de quantification selon le degré de sialylation et en (B2) la représentation graphique.

Comme illustré d'après les profils des spectres de masse, les résultats de quantification montrent que nous avons augmenté l'abondance des glycoformes B/B dans la fraction acide et que nous avons pu également enrichir un lot en glycoformes B/T et T/T ou B/Te.

De ce fait, en ayant augmenté le nombre d'antennes dans la fraction antennarité élevée, nous augmentons également le nombre d'acides sialiques (4S, 5S et 6S) et inversement dans la fraction antennarité faible, on augmente le nombre de glycoformes contenant moins d'acides sialiques (0S, 1S, 2S).

Néanmoins, en comparant ces résultats à ceux obtenus sur la r-hFSH de référence, l'enrichissement n'est pas flagrant en ce qui concerne l'antennarité puisqu'on a 64 % de B/B, 23 % de B/T et 13 % de T/T ou B/Te. De même pour le nombre d'acides sialiques, la différence n'est pas très élevée, si ce n'est pour les glycoformes contenant 0S, 1S et 2S.

III.2.2. Analyses LC-MS des mélanges de peptides issus de la digestion des fractions purifiées selon l'antennarité

La figure 19 représente les spectres de masse obtenus pour chacune des fractions en combinant l'ensemble des spectres de masse dans la zone d'élution des glycopeptides.

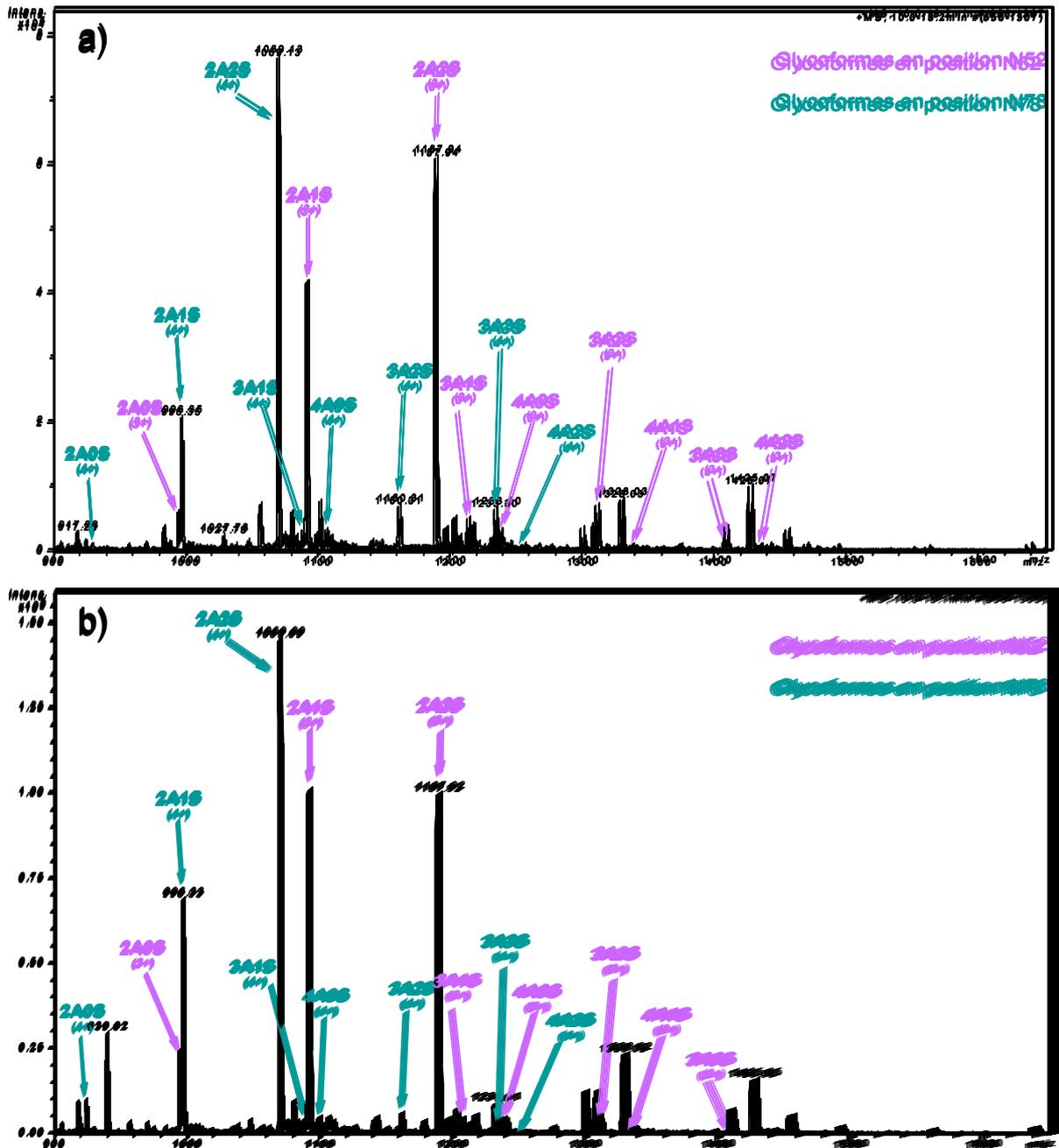


Figure 19 : Comparaison des profils des spectres de masse combinés dans la zone d'éluion des glycopeptides issus des analyses LC-MS de (a) la fraction antennarité élevée et (b) la fraction antennarité faible.

En comparant les profils des spectres de masse combinés, on identifie dans les deux les mêmes ions mais avec des abondances variables. En effet, on constate une augmentation des glycopeptides N52 et N78 contenant les N-glycanes 2A0S et 2A1S dans la fraction antennarité faible.

La figure 20 illustre les valeurs moyennes des résultats de quantification pour les glycopeptides N52 et les glycopeptides N78.

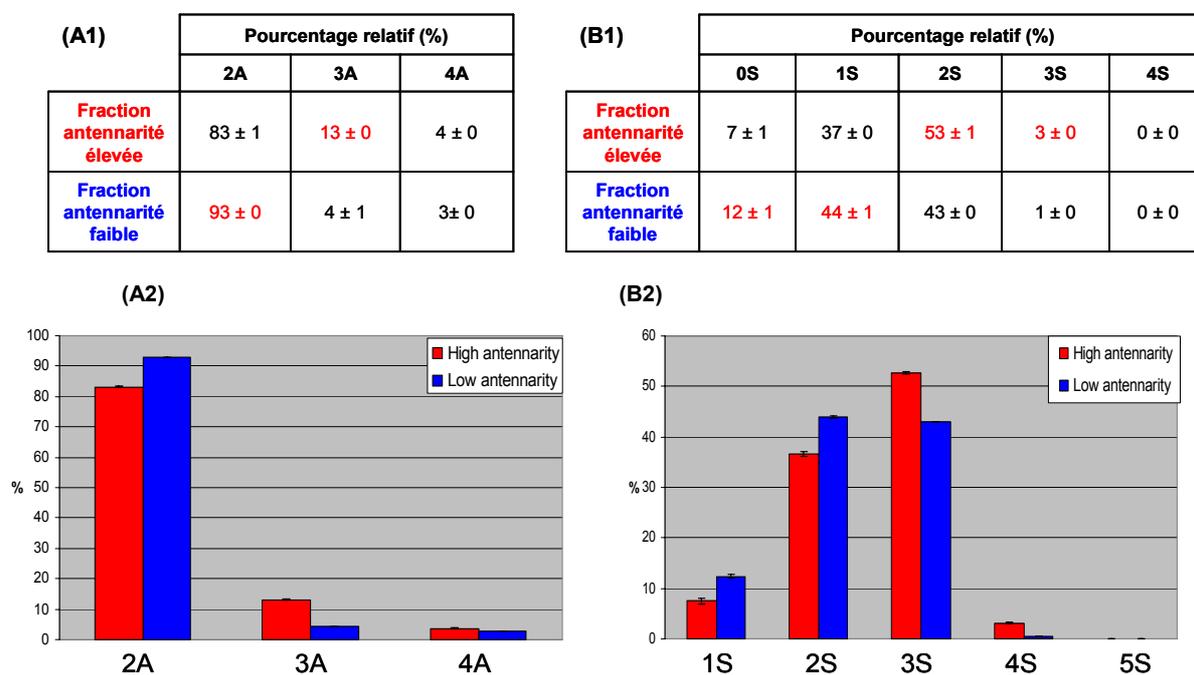
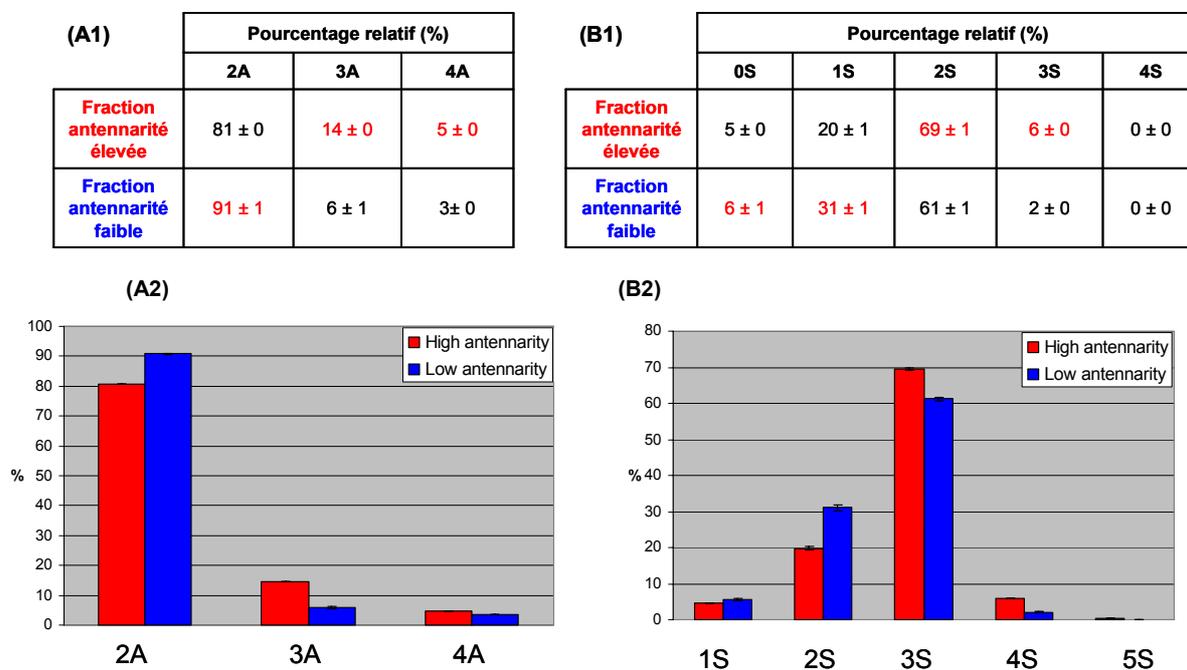
Quantification relative des glycopeptides N52**Quantification relative des glycopeptides N78**

Figure 20 : Résultats de quantification relative des glycopeptides N52 et N78 en fonction de l'antennarité et du degré de sialylation pour chacune des fractions avec en (A1) les résultats de quantification selon l'antennarité et en (A2) la représentation graphique, en (B1) les résultats de quantification selon le degré de sialylation et en (B2) la représentation graphique.

En comparant les fractions entre elles, qu'il s'agisse des glycopeptides N52 ou N78, on constate une augmentation des espèces biantennées dans la fraction antennarité faible et une augmentation des espèces triantennées et tétraantennées dans la fraction antennarité élevée. De ce fait, les espèces contenant 0S ou 1S sont plus abondantes dans la fraction faible antennarité et inversement ce sont les espèces contenant 2S et 3S qui sont plus abondantes dans la fraction antennarité élevée. Concernant les espèces 4S, aucun glycopeptide N52 ou N78 de ce type n'est observé.

En comparant ces résultats à ceux obtenus pour le mélange de peptides de la r-hFSH de référence, on constate que ce type de fractionnement a permis d'enrichir en glycopeptides N52 et N78 triantennés puisqu'on passe en moyenne de 7 % à 14 % dans la fraction haute antennarité. Par contre pour la fraction antennarité faible, on ne constate pas d'augmentation du pourcentage des glycopeptides N52 et N78 biantennés par rapport au lot de référence.

En ce qui concerne le degré de sialylation, on ne constate que de faibles variations par rapport au lot de référence. En effet, pour les glycopeptides N52, on a une légère augmentation de l'abondance des espèces 0S et 1S dans la fraction antennarité faible et mais aucune variation des espèces 2S et 3S dans la fraction antennarité élevée. De même pour les glycopeptides N78, on a une légère augmentation de l'abondance des espèces 1S dans la fraction antennarité faible puisqu'on passe de 27 % à 31 % et une légère augmentation des espèces 3S dans la fraction antennarité élevée puisqu'on passe de 3 % à 6 %.

IV. Conclusion

Cette étude nous a permis de montrer la **complémentarité de deux méthodes** pour la **caractérisation** et la **quantification relative des glycoformes** d'une **protéine glycosylée** contenant **plusieurs sites de glycosylation** occupés par **différents glycanes**.

L'analyse **ESI-MS** de la **r-hFSH entière** nous a permis d'étudier **l'hétérogénéité globale** de la sous-unité alpha et de déterminer quelles sont les glycoformes les plus abondantes et dans quelles proportions. Nous avons pu **quantifier les différentes glycoformes** avec une **précision de $\pm 2\%$** .

L'analyse **LC-MS du digeste de la r-hFSH** nous a permis d'étudier **la micro-hétérogénéité sur chaque site de glycosylation** de la sous-unité alpha (Asn⁵² et Asn⁷⁸) et de déterminer pour chaque site quels sont les N-glycanes majoritaires et en quelles proportions. Nous avons pu **quantifier les différents glycopeptides** avec une **précision de $\pm 1\%$ à $\pm 3\%$** .

Par la suite, ces méthodes d'analyses rapides et reproductibles ont été **appliquées à différentes fractions de r-hFSH enrichies** selon **l'antennarité**, le **nombre d'acides sialiques** ou encore les deux.

Avec ces méthodes de quantification relative, il nous a été possible de **montrer des variations** plus ou moins importantes dans la **composition** et **l'abondance** des différentes **glycoformes**.

En **perspective**, ces fractions de composition en glycoformes connues seront soumises à des **tests biologiques** afin de vérifier leur **efficacité** dans le cas d'infertilités chez la femme. En effet, le signal

biologique envoyé par la FSH permettant la nidation est optimale pour des glycoformes d'antennarité et de sialylation donnée en position Asn⁵².

La complémentarité des deux méthodes nous permettant de déterminer si une fraction est majoritairement biantennée ou triantennée en Asn⁵², il sera possible de **définir à quel moment ces fractions** doivent être **injectées durant le cycle menstruel** de la femme pour **aboutir à une nidation** ou **éviter toutes nidations supplémentaires** tout en maintenant une première nidation. Ceci permettrait **d'éviter des grossesses multiples** (3 à 5 enfants) qui sont trop souvent observées lors de ces traitements hormonaux.

BIBLIOGRAPHIE

Anobile C.J., Talbot J.A., McCann S.J., Padmanabhan V. and Robertson W.R. (1998). Glycoform composition of serum gonadotrophins through the normal menstrual cycle and in the post-menopausal state. *Mol Hum Reprod* **4**, 631-639.

Bishop L.A., Robertson D.M., Cahir N. and Schofield P.R. (1994). Specific roles for the asparagine-linked carbohydrate residues of recombinant human follicle stimulating hormone in receptor binding and signal transduction. *Mol Endocrinol* **8**, 722-731.

Fan Q.R. and Hendrickson W.A. (2005). Structure of human follicle-stimulating hormone in complex with its receptor. *Nature* **433**, 269-277.

Fox K.M., Dias J.A. and Van Roey P. (2001). Three-dimensional structure of human follicle-stimulating hormone. *Mol Endocrinol* **15**, 378-389.

Gervais A., Hammel Y.A., Pelloux S., Lepage P., Baer G., Carte N., Sorokine O., Strub J.M., Koerner R., Leize E. and Van Dorselaer A. (2003). Glycosylation of human recombinant gonadotrophins: characterization and batch-to-batch consistency. *Glycobiology* **13**, 179-189.

Gomord V. and Faye L. (2004). Posttranslational modification of therapeutic proteins in plants. *Curr Opin Plant Biol* **7**, 171-181.

Loumaye E., Dreano M., Galazka A., Howles C., Ham L., Munafò A., Eshkol A., Giudice E., De Luca E., Sirna A., Antonetti F., Giartosio C.E., Scaglia L., Duthu B., Cymbalista S. and Lepage P. (1998). Recombinant Human FSH Product Development Group. Recombinant follicle stimulating hormone: development of the first biotechnology product for the treatment of infertility. *European Society of Human Reproduction and Embryology* **4**, 862-881.

Ulloa-Aguirre A., Timossi C., Damian-Matsumura P. and Dias J.A. (1999). Role of glycosylation in function of follicle-stimulating hormone. *Endocrine* **11**, 205-215.

Zambrano E., Olivares A., Mendez J.P., Guerrero L., Diaz-Cueto L., Veldhuis J.D. and Ulloa-Aguirre A. (1995). Dynamics of basal and gonadotropin-releasing hormone-releasable serum follicle-stimulating hormone charge isoform distribution throughout the human menstrual cycle. *J Clin Endocrinol Metab* **80**, 1647-1656.

CHAPITRE 2

CARACTERISATION FINE DES N-GLYCOSYLATIONS: STRATEGIES BASEES SUR L'ANALYSE NANOLC-MS/MS DE GLYCOPEPTIDES

Le « Chapitre 2, partie Introduction bibliographique » nous a permis de faire un point sur l'apport de l'analyse protéomique pour la **caractérisation fine de protéines N-glycosylées**. Ainsi nous avons pu montrer que leur **étude n'est pas aisée** qu'il s'agisse soit de l'étude d'une protéine de séquence connue et dont le site de N-glycosylation est connue (les protéines recombinantes par exemple) ou soit de la caractérisation de nouvelles protéines N-glycosylées (recherche de sites de N-glycosylation, détermination de la structure des N-glycanes et étude de l'hétérogénéité des glycoformes).

De manière générale, l'identification de protéines en analyse protéomique comporte plusieurs étapes : la séparation des protéines, la digestion en peptides, l'analyse nanoLC-MS/MS des peptides et l'identification des protéines par « Peptide Fragment Fingerprint ». Dans le cas d'échantillons complexes contenant des peptides N-glycosylés nous allons être confrontés à plusieurs difficultés lors de l'utilisation de ce type d'approche:

- **La taille des peptides de digestion** : la N-glycosylation entraînant des différences de masse supérieures à 1000 Da, la masse des N-glycopeptides peut être très élevée et entraîner des difficultés lors de la détection par spectrométrie de masse et lors des étapes de fragmentation MS/MS.
- **Détection** : dans un mélange complexe les N-glycopeptides sont peu abondants par rapport aux peptides non-glycosylés et leur signal peut souvent être masqué par l'ensemble de ces peptides. Il faut donc améliorer leur détection par des méthodes d'enrichissement mais également optimiser les conditions d'analyses.
- **Identifier et caractériser les N-glycopeptides par MS/MS** : il est nécessaire d'avoir des spectres de fragmentation très informatifs pour pouvoir déterminer la structure du ou des N-

glycanes, le site de N-glycosylation et la séquence en acides aminés du peptide N-glycosylé. Ainsi les paramètres de fragmentation doivent être optimisés.

- **Interprétation des spectres de fragmentation** : actuellement il n'est pas possible de le faire à l'aide de logiciels automatisés comme dans le cas du « Peptide Fragment Fingerprint ». L'interprétation se fait manuellement. De ce fait il est important de pouvoir rapidement déterminer quel spectre de fragmentation correspond à celui d'un N-glycopeptide. Pour cela on utilise les ions diagnostiques.

Une **solution** pour contourner ces difficultés réside dans l'utilisation **d'étapes d'enrichissement** après digestion des protéines mais également d'utiliser des **colonnes chromatographiques adaptées** pour améliorer leur séparation donc leur détection. Néanmoins, toutes ces étapes entraînant des **pertes d'échantillons**, le défi réside dans une analyse nanoLC-MS/MS des mélanges de peptides et de N-glycopeptides **s'affranchissant de toutes ces étapes**.

En 2005, le laboratoire fait l'acquisition d'un nouveau système nanoLC-MS commercialisé par Agilent Technologies, la **nanoLC-Chip-MS**. De même, des chips remplies de **phase stationnaire de carbone graphite** sont développées. Celles-ci sont classiquement utilisées pour séparer les oligosaccharides mais peuvent également être utilisées pour **l'analyse de N-glycopeptides**. De ce fait durant ce travail de thèse, j'ai mis au point des méthodologies utilisant ce système LC et ces chips pour l'analyse de N-glycopeptides issus de mélanges complexes.

Nous décrirons dans un **premier chapitre** le **principe, les avantages de ce système chromatographique** couplé à une trappe à ions (IT), les avantages de l'utilisation de chip carbone graphite ainsi que la méthodologie que nous avons choisi d'employer pour l'analyse des N-glycopeptides.

Puis nous présenterons dans un **second chapitre** trois **études** qui ont été effectuées au laboratoire pour **caractériser des N-glycopeptides** par analyses **nanoLC-Chip-MS/MS sur une trappe à ions (IT)**. Ces études ont été menées sur un anticorps monoclonal purifié, des protéines séparées sur gels 1D et des protéines séparées sur gels 2D.

I. Méthodologie : Analyse de glycopeptides par nanoLC-Chip-MS/MS sur chips carbone graphite

I.1. Le système nanoLC-Chip-MS/MS

Classiquement, les systèmes nanoLC-MS couplés à une trappe à ions se composent d'une colonne d'enrichissement, d'une colonne analytique et d'une aiguille nanospray qui est introduite à l'entrée de la source électrospray du spectromètre de masse. Toutes ces composantes sont reliées entre elles par des connecteurs via des vannes. Ce système est un des plus utilisé aujourd'hui en analyse protéomique puisqu'il permet de travailler à des débits de l'ordre de centaines de nL/min et est beaucoup plus sensible que des systèmes LC-MS conventionnels.

Toutefois, un des plus grands inconvénients de ce système réside dans le nombre de connexions qui augmentent considérablement les volumes morts et entraînent une perte en résolution chromatographique mais également une perte en sensibilité. Un autre inconvénient majeur est le système d'aiguille nanospray en sortie de colonne qui nécessite d'être changé régulièrement.

De ce fait, des améliorations ont été apportées en mettant au point un nouveau système chromatographique la chip et son interface, le cube.

I.1.1. La chip et son interface, le cube

Les chips sont des dispositifs qui intègrent la pré-colonne, la colonne analytique et l'aiguille nanospray (figure 1).

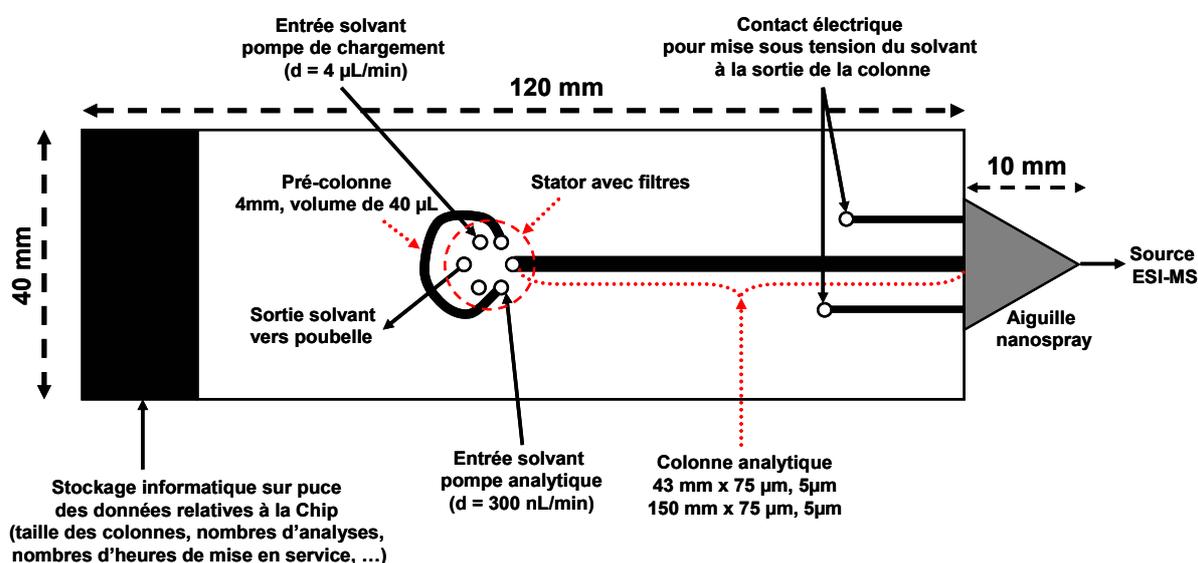


Figure 1 : Représentation schématique de la Chip et de ces composants. Elle intègre la pré-colonne, la colonne analytique, l'aiguille nanospray, le stator avec les 6 ports et des filtres dans un sandwich en feuilles de polyimide protégé par une carcasse métallique. L'ensemble mesure 120 mm de long et 40 mm de large.

Elles sont fabriquées à partir de films de polyimide sur lesquels les canaux, les chambres et les ports sont générés par ablation laser. Une fois les canaux créés, la pré-colonne et la colonne analytique sont remplies de phase stationnaire [Yin et al., 2005]. L'ablation laser est également utilisée pour tailler la chip à sa forme finale et pour former la pointe de l'aiguille nanospray qui une fois formée et métallisée.

La mise en place de la chip dans le système de chromatographie se fait via ce qu'on appelle le cube (figure 2).

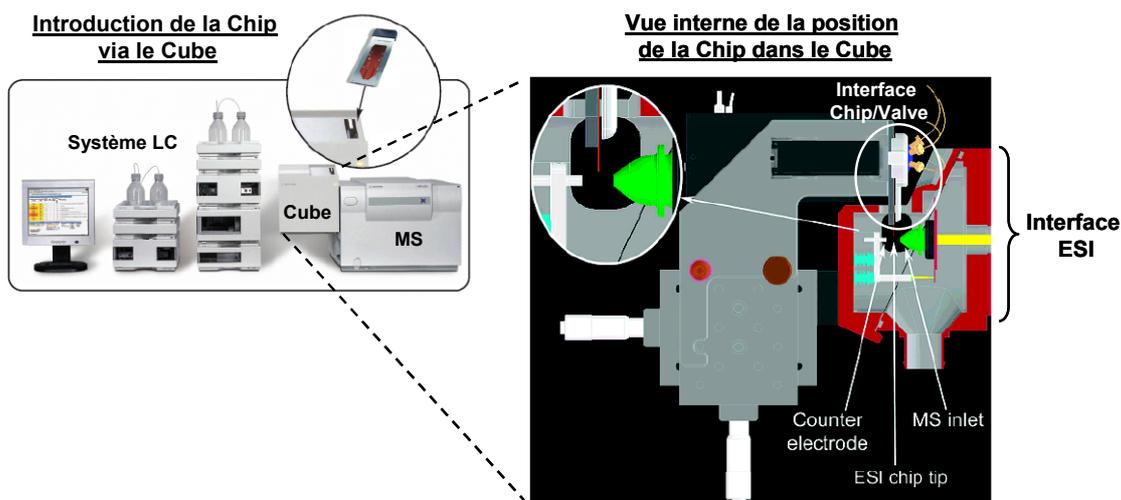


Figure 2 : Schéma représentant la mise en place de la chip dans le système LC via le cube [Yin et al., 2005].

La chip y est introduite, et par un mécanisme contrôlé par l'utilisateur, elle va venir se placer entre le rotor et le stator qui forment la valve permettant de faire la connexion entre la pré-colonne et la colonne analytique. Par un système de pression, la chip est prise en sandwich entre les deux. Elle est positionnée dans le cube de telle sorte que l'aiguille nanospray vienne se placer dans l'interface ESI face à l'entrée du spectromètre de masse. La position de l'aiguille peut être réglée manuellement pour obtenir un spray de qualité.

I.1.2. Capacité du système

En combinant sur un même dispositif la pré-colonne, la colonne et l'aiguille nanospray, le système nanoLC-Chip-MS offre de nombreux avantages.

En effet, les volumes morts sont considérablement diminués permettant de gagner en résolution et en sensibilité. De même, le système d'aiguille nanospray est également avantageux puisqu'il offre une stabilité de spray considérable.

Les chips couramment utilisées en analyse protéomique sont remplies de phase stationnaire inverse C18. Elles existent avec des colonnes analytiques de deux longueurs, 43 mm x 75 μ m et 150 mm x 75 μ m.

Classiquement les colonnes de 43 mm sont utilisées pour l'analyse de mélanges peu complexes, comme par exemple les extraits de peptides issus de spots de gels 2D. Quant aux colonnes de 150mm, elles sont plutôt utilisées pour l'analyse de mélanges plus complexes tels que les extraits de peptides issus de bandes de gels 1D [Yin and Killeen, 2007].

Après l'acquisition au laboratoire de ce système installé sur une trappe à ions, l'HCTultra de Bruker Daltonics, plusieurs études ont été effectuées pour optimiser les analyses sur ce type de systèmes pour obtenir les meilleurs résultats pour l'identification de protéines à partir d'analyses nanoLC-MS/MS de mélanges de peptides issus de spots de gels 2D.

Pour cela, nous avons testé différents gradients et nous avons optimisé les paramètres du spectromètre de masse (vitesse de balayage pour avoir une résolution isotopique, temps d'accumulation des ions dans la trappe pour éviter des effets d'espace charge, temps d'un cycle MS et d'un cycle MS/MS pour obtenir des spectres de qualité, ...).

Afin d'illustrer les performances du système nous comparerons les performances du système nanoLC-Chip-MS/MS sur les chips 43mm et 150 mm à celle du système classique nanoLC-MS/MS utilisant des colonnes capillaires de 150mm x 75 μ m et des aiguilles cotées pour générer le nanospray. Toutes les analyses ont été réalisées en couplage avec la même trappe à ions, l'HCTultra (Bruker Daltonics).

a) Conditions expérimentales optimales pour les deux systèmes

Instruments:

- Chromatographie: Agilent HPLC 1100 series équipée d'une pompe nanodébit (G2226A), d'une pompe capillaire (G1376A), d'un HPLC-Chip Cube (G4240A) ou d'un compartiment à colonnes (G1312A).
- Spectromètre de masse ESI-IT : HCTultra (Bruker Daltonics).

Analyses nanoLC-MS/MS:

- Pré-colonne: ZORBAX 300SB-C18, 300 μ m x 5mm, 5 μ m ; Colonne analytique: ZORBAX 300SB-C18, 75 μ m x 150mm, 3.5 μ m.
- Conditions de chargement: 20 μ L/min avec 100% de H₂O + 0.1% HCOOH pendant 5 minutes; volume injecté 3 μ L
- Conditions analytiques: débit = 230 nL/min; phase mobile: A = H₂O + 0.1% HCOOH, B = ACN + 0.1% HCOOH ; gradient: de 15% B à 34.5% B en 30 min, puis à 90% B en 1 min.
- Conditions MS: ICC 90000 ; Temps d'accumulation max. 500 ms ; Gamme de balayage 200-1500 m/z à une vitesse de scan de 8100 m/z / sec ; 4 spectres MS sont moyennés.
- Conditions MS/MS: Gamme de balayage 50-2800 m/z à une vitesse de scan de 26000 m/z / sec ; 3 ions précurseurs par MS sont sélectionnés (préférentiellement les doublement chargés); exclusion des ions après 2 spectres MS/MS pendant 3 min.

Analyses nanoLC- Chip-MS/MS:

- Pré-colonne: ZORBAX 300SB-C18, 4 mm, 40 nL, 5 μ m ; Colonne analytique: ZORBAX 300SB-C18, 75 μ m x 43 mm, 5 μ m ou ZORBAX 300SB-C18, 75 μ m x 150 mm, 5 μ m.
- Conditions de chargement: 3,75 μ L/min avec 100% of H₂O + 2% ACN + 0.1% HCOOH; pendant 3 minutes; volume injecté 3 μ L.
- Conditions analytiques: débit = 300 nL/min; phase mobile: A = H₂O + 2% ACN + 0.1% HCOOH, B = ACN + 2% H₂O + 0.1% HCOOH ; gradient: de 10% B to 40% B en 7 min ou 30 min pour la chip 150 mm, puis à 70% B en 2 min.

- Conditions MS: ICC 100000 ; Temps d'accumulation max. 200 ms ; Gamme de balayage 250-2500 m/z à une vitesse de scan de 8100 m/z / sec ; 4 spectres MS moyennés.
- Conditions MS/MS: Gamme de balayage 50-2800 m/z à une vitesse de scan de 26000 m/z / sec ; 3 ions précurseurs par MS sont sélectionnés (préférentiellement les doublement chargés); exclusion des ions après 2 spectres MS/MS pendant 1 min.

b) Performances du système nanoLC-Chip-MS/MS

Afin de comparer les performances des systèmes nous avons injecté 300 fmol d'un mélange de peptides issus de la digestion à la trypsine de sérum d'albumine bovine (BSA). La figure 3 illustre les trois profils LC-MS/MS obtenus sur les différentes colonnes après optimisation des gradients de chromatographie et les paramètres du spectromètre de masse, ainsi que les courants de l'ion doublement chargé de m/z 569,8 pour ces trois mêmes analyses.

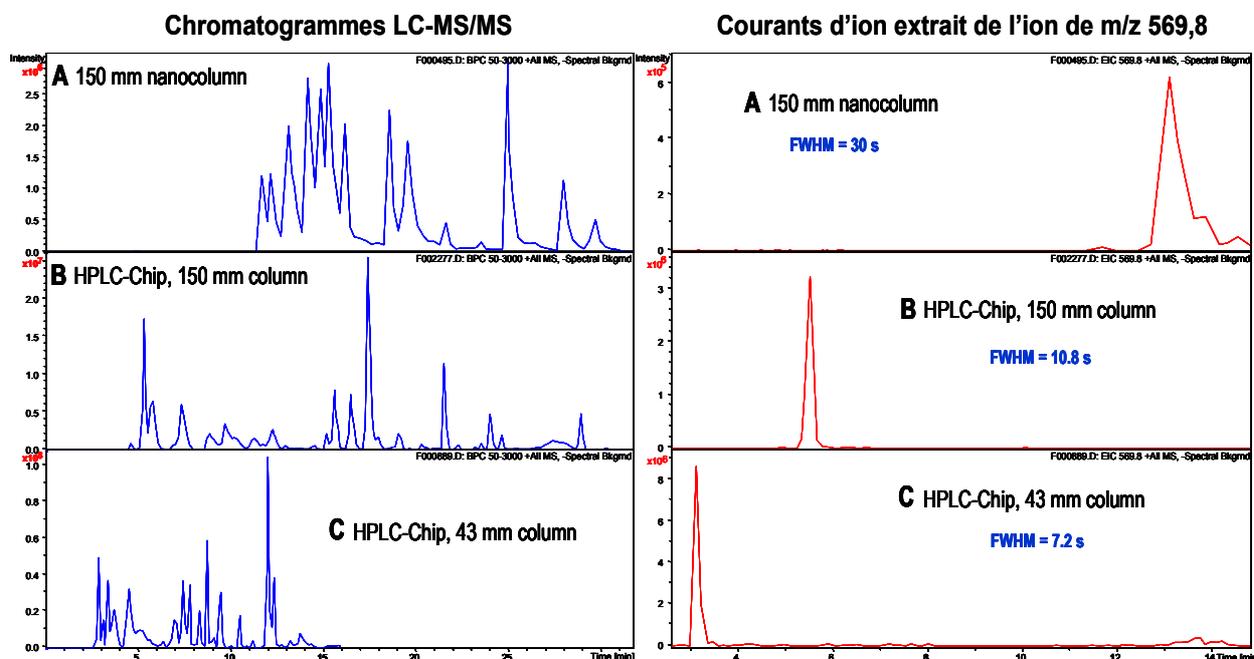


Figure 3 : Comparaison des analyses effectuées sur les deux systèmes nanoLC et sur trois colonnes différentes. Injection de 300 fmol d'un mélange standard de peptides issus de la digestion à la trypsine de la BSA.

En comparant les profils et les trois courants d'ions extraits, on constate que la **résolution chromatographique** est meilleure en utilisant le système nanoLC-Chip. En effet, le peptide séparé sur chip 43 mm a une largeur à mi hauteur (FWHM) de 7 secondes contre 30 secondes sur une nanocolonne de 150mm. De plus avec des chips de 43 mm, le temps d'analyse est considérablement diminué (15 minutes contre 45 minutes).

De même la **sensibilité** est considérablement augmentée. On gagne un facteur 5 à 10 entre l'analyse effectuée sur une chip de 43 mm et celle effectuée sur une nanocolonne de 150 mm. De même, entre une analyse faite sur une chip de 150mm et la nanocolonne, on gagne un facteur 5 à 10 avec une meilleure séparation chromatographique.

Afin d'étudier la limite de détection du système, plusieurs quantités d'un mélange de peptides issus de la digestion de la BSA ont été injectées en mode nanoLC-MS. Le système nanoLC-Chip permet de descendre à une limite de détection de 10 fmol en MS. Néanmoins, pour une analyse nanoLC-Chip-MS/MS, il faut travailler à 100 fmol pour avoir des spectres de fragmentation de bonne qualité et très informatifs.

L'ensemble de ces améliorations permet d'augmenter **le nombre et la qualité des spectres de fragmentation** mais également d'améliorer la confiance sur les identifications. Le tableau 1 illustre les résultats des identifications obtenus en soumettant les listes de masses au moteur de recherche MASCOT.

Colonnes	Nombre de peptides identifiés	Recouvrement de séquence de la protéine (%)	Erreur sur la MS (ppm)	Moyenne des scores MSMS obtenus pour 15 peptides communs aux 3 analyses
Nanocolonne 150mm	27	45	59	47
Chip 150mm	34	53	61	50
Chip 43mm	32	51	55	56

Tableau 1 : Comparaison de la qualité des identifications à partir des trois analyses LC-MS/MS d'un mélange standard de peptides de la BSA, en utilisant la stratégie PFF via le moteur de recherche MASCOT.

I.1.3. Conclusion

Ce système très prometteur qui offre un gain en résolution et en sensibilité, permet d'améliorer la qualité et le nombre de spectres de fragmentation avec des temps d'analyses considérablement diminués. Ainsi, il est possible d'augmenter le nombre de peptides identifiés et par conséquent d'améliorer la qualité des identifications mais également d'augmenter le nombre de protéines identifiées.

I.2. Les chips carbone graphite pour l'analyse de N-glycopeptides.

Généralement, les composés polaires tels que les polysaccharides sont séparés sur des colonnes échangeuses d'ions ou des colonnes en phase normale et en phase inverse.

Néanmoins, afin de permettre d'améliorer leur séparation et leur détection, il est souvent nécessaire de dériver les oligosaccharides avec des composés hydrophobes tels que le 2-aminobenzamide (2-AB) par exemple [Wuhrer et al., 2005].

Dans les années 1980, Knox et al. développe des phases de carbone graphite avec des petites tailles de particules pour colonnes de chromatographie liquide haute performances (HPLC). Cette phase qui semble avoir des propriétés d'adsorption similaires à celles des phases inverses, permet de séparer les composés polaires par interactions hydrophobes mais également par des interactions donneurs-accepteurs d'électrons [Hanai, 2003], [Hanai, 2004], [Forgacs, 2002]. Il n'est donc pas

nécessaire de dériver les polysaccharides. De même ces phases sont très résistantes aux pH acides et aux sels.

Depuis, de nombreuses études ont été effectuées qui utilisent ce type de phases pour l'analyse de polysaccharides. En 1990, Koizumi utilise ces colonnes HPLC pour séparer des mono et des oligosaccharides [Koizumi et al., 1990]. En 1992, Davies et al. utilisent ces colonnes pour séparer des oligosaccharides alditols et des glycopeptides. Ces études ont montré qu'il est possible de séparer des isomères ou des polysaccharides de structures très proches sans utiliser des étapes de dérivation [Davies et al., 1992], [Davies et al., 1993], [Koizumi, 1996]. En 1999, Kawasaki utilise ces colonnes pour l'analyse LC-MS et LC-MS/MS de carbohydrates qui lui permet de déterminer la distribution des oligosaccharides mais également de déterminer leur structure fine à partir d'un échantillon très hétérogène [Kawasaki et al., 1999].

Avec la commercialisation du système LC-Chip en 2005, Agilent Technologies développe des chips contenant une pré-colonne et une colonne analytique remplies toutes deux avec de la phase de carbone graphite. En combinant les avantages de ce système et la sélectivité de cette phase, celle-ci montrent rapidement les avantages pour l'analyses nanoLC-MS/MS de mélanges complexes de N-glycanes.

En effet, une étude a montré qu'il est possible de détecter des N-glycanes au niveau sub-picomolaire à partir d'un mélange de 10 femtomol [Ninonuevo et al., 2005], [Perkins, 2006].

I.3. Méthodologie

Comme nous avons pu le voir dans le chapitre 2 de l'Introduction bibliographique (p. 65-68), l'analyse LC-MS/MS de N-glycopeptides présente un intérêt tout particulier dans l'étude de glycoprotéines.

En effet, de manière générale les spectres de fragmentation MS/MS des N-glycopeptides contiennent trois types d'ions fragments : les pertes des sucres sur le N-glycopeptide, des ions fragments du N-glycane et un ion correspondant à la masse du peptide non-glycosylé.

Ainsi, un spectre de fragmentation MS/MS d'un N-glycopeptide permet d'une part de déterminer l'enchaînement des monosaccharides donc la structure du N-glycane, et d'autre part il peut donner des informations sur la nature du site de N-glycosylation d'une protéine.

Dans certains cas, il est possible d'observer dans le spectre MS/MS des ions fragments correspondants à la chaîne peptidique. Or ces ions sont souvent difficilement détectables dans le spectre MS/MS car la liaison glycosidique est beaucoup plus labile que la liaison peptidique. Toutefois, pour déterminer la séquence en acides aminés du peptide N-glycosylé, il est possible d'effectuer une étape de MS³ sur un des ions fragments le plus intense qui correspond au peptide contenant un GlcNAc lié à l'asparagine.

Néanmoins, la caractérisation de N-glycopeptides par analyse nanoLC-MS/MS nécessite de nombreuses optimisations au niveau de la préparation des échantillons (digestion, enrichissement, dérivation) et des conditions d'analyses (séparation, paramètres du spectromètre de masse) pour améliorer leur détection.

Durant ce travail de thèse, notre objectif visait à caractériser des N-glycoprotéines en analysant des N-glycopeptides sans passer par des étapes d'enrichissement ou de dérivations. Pour cela nous avons utilisé le système nanoLC-Chip-MS couplé à une trappe à ions avec des chips carbone graphite beaucoup plus sélectives vis à vis des N-glycopeptides. Nous avons procédé de la manière suivante :

- Dans un premier temps, les mélanges de protéines et de N-glycoprotéines ont été **digérés à l'aide de la trypsine**. Puis les mélanges de peptides et de N-glycopeptides ont été analysés par **nanoLC-Chip-MS/MS** sur des chips **carbone graphite**.
- Dans un second temps il s'agissait de rapidement **détecter les spectres MS/MS correspondants à des N-glycopeptides**. Pour cela nous avons utilisé les **ions fragments du N-glycane** présents dans les spectres MS/MS. En effet, quelle que soit la nature du N-glycane, on observe toujours une série de mêmes ions fragments monochargés (204, 366, 528) qui correspondent à l'association d'hexoses et de N-acétylhexosamines. Ces ions fragments sont utilisés comme ions diagnostiques et en les extrayant de tous les spectres MS/MS, il est possible de rapidement détecter les N-glycopeptides dans une analyse.
- Puis tous **les spectres MS/MS contenant des ions diagnostiques ont été interprétés manuellement**. Les ions fragments correspondant à des **pertes de sucres** sur le N-glycopeptide ont été utilisés d'une part pour remonter à l'enchaînement des sucres et ainsi à la **structure du N-glycane** et d'autre part de déterminer **la masse du peptide non-glycosylé**. Celle-ci a d'ailleurs été confirmée par la présence dans les spectres MS/MS de l'ion correspondant au peptide non-glycosylé.

Une fois la structure du N-glycane et la masse du peptide non-glycosylé déterminées, il est nécessaire de pouvoir attribuer celles-ci à une protéine. Ainsi, **en parallèle** des analyses sur chip carbone graphite, nous avons utilisé une **stratégie d'analyse protéomique classique** :

- Les **mêmes mélanges** de peptides et de N-glycopeptides sont analysés sur des chips **phase inverse C18** et les **protéines identifiées par stratégie « PFF »**.
- Identification des **sites potentiels de N-glycosylation sur ces protéines**.
- **Digestion *in silico* de ces protéines** pour calculer la **masse théorique des peptides contenant ces sites de N-glycosylation**.

En combinant les **informations de structure des N-glycanes** et en comparant la **masse calculée du peptide non-glycosylé aux masses théoriques des peptides** contenant un site potentiel de N-glycosylation, il nous est alors possible **d'attribuer un N-glycane à un peptide d'une protéine si celui-ci est spécifique de cette protéine**.

La figure 4 illustre la manière dont nous avons procédé dans cette méthodologie pour identifier les N-glycopeptides après qu'ils aient été analysés par nanoLC-MS/MS sur chip carbone graphite.

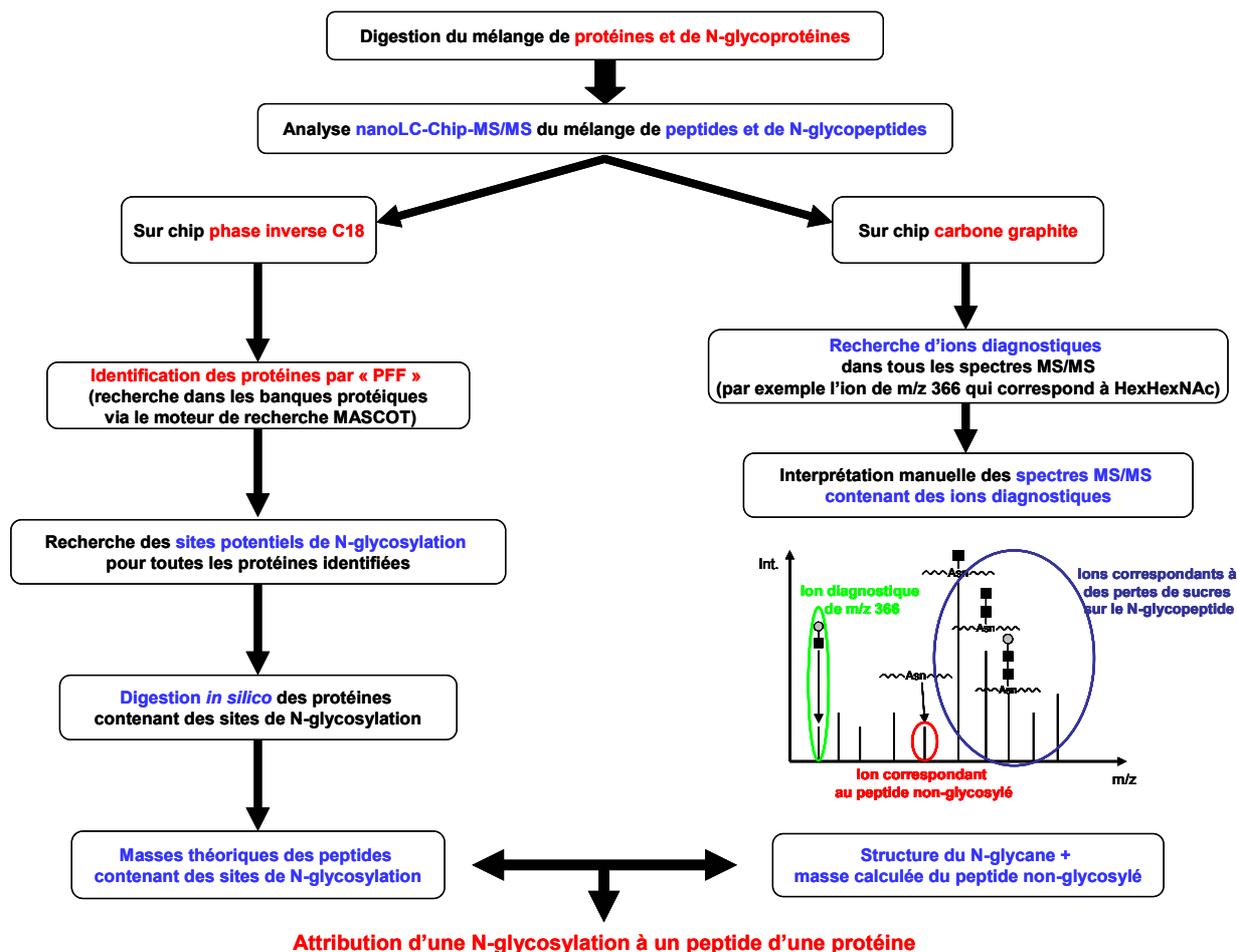


Figure 4 : Illustration de la stratégie pour la caractérisation fine de glycoprotéines basée sur l'analyse nanoLC-Chip-MS/MS de mélanges complexes de peptides.

Cette stratégie a été appliquée aux trois études suivantes menées au laboratoire:

- **Analyse d'anticorps thérapeutiques, les MAbs :** Ce travail consistait à caractériser l'ensemble des glycoformes de la chaîne lourde d'un anticorps monoclonal. Dans le cadre de cette étude la nature de la protéine était connue ainsi que le peptide contenant le site de N-glycosylation. Nous avons de ce fait effectué une étude comparative pour l'analyse de N-glycopeptides sur colonne C18 ou carbone graphite.
- **Le parasite *Toxoplasma gondii* :** La seconde étude consistait à mettre en évidence la présence de protéines N-glycosylées dans le parasite *Toxoplasma gondii*. Les protéines ont été séparées sur gel 1D, digérées, analysées sur colonne C18 pour identifier dans un premier temps les protéines puis analysées sur chips carbone graphite pour identifier des protéines N-glycosylées.
- **Le pollen de l'olivier :** La troisième étude a été menée sur une protéine du pollen de l'olivier, ole e1 qui semble impliquée dans les allergies. Il s'agissait ici d'étudier les différentes formes de cette protéine selon la souche de pollen. Pour cela différentes souches ont été séparées sur

gel 2D, les spots d'intérêt ont été digérés, analysés une première fois sur C18 pour identifier les différentes formes puis sur chips carbone graphite pour identifier les N-glycopeptides. Pour chacune de ces études, les conditions chromatographiques et les paramètres du spectromètre de masse ont été optimisés pour favoriser la séparation ainsi que le détection des N-glycopeptides mais également d'éviter toute fragmentation en source des N-glycopeptides qui fausserait les résultats de caractérisations. Les paramètres de fragmentation ont également été optimisés pour obtenir des spectres MS/MS les plus informatifs.

II. Résultats : Exemple de trois applications

II.1. Analyse d'anticorps thérapeutiques : les MAb

Cette étude a fait l'objet d'une collaboration avec l'équipe d'Alain Beck du « Centre d'immunologie Pierre Fabre » (Saint-Julien-en-Genevois, France). Les résultats ont été publiés en 2008 dans « Journal of Chromatography B ». L'article est disponible en Annexe 1.

II.1.1. Contexte de l'étude

Dans le domaine des protéines recombinantes à visée thérapeutique, les **anticorps monoclonaux** (MAbs) représentent une nouvelle **classe de drogues des plus importantes**. Aujourd'hui plus de 20 anticorps monoclonaux sont approuvés et des centaines sont en phase clinique dans différents domaines tels que l'oncologie, les maladies inflammatoires, les transplantations d'organes, la cardiologie, les infections virales, les allergies, le développement et la réparation des tissus. Leur efficacité résulte de leur capacité à former **un complexe avec une molécule cible** (un antigène) qui va **activer des mécanismes biologique** qui auront comme effet **d'éliminer cette cible** [Carter, 2006], [Jefferis, 2005]. Il existe 5 classes d'anticorps humains, les immunoglobulines G, M, A, D et E. Chacun d'eux ont des structures et des fonctions différentes [Arnold et al., 2007].

a) Une classe de MAb : Les immunoglobulines G (IgG)

La plupart des anticorps monoclonaux approuvés à ce jour font partie de la classe des **immunoglobulines G (IgG)**. Les IgG sont des **glycoprotéines tetramériques** de poids moléculaire d'environ **150 kDa** qui se composent de **deux chaînes polypeptidiques légères** (LC) (≈ 25 kDa) et **deux chaînes polypeptidiques lourdes** (HC) (≈ 50 kDa) qui sont associées de manière covalente (liaisons disulfures) ou non-covalente. L'ensemble forme trois parties protéiques indépendantes connectées par un « linker » flexible, la région pivotante (Hi) (figure 5A) [Jefferis, 2005].

Les LC et les HC sont liées entre elles par une liaison disulfure tandis que les HC sont liées entre elles par deux ou trois liaisons disulfures toutes localisées dans la région pivotante. Les autres ponts disulfures (12 ou 14) sont intramoléculaires et permettent de délimiter six domaines globulaires : un variable (V_L) et un constant (C_L) pour les LC, et un variable (V_H) et trois constants (C_{H1} , C_{H2} , C_{H3}) pour les HC (figure 5B) [Arnold et al., 2007].

Deux de ces parties, les domaines Fab, ont des structures identiques et possèdent chacune un site spécifique de liaison pour un antigène. La troisième partie, le **domaine Fc**, contient des sites **d'interaction pour des ligands** qui **activent les mécanismes d'élimination** de la molécule cible.

Le domaine Fc est un **homodimère** composé de régions pivotantes, de **domaines C_{H2}** non liés et de **domaines C_{H3}** liés de manière non-covalente. Les domaines **C_{H2} sont N-glycosylés** en position **Asn²⁹⁷**. Un MAb contient donc généralement **deux sites de N-glycosylation**.

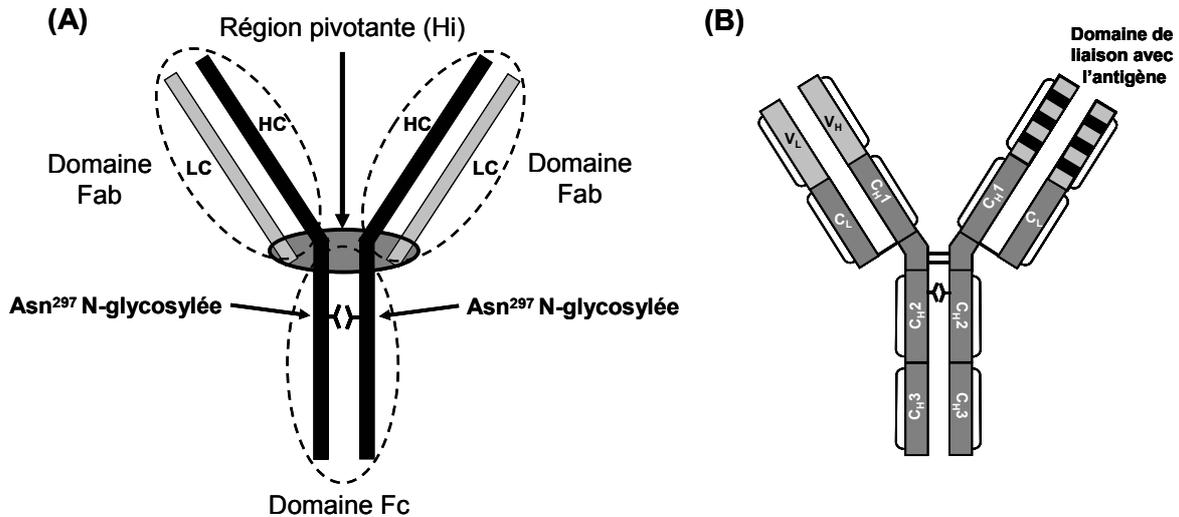


Figure 5 : Structure d'un MAb.

(A) Représentation schématique d'un MAb avec 2 chaînes lourdes (HC) et 2 chaînes légères (LC) reliées entre elles par des liaisons disulfures situées au niveau de la région pivotante. L'ensemble forme trois parties, 2 domaines Fab et un domaine Fc contenant les sites de glycosylation.

(B) Schéma des différents domaines globulaires pour chacune des chaînes lourdes et légères. Le domaine globulaire constant C_{H2} des chaînes lourdes contient l'Asn²⁹⁷ qui est N-glycosylée.

La plupart des MAbs commercialisés sont exprimés dans des lignés de cellules d'ovaires d'hamster chinois (CHO) ou dans des cellules myelomes murines (NS0 ou SP2/0). Dans ce type de cellules, les **N-glycanes** sont de type **bi-antennés** avec des **fucoses** positionnés sur le tronc et un **nombre variable de galactoses** à l'extrémité du N-glycane. Les trois formes majoritaires sont **G0F** (aucun galactose), **G1F** (un galactose) et **G2F** (deux galactoses) (figure 6 de l'article publié disponible en Annexe 1) [Jefferis, 2006].

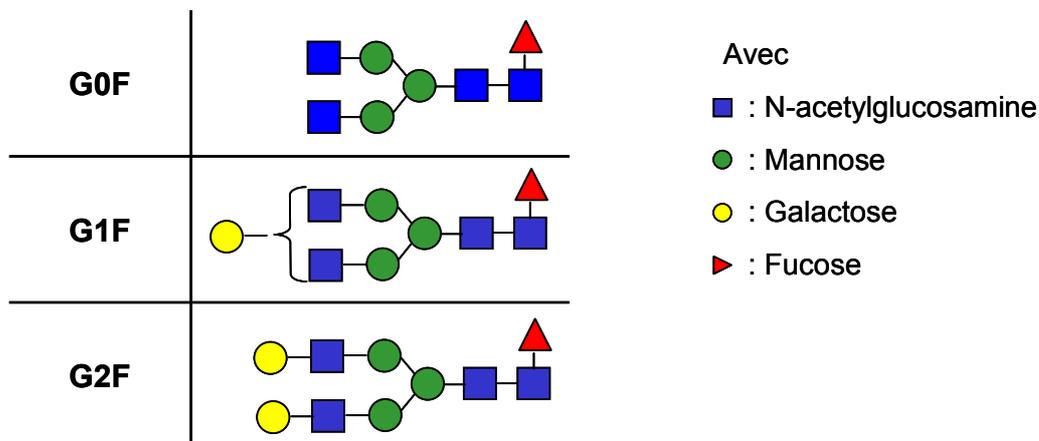


Figure 6 : Structure des trois N-glycanes les plus couramment observés sur les MAbs exprimés dans des lignés de cellules CHO, NS0 ou SP2/0.

b) Rôle de la N-glycosylation des MAbs

La nature de la glycosylation intervient dans les **caractéristiques structurales de l'anticorps** qui sont essentielles pour lier un ligand et activer les mécanismes d'élimination. De ce fait, selon la nature de la N-glycosylation, celle-ci aura un **impact direct sur la sécrétion, la stabilité, l'immunogénicité et l'activité de l'anticorps** [Jefferis, 2005]. Ainsi pour pouvoir contrôler les effets d'un anticorps dans un système biologique, il est aujourd'hui nécessaire de pouvoir définir/contrôler à tout stade de la production des MAbs la structure des N-glycanes formés sur l'anticorps.

Le processus de N-glycosylation étant dépendant de nombreux facteurs tels que : le choix du système de production des IgG, les conditions de culture des cellules, la sélection des clones, dans des conditions non optimales, les cellules peuvent produire des produits glycosylés anormaux qui seront défaillants [Jefferis, 2005]. C'est pourquoi, il est primordial de bien choisir les clones de cellules et les conditions de cultures de celles-ci.

Néanmoins, **à tout stade du développement**, pour des clones de cellules issus d'une même lignée de cellules, il peut y avoir des **divergences dans la composition en glycoformes** des MAbs et c'est pourquoi il est primordiale d'avoir des outils analytiques pour **vérifier la composition exacte** et la **distribution des glycoformes** à tout stade de la recherche, du développement et de la production des MAbs.

c) Objectif : caractérisation complète des MAbs

La spectrométrie de masse est devenue un outil de choix pour la caractérisation des MAbs et à ce jour de nombreuses méthodologies ont été développées. Certaines utilisent des approches simples, d'autres utilisent la combinaison de plusieurs approches pour une caractérisation complète. Ces approches combinent en général l'analyse des glycanes libres, l'analyse de protéines entières (Top-down) et l'analyse de mélanges de peptides (Bottom-up) [Srebalus Barnes and Lim, 2007], [Mechref and Novotny, 2002].

En collaboration avec le Centre d'Immunologie Pierre Fabre nous avons effectué une étude comparative de trois stratégies basées sur la spectrométrie de masse pour caractériser et étudier l'hétérogénéité des MAbs. Les trois stratégies sont les suivantes :

- **Approche Top-down** : Analyse LC-MS sur un ESI-TOF des **MAbs intacts et réduits**.
- La **caractérisation structurale de N-glycanes** libres par ESI-MS/MS sur une trappe à ions.
- **Approche Bottom-up** : Analyses nanoLC-Chip-MS/MS sur une trappe à ions des **N-glycopeptides** de digestion.

II.1.2. Méthodologie

a) Préparation des échantillons

L'ensemble de ces études a été effectué sur des MAbs exprimés dans des lignées de cellules eucaryotes. Les MAbs humanisés A2CHM et A3BHM ont été purifiés à partir du surnageant de cellules CHO et NS0, tandis que le MAb 7C10 a été purifiés d'un hybridome murin. Ces MAbs sont séquencés et leurs sites de glycosylation ainsi que la nature des N-glycanes majoritaires sont bien définis. Les échantillons ont été préparés au Centre d'Immunologie Pierre Fabre. L'ensemble des informations relatives à la préparation des échantillons est disponible dans l'article en annexe 1.

b) Analyses LC-MS des MAbs intacts et réduits

Les MAbs étant des molécules de poids moléculaires de ≈ 150 kDa, leur séparation par HPLC n'est pas chose aisée. Néanmoins, depuis quelques années différentes chimies et conditions chromatographiques ont été explorées pour l'étude des MAbs. C'est le cas par exemple des colonnes de phase inverse [Dillon et al., 2006].

Dans cette étude, les MAbs ont été séparés sur des colonnes de phase inverse C8 en utilisant un système HPLC composé d'un module de séparation Waters Alliance 2695 équipé d'un compartiment thermostaté et d'un détecteur à photodiodes 996. L'analyse par spectrométrie de masse a été effectuée sur un instrument ESI-TOF, le LCT de Waters. Notre choix s'est porté sur un système ESI-TOF car il offre une résolution et une précision de masse très élevée. Des études ont montré qu'en utilisant un instrument d'une résolution de 10000, il est possible d'obtenir une précision de mesure sur la masse d'un anticorps intact de l'ordre de 30 ppm [Gadgil et al., 2006]. L'ensemble des conditions chromatographiques et des paramètres du spectromètre de masse sont disponibles dans la partie expérimentale de l'article (Annexe 1).

c) Analyses par ESI-MS/MS des N-glycanes libres

Comme nous avons pu le voir dans le chapitre 2 de l'Introduction bibliographique (p. 62-64), l'analyse des N-glycanes par des expériences de MSⁿ est cruciale puisqu'elle permet de déterminer la structure fine de ceux-ci.

Néanmoins plusieurs paramètres sont à prendre en compte pour des études les plus complètes. Le premier point important est le choix de l'enzyme de déglycosylation qui idéalement doit décrocher tout type de sucre. Dans cette étude, la nature des glycanes étant connue, notre choix s'est porté sur la PNGase F qui laisse les N-glycanes intactes. Le second point important réside dans leur détection et leur fragmentation par spectrométrie de masse. En effet, selon le mode de fragmentation et l'énergie mise en jeu, le type d'ions fragments générés ne sera pas le même (fragments au niveau des liaisons interglycosidiques ou intracycliques).

Ainsi, pour avoir des spectres MS/MS les plus informatifs, il est souvent nécessaire de dériver les N-glycanes. Ces dérivations permettent également d'améliorer leur ionisation, donc leur détection. Toutefois, ce type d'approche rajoute des étapes de purification qui peuvent entraîner la perte en échantillon [Harvey, 2000a], [Harvey, 2000b].

Dans cette étude nous avons choisi de travailler en infusion sur une trappe à ion (Esquire 3000+, Bruker Daltonics) qui permet de combiner de multiples étapes de MSⁿ. Les N-glycanes libres ont été dessalés puis infusés sans être dérivés ni séparés au préalable. L'ensemble des informations relatives à la préparation des échantillons et les optimisations des paramètres du spectromètre de masse sont disponible dans la partie expérimentale de l'article (Annexe 1).

d) Analyses des N-glycopeptides par nanoLC-Chip-MS/MS

Dans cette étude, nous avons choisi de digérer l'anticorps A2CHM à la trypsine. Cette enzyme permet de générer des glycopeptides détectables dans les gammes de m/z analysables. Le mélange total de peptides et de N-glycopeptides a ensuite été analysé en nanoLC-MS/MS en utilisant le système nanoLC-Chip (Agilent Technologies) couplé à une trappe à ions (HCTplus, Bruker Daltonics). Nous avons utilisé trois types de chips équipées avec trois colonnes différentes : une C18 de 43 mm, une C18 de 150 mm et une carbone graphite de 43 mm. Afin de montrer les performances des trois systèmes, le même mélange de peptides et de N-glycopeptides est analysé sur les trois colonnes en utilisant les mêmes conditions chromatographiques et les mêmes paramètres sur le spectromètre de masse.

L'ensemble des informations relatives à la préparation des échantillons, les conditions chromatographiques et les optimisations des paramètres du spectromètre de masse sont disponible dans la partie expérimentale de l'article (Annexe 1). Tous les spectres MS/MS ont été interprétés manuellement.

II.1.3. Résultats publiés

Les résultats obtenus ont fait l'objet d'une publication dans « Journal of Chromatography B » en 2008 (Annexe 1). Ceux-ci seront brièvement décrits ci-dessous.

Cette étude a permis dans un premier temps de montrer la complémentarité de ces trois stratégies basées sur la spectrométrie de masse comme outil analytique pour une caractérisation fine des différents glycoformes de MAbs mais également pour étudier leur hétérogénéité.

- **L'analyse des MAbs intacts** a permis de mesurer de manière rapide et directe des glycoformes de masses de ≈ 150 kDa dans un mélange très hétérogène avec une précision inférieure à 100 ppm. Ce type d'analyse offre donc la possibilité de caractériser de manière globale un anticorps. Néanmoins, les chaînes lourdes glycosylées étant symétriques, l'analyse directe d'un Mab conduit à un spectre MS très complexe dû à l'hétérogénéité des combinaisons de glycoformes sur les deux sites ne permettant pas de caractériser les glycoformes précisément sur chaque site. Une alternative réside dans l'analyse des **MAbs après réduction des ponts disulfures**. Celle-ci a permis d'améliorer la résolution et de quantifier de manière relative les glycoformes de chaque chaîne lourde.
- **L'analyse ESI-MS/MS de N-glycanes libres** nous a permis de caractériser structurellement les formes glycosylées les plus abondantes d'un Mab qui sont G0F, G1F et G2F. De plus,

l'interprétation du spectre MS/MS de G2F nous a permis mettre en évidence la présence d'une nouvelle forme isobare de G2F. Celle-ci correspond à une forme hybride.

- Enfin, **un mélange de peptides et de N-glycopeptides issu de la digestion à la trypsine d'un MAb, A2CHM, a été analysé par nanoLC-Chip-MS/MS** sur colonnes de phase inverse et sur colonnes de carbone graphite. En comparant les résultats nous avons pu montrer les avantages des colonnes en carbones graphites par rapport aux phases inverse C18 (meilleure rétention des glycopeptides, ionisation améliorée, gain en sensibilité, amélioration de la qualité des spectres MS/MS). Ainsi, nous avons pu caractériser les formes majoritaires G0F, G1F et G2F, la forme isobare de G2F, mais également caractériser et mettre en évidence 8 autres glycoformes minoritaires. Cette approche a su montrer tout son intérêt puisqu'elle permet en une analyse de rapidement détecter les N-glycopeptides à partir des spectres MS/MS en utilisant les ions diagnostiques, de déterminer le site de glycosylation à partir des ions fragments du peptide non-glycosylé et de déterminer l'enchaînement des monosaccharides permettant de remonter à la structure des N-glycanes.

II.1.4. Perspectives

Cette étude nous a permis de montrer la **complémentarité de différentes approches** (top-down et bottom-up) pour la **caractérisation fine des glycoformes d'anticorps monoclonaux**.

En utilisant une nouvelle stratégie basée sur l'analyse des N-glycopeptides par nanoLC-Chip-MS/MS utilisant des chip carbone graphite, il nous a été possible en une analyse de **caractériser 12 glycoformes** et de les **quantifier de manière relative**. Cette approche a su donc montrer tout son intérêt pour un screening en routine rapide et efficace des anticorps monoclonaux durant les différentes étapes de leur production et du développement pharmaceutique.

De plus cette méthodologie sera prochainement utilisée pour caractériser des anticorps produits dans d'autres systèmes tels que des lignées de cellules humaines, des levures, des cellules d'insectes, des plantes transgéniques et bien d'autres.

ANNEXE 1: Publication

The way forward, enhanced characterization of therapeutic antibody glycosylation: Comparison of three level mass spectrometry-based strategies

Elsa Wagner-Rousset, Audrey Bednarczyk, Marie-Claire Bussat, Olivier Colas, Nathalie Corvaia, Christine Schaeffer, Alain Van Dorselaer, Alain Beck, *Journal of Chromatography B*, 2008, 872, 23-27.

**The way forward, enhanced characterization of therapeutic antibody glycosylation
: comparison of three level mass spectrometry-based strategies**

Elsa WAGNER-ROUSSET, Audrey BEDNARCZYK, Marie-Claire BUSSAT, Olivier COLAS,
Nathalie CORVAÏA, Christine SCHAEFFER, Alain VAN DORSSELAER and Alain BECK

Journal of chromatography B, 2008, vol. 872, pages 23-37

Annexe 1 : pages 134-... :

La publication présentée ici dans la thèse est soumise à des droits détenus par un éditeur commercial.

Les utilisateurs de l'ULP peuvent consulter cette publication sur le site de l'éditeur :

<http://dx.doi.org/10.1016/j.jchromb.2008.03.032>

La version imprimée de cette thèse peut être consultée à la bibliothèque ou dans un autre établissement via une demande de prêt entre bibliothèques (PEB) auprès de nos services :

<http://www-sicd.u-strasbg.fr/services/peb/>

II.2. Le parasite *Toxoplasma gondii* : identification de protéines glycosylées

Cette étude a fait l'objet d'une collaboration avec Stanislas Tomavo de l'Unité de Glycobiologie Structurale et Fonctionnelle, CNRS UMR 8576, Université des Sciences et Technologies de Lille, Villeneuve d'Ascq. Les résultats ont été publiés en 2008 dans « Molecular and Cellular Proteomics ». L'article est disponible en Annexe 2.

II.2.1. Contexte de l'étude

Toxoplasma gondii, agent de la toxoplasmose, est un parasite protozoaire intracellulaire. Ce parasite est une des **causes majeures des infections du système nerveux central** chez les patients **atteints par le virus HIV**. Il est également un **pathogène opportuniste** chez les patients ayant subi des traitements contre les cancers, des transplantations d'organes ou encore chez les nouveau-nés qui sont infectés dans l'utérus. Il est un membre du phylum Apicomplexa qui inclut de nombreux autres pathogènes, parmi lesquels *Plasmodium*, un des plus mortel chez l'humain et les animaux d'élevage.

Aujourd'hui le défi réside dans une **meilleure compréhension des mécanismes d'invasion et d'interaction hôte-parasite** pour pouvoir trouver des **traitements efficaces**. Celui-ci passe par une caractérisation des protéines exprimées par ce parasite.

a) Le parasite *Toxoplasma Gondii*

Les Apicomplexa sont des parasites intracellulaires qui doivent vivre à l'intérieur d'une cellule pour survivre. Une fois installé dans la cellule-hôte, celle-ci lui assure de larges ressources en nutriments ainsi qu'une protection contre le système immunitaire de l'hôte [Carruthers, 2002]. Comme toute cellule eucaryote, il possède la machinerie nécessaire à la synthèse et au transport de protéines. Celle-ci comprend un noyau, un réticulum endoplasmique (ER), un appareil de Golgi réticulé ainsi qu'un appareil mitochondrial.

Le parasite *Toxoplasma gondii* se lie et pénètre tout type de cellules mammifères en utilisant une **machinerie moléculaire** unique, le « **Glideosome** », un complexe macromoléculaire composé de **protéines adhésives** qui sont libérées par différents organelles (les micronèmes, les Rhoptries et les granules denses) et **transloquées** à un endroit donné par l'action d'un **système composé de filaments d'actin et de myosins** ancré dans le complexe intermembranaire [Keeley and Soldati, 2004] (figure 7). Une fois à l'intérieure de la cellule hôte, le parasite s'entoure d'une vacuole très résistante dans laquelle il se réplique. Après plusieurs cycles de réplication, les parasites vont émergés de la cellule hôte en utilisant les mêmes mécanismes que pour l'envahir.

Son cycle de vie se compose de trois stades majeurs, le sporozoïte (cellules qui vont infectées un nouvel hôte), le tachyzoïte (forme rapide de réplication) et le bradyzoïte (forme de division lente inactive). Bien que le parasite ait un cycle sexuel dans les cellules hôtes de chats, dans les hôtes mammifères non-félins, il existe sous les deux formes asexuées, le tachyzoïte et le bradyzoïte.

Toxoplasma gondii engendre des infections par ingestion orale de kystes infectés provenant de félins ou présent dans la nourriture mal cuite. Ces kystes très résistants vont survivre durant leur passage dans l'estomac. Une fois arrivés dans l'intestin, ils libèrent des bradyzoïtes. Ces bradyzoïtes vont immédiatement se différencier en tachyzoïtes qui vont s'éparpiller dans tout le corps. La phase aiguë de l'infection se déclenche alors lorsque les tachyzoïtes se répliquent dans différents tissus. Pour résister aux défenses immunitaires des hôtes, les tachyzoïtes vont se différencier en bradyzoïtes qui forment un mur de kystes qui entoure et fortifie le parasite. L'apparition de ces bradyzoïtes entraîne une diminution de la réplication du parasite et marque le début de la phase chronique de l'infection, une phase dormante.

Néanmoins, lorsque le système immunitaire de l'hôte commence à montrer des faiblesses, le parasite reprend sa forme active, les tachyzoïtes, qui se répliquent. Cette réactivation peut se faire dans tout type de tissus mais l'un des sites les plus touché est le système nerveux central. Elle déclenche des encéphalites aiguës chez les patients atteints du SIDA qui peuvent être fatales [Carruthers, 2002].

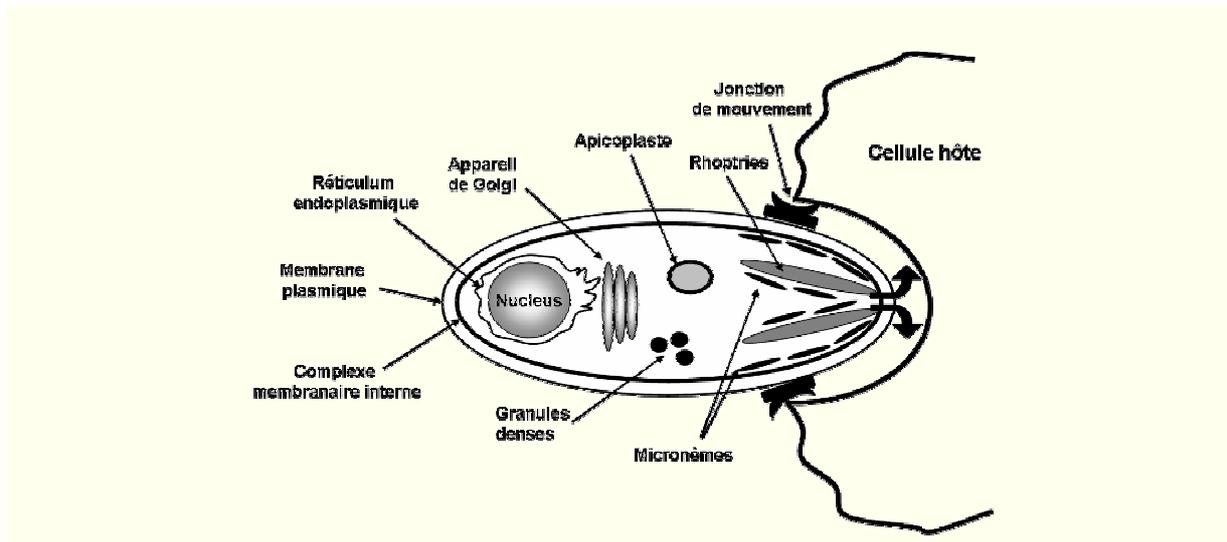


Figure 7 : Schéma représentant le parasite et son mécanisme d'invasion dans la cellule hôte.

A ce jour, la nature et le rôle exact de toutes les protéines exprimées par les différentes organelles sont encore mal définis. Néanmoins, des études biochimiques et génétiques ont permis de suggérer que : les protéines des granules denses semblent modifier la vacuole pour permettre au parasite de récupérer les nutriments de la cellule hôte, les protéines du Rhoptry semblent être impliquées dans la biogenèse et l'association organelle de la vacuole, les protéines des micronèmes semblent avoir une fonction durant la fixation et la pénétration du parasite dans la cellule cible.

b) Les N-glycoprotéines

Les N-glycoprotéines pouvant avoir de nombreuses fonctions biologiques dans la stabilité, le temps de vie, la sécrétion, des hypothèses ont été émises selon lesquelles la glycosylation des protéines du parasite pourrait jouer un **rôle dans la reconnaissance de la cellule hôte et durant son invasion**.

Néanmoins ce type d'hypothèse fut et reste controversé. En effet, dans les années 1980, l'équipe de Hardman démontre que les tachyzoïtes de *Toxoplasma gondii* sont incapables de se lier à des lectines qui sont généralement utilisées pour identifier des glycoprotéines de surface. Au même moment, les équipes de Mauras et Sharma et Johnson détectent la liaison de la concanavalin A à des tachyzoïtes. La présence de glycoprotéines dans le tachyzoïte du parasite fut confirmé dans les années 1993 par l'équipe de Swartz et al. [Odenthal-Schnittler et al., 1993].

c) Objectif : caractériser des protéines N-glycosylées

L'objectif de cette collaboration était **d'étudier et d'élucider le chemin de biosynthèse des N-glycosylations des formes virulentes** (tachyzoïtes) et **avirulentes** (bradyzoïtes) du parasite *Toxoplasma gondii*.

Dans ce but, l'équipe de Stanislas Tomavo a choisi de faire un certain nombre d'expériences biologiques pour montrer la présence des N-glycoprotéines. En parallèle, notre laboratoire s'est appliqué à **identifier ces N-glycoprotéines** par des approches protéomiques classiques (séparation sur gel 1D, digestion des protéines, analyses nanoLC-MS/MS, identification des protéines). Une autre partie du travail a consisté à **identifier les sites de glycosylation** de ces protéines ainsi que la **structure des glycanes** sur ces sites. Pour cela les extraits de mélange de peptides issus de la digestion de protéines ont été analysés par nanoLC-Chip-MS/MS sur colonnes de carbone graphite.

II.2.2. Méthodologie

Concernant les expériences biologiques menées par les collaborateurs, l'ensemble des informations relatives à la préparation des échantillons, l'extraction des protéines, le décrochage des sucres, les western-blots, les expériences d'immunofluorescences, l'enrichissement en glycoprotéines, celles-ci ne seront pas vues en détail ici mais sont décrites en détail dans l'article publié disponible en annexe 2.

Concernant la stratégie protéomique, les protéines ont été séparées sur gel 1D puis deux types d'analyses nanoLC-MS/MS ont été menées. Celles-ci sont brièvement décrites ci-dessous.

a) Séparation des protéines

Les collaborateurs après avoir extraits les protéines de *Toxoplasma gondii*, ont enrichi les extraits en protéines N-glycosylées par passage sur des billes de concanavalin A. Les protéines ont ensuite été séparées sur un gel 1D et colorées au bleu de Coomassie. Les protéines présentes dans les différentes bandes de gel 1D ont été réduites et alkylées, digérées à la trypsine et les peptides ont été extraits des bandes de gel 1D.

b) Identification des protéines par nanoLC-Chip-MS/MS sur colonnes de phase inverse C18

Les bandes de gel 1D pouvant contenir jusqu'à 20 protéines, nous pouvons être en présence de mélanges très complexes de peptides. De ce fait, chaque extrait de peptides a été analysé par nanoLC-Chip-MS/MS sur des chip de 150 mm contenant de la phase C18. Les données de masse ont été interprétées par stratégie « Peptide Fragment Fingerprint » en recherchant dans la banque NCBI via le moteur de recherche MASCOT. Les banques protéiques de *Toxoplasma Gondii* n'étant pas encore disponibles, une stratégie complémentaire a été développée au laboratoire. En parallèle, une base de données du génome complet a été créée au laboratoire et importée dans le moteur de recherche MASCOT. L'ensemble des informations relatives aux conditions chromatographiques utilisées et aux paramètres du spectromètre de masse ainsi que les modes d'interprétation détaillés sont disponibles en annexe 2.

c) Identification de protéines N-glycosylées par nanoLC-Chip-MS/MS sur colonnes de carbone graphite

Chaque extrait de peptides a été analysé par nanoLC-Chip-MS/MS sur des chip de 43 mm contenant de la phase carbone graphite. Les paramètres de transmission des ions ont été optimisés pour les haut m/z . Pour chaque analyse, les courants des ions diagnostiques ont été extraits de tous les spectres MS/MS afin de déterminer quels spectres MS/MS correspondent à ceux de glycopeptides.

Pour chaque protéine identifiée au préalable, les sites potentiels de glycosylation ont été prédits à l'aide d'un logiciel bioinformatique, puis les protéines ont été digérées *in silico* pour déterminer la masse des peptides tryptiques contenant les sites potentiels de glycosylation. Ensuite les spectres MS/MS de glycopeptides ont été interprétés manuellement et validés en tenant compte des trois types d'ions fragments classiquement générés : pertes de sucres sur le glycopeptide, fragments du glycane et présence de l'ion de m/z correspondant au peptide non-glycosylé. L'ensemble des informations relatives aux conditions chromatographiques utilisées et aux paramètres du spectromètre de masse ainsi que les modes d'interprétation détaillés sont disponibles en annexe 2.

II.2.3. Résultats publiés

Les études biologiques menées par les collaborateurs ont permis de mettre en évidence la présence de N-glycosylations de type **high-mannose** sur des protéines qui font partie du **glideosome**, ce complexe macromoléculaire permettant au parasite de pénétrer et de sortir de la cellule hôte. En effet, ces expériences ont permis de montrer la présence de **7 types de N-glycanes** dans *Toxoplasma gondii* (Man₃GlcNAc₂, Man₄GlcNAc₂, Man₅GlcNAc₂, Man₆GlcNAc₂, Man₇GlcNAc₂, Man₈GlcNAc₂, Man₉GlcNAc₂) dont trois sont majoritaires, les N-glycanes Man₆GlcNAc₂, Man₇GlcNAc₂ et Man₈GlcNAc₂.

L'analyse nanoLC-Chip-MS/MS des extraits de peptides issus de la digestion à la trypsine sur colonnes C18 a permis d'**identifier 15 protéines impliquées dans le mécanisme de pénétration/éjection** du parasite dans la cellule hôte. En effet, nous avons pu identifier des protéines telles que la myosin A, la membrane anchor myosine XIV aussi connue sous le nom de GAP-50,

l'actin et la membrane skeleton protein IMC2A [Gaskins et al., 2004], [Meissner et al., 2002]. Nous avons pu également identifier des protéines qui sont sécrétées par les organelles Rhoptries. Certaines de ces protéines sont impliquées dans la formation de la jonction qui permet au parasite de pénétrer la cellule, d'autres sont impliquées dans la formation de la vacuole dans laquelle le parasite se réplique. Toutes ces protéines contiennent des sites consensus de N-glycosylation.

En analysant ces mêmes extraits de peptides par analyse nanoLC-Chip-MS/MS sur colonnes carbone graphite, nous avons pu montrer que la protéine de la famille des myosines, la membrane anchor myosine XIV aussi connue sous le nom de la **GAP-50**, est **N-glycosylée en position Asn¹³⁶** et qu'elle existe sous **deux formes** de type **hautement mannosé**. Une forme porte un N-glycane Man₇GlcNAc₂ et l'autre forme porte un N-glycane Man₈GlcNAc₂. Cette identification est primordiale puisque cette protéine est connue pour ancrer la myosin A, sa chaîne légère associée TgMLC1, l'actin et la TgGAP45 dans le complexe intermembranaire qui forme le glideosome.

II.2.4. Conclusion et perspectives

Dans cette **étude préliminaire** du parasite *Toxoplasma gondii*, l'utilisation du système nanoLC-Chip couplé à la spectrométrie de masse a permis de mettre en évidence la nature de **15 N-glycoprotéines** présentes dans le parasite.

De plus, dans cette étude nous avons tenté de caractériser ces protéines N-glycosylées en déterminant leurs sites de N-glycosylation et la structure des N-glycanes sur ces sites. En utilisant les systèmes nanoLC-Chip-MS/MS avec des colonnes de carbone graphite nous avons pu mettre en évidence **deux formes N-glycosylées de la protéine GAP-50** de type hautement mannosé.

S'agissant d'une étude préliminaire, les protéines ont été digérées à la trypsine, enzyme couramment utilisée. Néanmoins, en étudiant les séquences en acides aminés des protéines identifiées par analyses nanoLC-Chip-MS/MS sur C18, nous avons pu constater que les peptides générés après digestion à la trypsine contenant des sites potentiels de N-glycosylation ont pour la plupart des masses trop élevées pour qu'ils soient détectables dans les gammes de *m/z* du spectromètre de masse utilisé. Ainsi il n'est pas exclu d'être en **présence d'autres protéines glycosylées**.

En **perspective**, il s'agira de pouvoir aller vers une caractérisation fine de **l'ensemble des N-glycoprotéines de *Toxoplasma Gondii***.

Dans ce but, les protocoles de digestions des protéines ainsi que les conditions chromatographiques et les paramètres du spectromètre de masse devront être optimisés afin d'identifier les protéines N-glycosylées, leurs sites de N-glycosylation et les structures des N-glycanes.

Enfin, toutes les **protéines identifiées** dans cette étude étant impliquées dans la survie du parasite mais également dans les mécanismes de pénétration et d'interaction avec la cellule hôte, celles-ci pourraient être utilisées comme nouvelles **cibles thérapeutiques** pour combattre le parasite.

Cependant, pour cela il sera nécessaire **d'explorer les fonctions biologiques** de celles-ci selon la **nature du N-glycane** dans les différents phénomènes tels que le mécanisme de pénétration, la jonction de mouvement, la sécrétion des protéines, les interactions protéines-protéines.

ANNEXE 2: Publication

Proteomic and Glycomic Analysis of N-Glycosylated Structures Involved in *Toxoplasma gondii*-Host Cell Interactions

Syvain Fauquenoy, Willy Morelle, Agnès Hovasse, Audrey, Bednarczyk, Christian Slomianny, Christine Schaeffer, Alain Van Dorsselaer and Stanislas Tomavo, *Molecular and Cellular Proteomics*, 2008, 7, 891-910.

**Proteomics and glycomics analyses of N-glycosylated structures involved in
Toxoplasma gondii-host cell interactions**

Sylvain FAUQUENOY, Willy MORELLE, Agnès HOVASSE, Audrey BEDNARCZYK, Christian SLOMIANNY, Christine SCHAEFFER, Alain VAN DORSSELAER and Stanislas TOMAVO

Molecular & cellular proteomics, 2008, vol. 7, pages 891-910

Annexe 2 : pages 142-... :

La publication présentée ici dans la thèse est soumise à des droits détenus par un éditeur commercial.

Les utilisateurs de l'ULP peuvent consulter cette publication sur le site de l'éditeur :

<http://dx.doi.org/10.1074/mcp.M700391-MCP200>

La version imprimée de cette thèse peut être consultée à la bibliothèque ou dans un autre établissement via une demande de prêt entre bibliothèques (PEB) auprès de nos services :

<http://www-sicd.u-strasbg.fr/services/peb/>

II.3. Le pollen de l'olivier : Etude de la N-glycoprotéine allergène Ole e1 issue de différentes variétés cultivées d'olivier.

Cette étude a fait l'objet d'une collaboration avec Antonio J. Castro de l'Université de Granada, en Espagne (Department of Biochemistry, Cellular and Molecular Biology of Plants, Estación Experimental del Zaidín, CSIC). Un article est en cours de rédaction.

II.3.1. Contexte de l'étude

Les **allergies** font partie des problèmes de santé des plus importants actuellement. On estime que **plus de 20% de la population humaine** est touchée par des allergies. Celles-ci se contractent par ingestion, inhalation ou contact avec des antigènes issus de différentes sources telles que les pollens de plantes, les poils d'animaux ou encore les aliments. Chacune de ces sources contient de nombreux allergènes responsables de cette hypersensibilité contre une espèce particulière. Parmi les allergies les plus répandues, on compte celle **due au pollen de l'olivier** (*Olea europaea* L.) qui provoque des allergies respiratoires importantes chez les populations vivant dans les régions géographiques où cette plante est fortement cultivée, telles que les pays méditerranéens, l'Amérique et l'Australie.

Actuellement, on dénombre **10 allergènes** (Ole e1 à Ole e10) du pollen *Olea europaea* L.. L'allergène **majoritaire** est **Ole e1** et on estime qu'il affecte plus de 70% des patients hypersensibles au pollen de l'oliviers [Villalba et al., 1990], [Rodriguez et al., 2002].

De ce fait, **l'étude des protéines allergènes** du pollen et plus particulièrement **Ole e1** est primordiale pour pouvoir **trouver des traitements** plus efficaces contre ces allergies.

a) La protéine Ole e1

Ole e1 est une **N-glycoprotéine** de pI acide ayant des masses moléculaires variant de 18 à 22 kDa. Sa séquence en acide aminé a été déterminée en 1993 par Villalba et al. [Villalba et al., 1993].

Néanmoins cette protéine présente de **nombreuses micro-hétérogénéité** dans sa séquence nucléotidique [Villalba et al., 1994], [Lombardero et al., 1994] qui entraînent des changements dans la composition en acides aminés de la protéine naturelle [Villalba et al., 1993].

Il a d'ailleurs été montré que ces mutations avaient un **effet** dans la **capacité de l'allergène** à être **reconnu par l'anticorps** [Gonzalez et al., 2002], [Gonzalez et al., 2006]. De plus, ces polymorphismes peuvent dépendre de **l'origine génétique du pollen** [Hamman-Khalifa et al., 2008].

Des études ont montré que selon que les **pollens** proviennent d'une **variété d'olives cultivées** ou d'une autre, leur capacité à se lier à un anticorps IgE sera variable et de ce fait, la **réponse allergénique va différer**.

De même des études quantitatives ont permis de montrer que **selon la variété d'olives, l'abondance de la protéine Ole e1** n'est **pas la même**. Il est donc tout à fait envisageable que la quantité de cet allergène soit impliquée dans la variation de la réactivité des patients à une variété d'olives cultivées [Alche et al., 2007], [Castro et al., 2003].

De ce fait il est primordial de pouvoir **étudier les différentes formes mutées** de cette protéine **selon la variété d'olives** dont elle provient pour pouvoir mieux comprendre son **implication** ainsi que ces **effets sur les allergies**.

b) Rôle de la N-glycosylation

La protéine Ole e1 existe sous **deux formes**, la forme **non-glycosylée** et la forme **N-glycosylée** avec un site de glycosylation en position **Asn¹¹¹**. Par des études de déglycosylation à la PNGase F et à l'Endo H, la nature de ces N-glycanes a pu être déterminée. Ole e1 contient deux type de N-glycanes, les **hautement mannosés** et les **types complexes** dans des proportions équivalentes avec des formes mineures contenant un fucose en position $\alpha(1,3)$ du GlcNAc lié à l'Asn et un xylose en position $\beta(1,2)$ d'un mannose du squelette commun à tous les types de N-glycanes (figure 8) [Rodriguez et al., 2002].

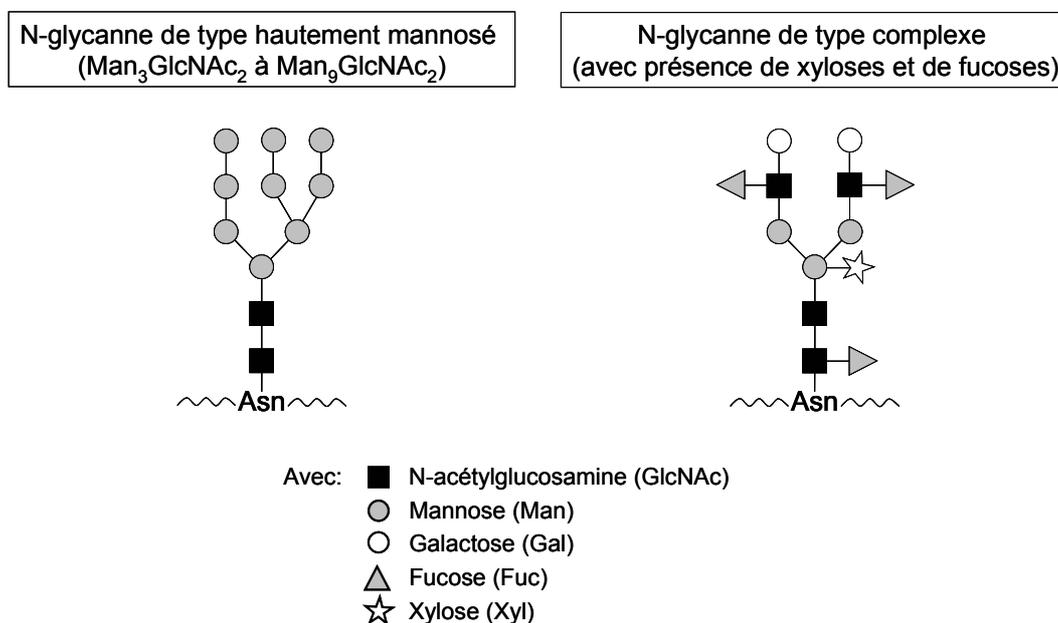


Figure 8 : Structure des deux types de N-glycanes observés sur la protéine Ole e1.

Ces N-glycanes pourraient **être impliquée** dans la **réponse de l'allergène à l'IgE**. En effet, en 1994, Batanero et al. montrent que la N-glycosylation de la protéine Ole e1 est impliquée dans la capacité de celle-ci à être reconnue par le système immunitaire. Ils ont montré qu'en déglycosylant Ole e1 à la PNGase F, celle-ci voyait sa capacité à se lier à un anticorps IgE diminuée. Par opposition, des études de déglycosylation à l'Endo H ont montré qu'il n'y avait pas d'effet sur la fixation de Ole e1 à l'anticorps IgE.

La PNGase F décrochant plus de 95 % des différents types de N-glycanes en les laissant intactes, alors que l'Endo H, beaucoup plus spécifique, ne décroche pas les N-glycanes de type complexe, il a pu en être déduit que les N-glycanes de type complexe jouaient un rôle important dans les allergies

[Batanero et al., 1994]. Plus récemment, des études ont permis de montrer que les fucoses et les xyloses étaient impliqués dans la fixation à l'anticorps IgE [van Ree et al., 2000].

Néanmoins, à l'heure actuelle, même si la nature de ces N-glycanes est connue, leur **structure exacte, le nombre de glycoformes et leur abondance selon la variété de pollen ne sont pas encore connue** et jusqu'à présent aucune étude n'a été réalisée par spectrométrie de masse en tandem pour caractériser les différentes glycoformes de la protéine Ole e1.

C'est pourquoi, aujourd'hui, l'étude des différentes glycoformes de la protéine Ole e1 est particulièrement importante.

c) Objectif : caractériser les mutations de la protéine Ole e1 et les différentes glycoformes dans différentes variétés cultivées d'olivier

L'objectif de cette collaboration était d'identifier les différentes formes (**mutations et glycoformes**) de la protéine **Ole e1** issue du pollen de **différentes variétés cultivées d'olivier**.

Dans ce but, il a été choisi de séparer les protéines sur gels d'électrophorèse 2D. Les spots présents dans la zone des 20 kDa des différents **gels 2D** ont été découpés et analysés par **nanoLC-MS/MS**.

Dans un premier temps, afin d'identifier les différentes **mutations** de Ole e1, les protéines des spots ont été digérées à la trypsine puis analysées par **nanoLC-Chip-MS/MS** sur colonne **C18**. Dans un second temps, afin d'identifier les sites de **N-glycosylation** de ces protéines ainsi que la structure des N-glycanes sur ces sites, ces mêmes extraits de peptides issus de la digestion à la trypsine des protéines ont été analysés par **nanoLC-Chip-MS/MS** sur colonne de **carbone graphite**.

II.3.2. Méthodologie

Concernant les expériences biologiques menées par les collaborateurs, l'ensemble des informations relatives à la préparation des échantillons, l'extraction des protéines, le décrochage des sucres, les western-blots, les expériences d'immunofluorescences, l'enrichissement en glycoprotéines, celles-ci ne seront détaillées dans l'article en cours de rédaction.

Concernant la stratégie protéomique, les protéines ont été séparées sur gel 2D puis deux types d'analyses nanoLC-MS/MS ont été menées. Celles-ci sont brièvement décrites ci-dessous.

a) Séparation des protéines sur gel 2D

Les collaborateurs après avoir extraits les protéines du pollen de trois variétés cultivées d'oliviers (Picual, Arbequina et Bella de Espana), ont séparé celles-ci sur un gel 2D. Afin de vérifier la reproductibilité de la technique séparative, les gels 2D ont été réitérés trois fois de suite.

Dans un premier temps les gels 2D ont été colorés par fluorescence afin de pouvoir effectuer une étude quantitative en utilisant le logiciel PDQuest.

Dans un second temps, une seconde série de gels 2D a été colorée à l'argent pour l'analyse des spots d'intérêt par spectrométrie de masse.

Puis la protéine d'intérêt Ole e1 est localisée dans le gel 2D par des expériences d'immunoblot avec des anticorps dirigés contre Ole e1. Les zones du gel 2D ayant révélé la présence de la protéine Ole e1 sont découpés sur les trois gels 2D colorés à l'argent et les protéines présentes dans les spots ont été réduites, alkylées, digérées à la trypsine puis les peptides ont été extraits des spots de gel.

b) Identification des protéines par nanoLC-Chip-MS/MS sur colonnes de phase inverse C18

Chaque extrait de peptides a été analysé par nanoLC-Chip-MS/MS sur des chip de 43 mm contenant de la phase inverse C18. Les données des masses ont été interprétées en utilisant la stratégie « Peptide Fragment Fingerprint » en recherchant dans la banque NCBIInr via le moteur de recherche MASCOT. Notre choix s'est porté sur cette banque car celle-ci contient à la fois les séquences des protéines issues de la translation automatique des banques nucléotidiques GenBank/EMBL/DDBJ et les séquences protéiques annotées de Swiss-Prot, de PIR (Protein Information Resource), de RefSeq et de PDB (Protein Data Bank). Elle est donc très complète mais présente également des redondances.

L'ensemble des informations relatives aux conditions chromatographiques utilisées et aux paramètres du spectromètre de masse ainsi que les modes d'interprétation détaillés sont disponibles en annexe 3.

c) Identification de protéines N-glycosylées par nanoLC-Chip-MS/MS sur colonnes de carbone graphite

Chaque extrait de peptides a été analysé par nanoLC-Chip-MS/MS sur des chip de 43 mm contenant de la phase carbone graphite. Les paramètres de transmission des ions ont été optimisés pour les m/z élevés. Pour chaque analyse, les courants des ions diagnostiques (m/z 204, 366, 528, ...) ont été extraits de tous les spectres MS/MS afin de déterminer les spectres MS/MS correspondant à des N-glycopeptides.

Pour chaque protéine identifiée au préalable, nous avons recherché les sites potentiels de N-glycosylation puis les protéines ont été digérés *in silico* pour déterminer la masse des peptides tryptiques contenant ces sites potentiels de N-glycosylation. Ensuite les spectres MS/MS de glycopeptides ont été interprétés manuellement et validés en tenant compte des trois types d'ions fragments classiquement générés : pertes de sucres sur le glycopeptide, fragments du glycane et présence de l'ion de m/z correspondant au peptide non-glycosylé. L'ensemble des informations relatives aux conditions expérimentales sera détaillé dans l'article en cours de rédaction.

II.3.3. Résultats en cours de publication

a) Séparation des protéines sur gel 2D

Ces étapes ayant été réalisées chez les collaborateurs, celles-ci ne seront que brièvement décrites ci-dessous.

Les trois extraits protéiques issus des trois variétés d'oliviers (Picual, Arbequina et Bella de España) ont été séparés par électrophorèse sur gel 2D en utilisant les mêmes conditions expérimentales (même quantité déposées, ...). Un des premiers objectifs était de montrer des différences entre les variétés en termes d'abondance et de nombre d'isoformes de la protéine Ole e1.

Dans ce but, les gels 2D correspondant à trois variétés de pollen ont été marqués par fluorescence et quantifiés à l'aide de PDQuest. Il a pu être montré que la variété Picual contenait un pourcentage bien plus élevé de la protéine Ole e1 (figure 9).

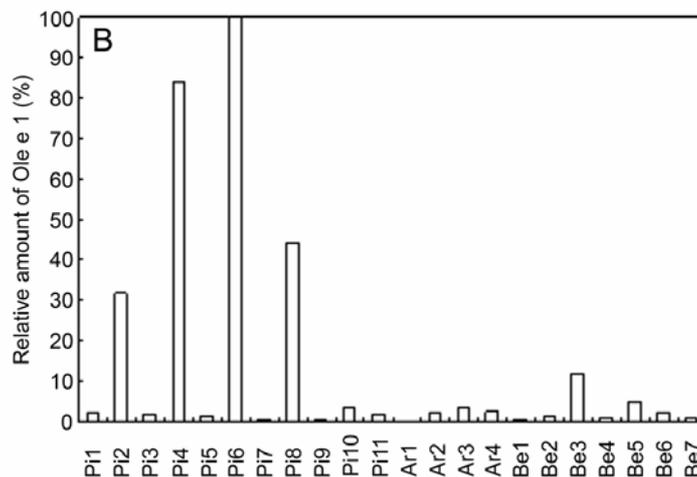


Figure 9 : Abondance relative en % de la protéine Ole e1 dans les différents spots de gel 2D issus des trois variétés de d'oliviers Picual (Pi), Arbequina (Ar) et Bella de España (Be).

Dans un second temps, les expériences de marquage par anticorps a permis de déterminer quels spots des gels 2D contenaient les différentes formes de la protéine Ole e1. Il a pu être constaté que la variété Picual contenait un nombre plus important d'isoformes que dans les variétés Arbequina et Bella de España.

Une fois ces spots identifiés, une autre série de gels 2D a été colorée à l'argent pour pouvoir être analysable par spectrométrie de masse. La figure 10 illustre la comparaison des profils de ces trois gels colorés à l'argent correspondant à trois variétés de pollen de l'olivier ainsi que la zone d'intérêt (15 à 20 kDa) et les spots ayant été découpés pour les analyses nanoLC-Chip-MS/MS.

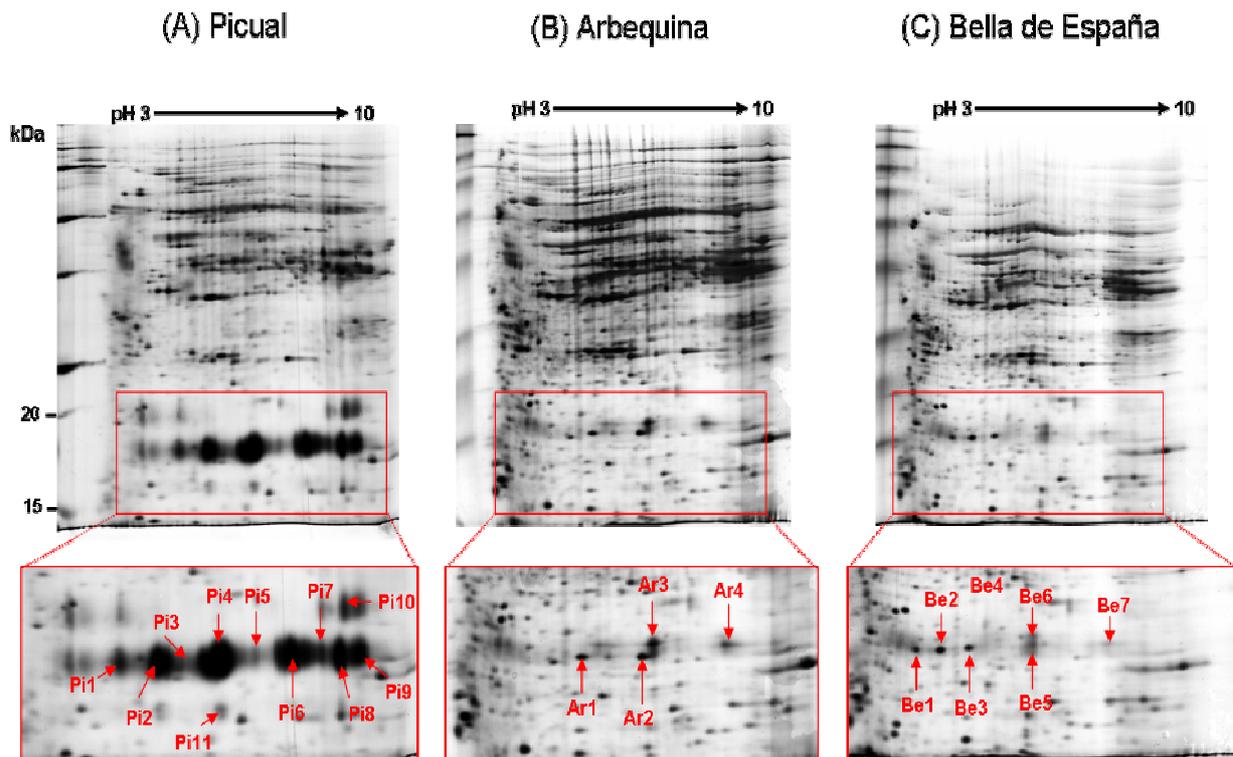


Figure 10 : Comparaison des profils des gels 2D des extraits d'allergènes issus de différentes variétés d'oliviers avec (A) Picual, (B) Arbequina, (C) Bella de España et zoom dans la zone d'intérêt (15 - 20 kDa) pour chacun des gels 2D. Les spots d'intérêt ayant été analysés par nanoLC-Chip-MS/MS sont annotés.

b) Mutations identifiées par analyses nanoLC-Chip-MS/MS

Les extraits de peptides issus des 22 spots ont été analysés par nanoLC-Chip-MS/MS sur des chips contenant de la phase inverse C18. Les données de masses générées ont été interprétées par stratégie PFF en recherchant dans la banque NCBIInr via le moteur de recherche MASCOT.

Dans la mesure où les séquences en acides aminés des différentes formes de la protéine Ole e1 ne sont que prédites à partir du gène et qu'elles n'ont jamais pu être vérifiées, nous avons choisi de rechercher dans une banque contenant à la fois les séquences déduites de la translation automatique des banques nucléotidiques et les séquences annotées dans Swiss-Prot.

Après exploitation des résultats des recherches, nous avons pu identifier un certain nombre de mutations sur la protéine Ole e1 mais également montrer les différences dans ces mutations selon la variété d'oliviers étudiée.

En effet, pour chacun des spots nous avons pu identifier jusqu'à 10 formes différentes de la protéine Ole e1 avec des spectres de fragmentation MS/MS de très bonne qualité. A titre d'exemple, nous présenterons dans la figure 11 deux spectres MS/MS correspondants à un peptide contenant une mutation dans sa séquence en acides aminés. Ces deux spectres sont issus de l'analyse nanoLC-Chip-MS/MS de l'extrait de peptides issu du spot Pi1.

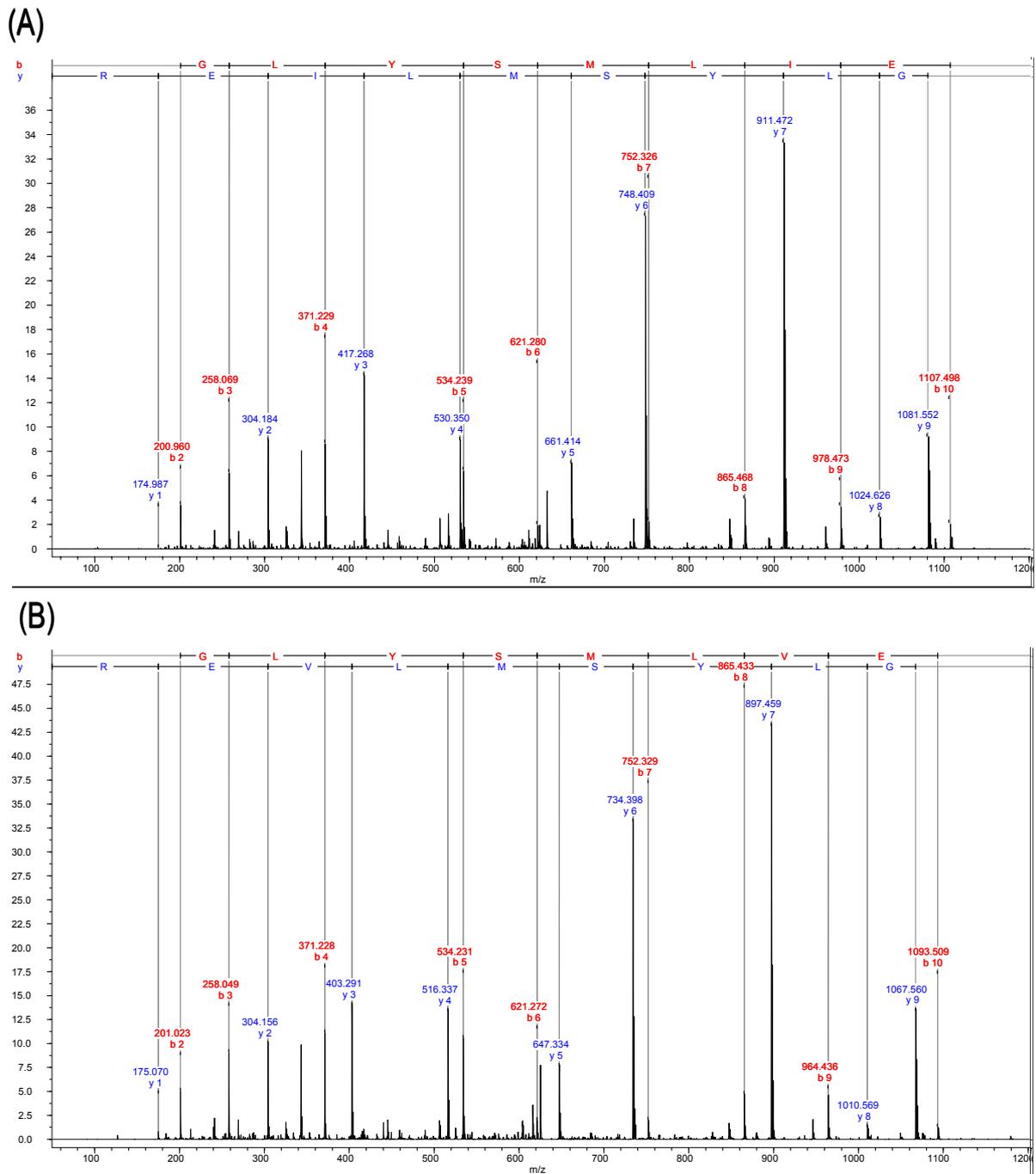


Figure 11 : Comparaison des spectres MS/MS de deux peptides de Ole e1 portant la mutation I↔V en position 69 avec (A) spectre du peptide AEGLYSMLIER et (B) spectre du peptide AEGLYSMLYER. Les deux spectres sont issus de l'analyse nanoLC-Chip-MS/MS de Pi1.

En combinant l'ensemble des résultats obtenus par analyse nanoLC-Chip-MS/MS pour chacune des variétés, nous avons pu identifier au total 30 mutations sur la protéine Ole e1. Le tableau 2 regroupe la liste des 36 peptides issus de la digestion à la trypsine identifiés avec un ou plusieurs acides aminés modifiés et dans quelles variétés d'oliviers ceux-ci sont identifiés.

Peptides	Position/Variant	Variétés		
		Picual	Bella de España	Arbequina
AGFITELSEFIPGASVR	28↔28/T↔N	+	+	+
AGFITELSEFIPGASLR	30↔30/L↔F	+	+	+
AGFINELSEFIPGASVR	38↔38/S↔G	+	+	-
AGFITELSEFIPGAGVR	39↔39/V↔L	+	-	-
FITELSEFIPGAGVR		+	+	+
FITEFSEFIPGAGVR		+	+	-
DGENGSITFTEVGYTR	45↔45/D↔E	+	-	-
DGENGKITFTEVGYTR	46↔46/G↔K	+	+	-
DKENGDTVTFTEVGYTR	47↔47/E↔K	+	+	+
DKENGDTVTFTEIGYTR	50↔50/S↔D↔K	+	+	+
EKENGDTFTEIGYTR	51↔51/I↔V	+	+	+
ENGDITFTEVGYTR	56↔56/ I↔V	+	-	-
KNGDITFTEVGYTR		+	+	+
AEGLYSMLVER	69↔69/I↔V	+	+	+
AEGLYSMLIER		+	+	+
DHKNEFCEITLISSGR	75↔75/N↔D	+	+	+
DHKNEFCEITLLSSSR	83↔83/I↔L	+	+	+
DHKNEFCEITLISSGSK	86↔86/G↔S	+	+	+
DHKDEFCEITLISSSR	87↔87/R↔S	-	+	-
KDCDEIPVEGWVKPSLK	91↔91/D↔N	+	+	?
KDCDEIPIEGWAKPSLK	95↔95/I↔V↔T	+	+	?
KDCDEIPTEGWVKPSVK	99↔99/V↔A↔G	+	-	?
KDCNEIPTEGWAKPSLK	103↔103/V↔L	+	-	?
KDCNEIPTEGWVKPSLK		+	-	?
DCNEIPTEGWGKPSLK		+	-	?
FKLNTVNGTTR^{c)}	106↔106/M↔K ↔I	+	+	+
FILNTVNGTTR	108↔108/N↔T	+	+	+
FILNLTNGTTR^{c)}	110↔110/V↔L	+	+	+
FILSTVNGTTR		-	-	-
FMLNTVNGTTR		+	-	-
TINPLGFFKK	117↔117/I↔V	+	?	?
TINPLGFYKK	123↔123/F↔Y	+	?	?
TVNPLGFFKK		+	?	?
TVNPLGFYKK		+	?	?
CAQVYNK	132↔132/A↔P	+	?	?
CPQVFNK	135↔135/F↔Y	+	?	?

Tableau 2 : Liste des peptides de Ole e1 identifiés après analyse nanoLC-Chip-MS/MS de spots de gels 2D issus de différentes variétés cultivées d'oliviers. Les acides aminés présentant une micro-hétérogénéité sont en gras.

La présence (+) ou l'absence (-) des différents peptides dans les variétés est indiquée. Dans le cas où aucune autre variante du peptide n'a pas été identifiée nous avons mis un (?).

En comparant les peptides identifiés selon les variétés d'oliviers cultivées, on constate que certaines mutations ne sont identifiées que dans une variété, c'est le cas par exemple de l'acide aminé en position 3 qui n'est identifié que dans Picual. Certains de ces N-glycanes ne sont d'ailleurs identifiés que dans Picual, d'autres ne sont identifiés que dans Bella de España.

c) *N-glycopeptides identifiés par analyses nanoLC-Chip-MS/MS*

Après avoir identifié les différentes formes de la protéine Ole e1, nous nous sommes intéressés à la recherche de N-glycosylations sur celles-ci. Dans ce but, les mêmes extraits de peptides issus de la digestion à la trypsine des différents spots ont alors été analysés par nanoLC-Chip-MS/MS sur des colonnes contenant de la phase de carbone graphite.

Avant d'exploiter manuellement les spectres de fragmentation MS/MS, et afin de faciliter la recherche des spectres de fragmentation des N-glycopeptides, nous avons dans un premier temps recherché et calculé les masses théoriques des peptides issus de la digestion à la trypsine contenant le site de N-glycosylation Asn¹¹¹ en tenant compte des acides aminés pouvant être modifiés en position 106 et 110. Nous avons en tout 5 peptides de séquences en acides aminés légèrement différentes qui contiennent le site de N-glycosylation Asn¹¹¹. Parmi ces 5 peptides, l'un contient une lysine. De ce fait, dans le cas de la digestion à la trypsine nous pouvons avoir le peptide LNTVNGTTR ou le peptide avec une coupure manquée FKLNTVNGTTR.

Comme nous avons pu le voir dans l'Introduction bibliographique, généralement les spectres de fragmentation MS/MS de N-glycopeptides contiennent trois types d'ions fragments : des ions correspondant à des pertes de sucres sur le glycopeptide, des ions correspondant à des fragments de sucres et des ions correspondant au peptide non glycosylé. De plus, dans un spectre MS/MS de N-glycopeptide, nous avons souvent un ion très intense qui correspond au peptide avec un GlcNAc lié à l'asparagine.

De ce fait, avant la recherche et l'exploitation manuelle des spectres de fragmentation MS/MS des N-glycopeptides et afin de faciliter les identifications, nous avons calculé, pour chacun de ces 5 peptides possibles, les masses théoriques des ions monochargés correspondants au peptide sans N-glycane et au peptide contenant un GlcNAc lié sur l'Asn¹¹¹ (tableau 3).

Séquence des peptides	Masse monoisotopique M (Da)	Masse monoisotopique de l'ion monochargé (M+H) ⁺ (Da)	Masse monoisotopique de l'ion monochargé (M+H+GlcNAc) ⁺ (Da)
LNTVNGTTR	974,51	975,52	1178,52
FKLNTVNGTTR	1249,67	1250,68	1453,68
FILNTVNGTTR	1234,67	1235,68	1438,68
FILNTLNGTTR	1248,68	1249,69	1452,69
FMLNTVNGTTR	1252,62	1253,63	1456,63

Tableau 3 : Liste de peptides contenant le site de glycosylation en Asn¹¹¹ issus de la digestion à la trypsine de la protéines Ole e1 en tenant compte des possibles modifications des acides aminés en position 106 et 110 avec : leurs séquences en acides aminés et les masses monoisotopiques théoriques du peptide, du peptide monochargé et du peptide monochargé avec un GlcNAc lié à l'Asn.

Une fois ces masses théoriques calculées, nous avons extrait de tous les spectres MS/MS les courants des ions fragments (M+H+GlcNAc)⁺. Ceci nous a permis de rapidement localiser les spectres

MS/MS de N-glycopeptides et d'avoir une première idée de la micro-hétérogénéité de la N-glycosylation pour chaque peptide possible.

De plus, dans chacun des spectres de fragmentation MS/MS, la présence d'ions diagnostiques correspondant à des fragments de sucres a également permis d'attribuer ces spectres à des N-glycopeptides. A titre d'exemple nous présentons le spectre de fragmentation MS/MS de l'ion doublement chargé de rapport m/z 1102,5 qui après interprétation a permis d'identifier le peptide LNTVNGTTR portant le N-glycane de type complexe $\text{GlcNAcMan}_3\text{XylGlcNAc}_2$ en Asn^{111} (figure 12).

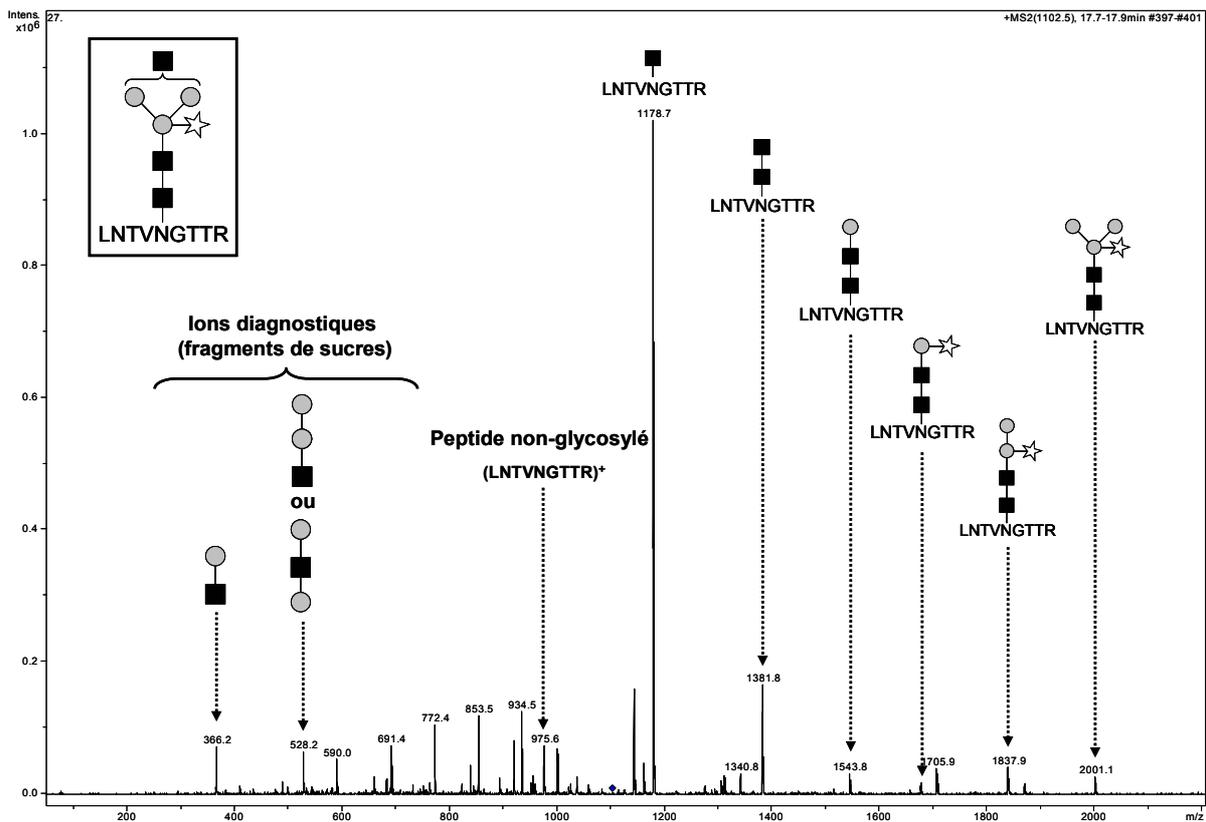


Figure 12 : Spectres de fragmentation MS/MS de l'ion parent doublement chargé de rapport m/z 1102,5 correspondant au peptide LNTVNGTTR portant le N-glycane de type complexe $\text{GlcNAcMan}_3\text{XylGlcNAc}_2$ en Asn^{111} . On observe les pertes successives de sucres sur le peptide glycosylé, la masse de l'ion monochargé correspondant au peptide non-glycosylé et des fragments de sucres, le tout permettant de remonter à la structure du N-glycane et la nature du peptide. Légende : carré noir, N-acétylglucosamine (GlcNAc) ; rond gris, Mannose (Man) ; étoile blanche, Xylose (Xyl).

En procédant de cette manière, dans certains spots de gel 2D, nous avons pu identifier plus de 20 glycoformes en tenant compte de la nature du peptide N-glycosylé. Par exemple, dans Pi1, nous avons identifiés les 4 types de peptides chacun portant 2 à 7 N-glycane différents.

Pour certains spots, nous n'avons identifié que très peu de peptides N-glycosylés (1 à 4), c'est le cas dans les spots Be1, Be2, Be3, Be4, Ar1, Ar2 et Ar4. Enfin dans le spot Pi11, nous n'avons identifié aucun peptide N-glycosylé.

Ainsi nous avons pu montrer dans un premier temps que les différentes formes de la protéine Ole e1 identifiées dans la variété d'olivier Picual présentent une micro-hétérogénéité plus élevée au niveau de leurs glycoformes. En effet, 12 N-glycanes différents ont été identifiés dans la variété Picual contre 10 dans Bella de España et 6 dans Arbequina.

En combinant l'ensemble des résultats obtenus pour chacune des variétés nous avons pu identifier 14 types de N-glycanes différents en tout qui se regroupe en deux types : les hautement mannosés et ceux de type complexe. Le tableau 4 énumère la liste des N-glycanes identifiés et dans quelles variétés d'oliviers ceux-ci sont identifiés.

N-glycanes	Variétés		
	Picual	Bella de España	Arbequina
Man ₄ GlcNAc ₂	+	-	-
Man ₅ GlcNAc ₂	+	-	+
Man ₆ GlcNAc ₂	+	+	-
Man ₇ GlcNAc ₂	+	+	+
Man ₈ GlcNAc ₂	+	+	+
Man ₉ GlcNAc ₂	+	-	-
GlcNAcMan ₃ GlcNAc ₂	+	-	-
GlcNAcMan ₃ GlcNAc ₂ + 180	+	+	-
GlcNAcMan ₃ XylGlcNAc ₂ + 2x180	-	+	-
GlcNAcMan ₃ XylGlcNAc ₂	+	+	+
GlcNAc ₂ Man ₃ XylGlcNAc ₂	+	+	+
GlcNAcMan ₃ XylGlcNAcFucGlcNAc	+	+	+
GlcNAc ₂ Man ₃ XylGlcNAcFucGlcNAc	+	+	-
Man ₃ XylGlcNAcFucGlcNAc	-	+	-

Tableau 4 : Liste des N-glycanes identifiés après analyse nanoLC-Chip-MS/MS de spots de gels 2D issus de différentes variétés cultivées d'oliviers. La présence (+) ou l'absence (-) des différents N-glycanes dans les variétés est indiquée. L'ajout de 180 indique qu'un résidu mannose (180,156 Da), glucose (180,160) ou galactose (180,080) est présent sur le N-glycane.

En comparant les N-glycanes identifiés selon les variétés d'oliviers cultivées, on constate que la variété Picual est celle qui contient le plus de N-glycanes différents. Certains de ces N-glycanes ne sont d'ailleurs identifiés que dans Picual, d'autres ne sont identifiés que dans Bella de España.

II.3.4. Conclusion et perspectives

Dans cette étude l'utilisation d'un système nanoLC-MS/MS plus sensible et plus résolutif a su montrer tout son intérêt.

En effet, l'objectif de cette étude était d'identifier d'une part les différentes formes de la protéine Ole e1 dans trois variétés d'oliviers mais également de déterminer la nature des N-glycanes pour chacune des formes.

De ce fait, cette étude différentielle des protéines des pollens de **trois variétés cultivés d'oliviers** a été réalisée en utilisant une stratégie basée sur la séparation des protéines sur gels 2D. Les différentes formes de **Ole e1** pouvant être **faiblement représentées** dans des extraits protéiques, les **gels 2D ont été colorés à l'argent**, coloration beaucoup plus sensible, puisqu'elle permet de détecter des quantités de l'ordre de 1 à 10 ng. Néanmoins, l'étude des extraits de peptides issus de spots colorés à l'argent par analyse nanoLC-MS/MS peut poser problèmes notamment dans le cas de spots faiblement colorés contenant peu de protéines.

Toutefois, dans notre cas, en utilisant le système **nanoLC-Chip** plus sensible, nous avons pu obtenir de très bons résultats d'autant plus que les **extraits de peptides ont été divisés en deux** et qu'une partie a été analysée sur phase inverse **C18** et l'autre partie sur colonne de **carbone graphite**.

Ainsi, nous avons pu, d'une part, identifier **30 acides aminés mutés** et **14 N-glycanes différents** au total, et d'autre part nous avons également montré des **différences dans la nature de Ole e1 selon les variétés cultivées d'oliviers**.

Cette étude constitue la première ayant utilisé l'analyse protéomique pour montrer la forte hétérogénéité de la protéine **Ole e1** tant au niveau des acides aminés mutés qu'au niveau des N-glycanes dans les pollens de plusieurs variétés d'oliviers.

Toutefois, à l'heure actuelle, même si nous avons pu identifier un certain nombre d'acides aminés mutés et plusieurs types de N-glycanes, il ne nous a pas été possible de caractériser finement les formes mutées de **Ole e1** (un ou plusieurs acides aminés mutés sur une forme) dû aux fortes homologies de séquences en acides aminés. De même, il ne nous a pas été possible de caractériser finement chaque glycoforme de **Ole e1 (quels N-glycanes en position Asn¹¹¹ sur quelle forme mutée)**.

En perspective, il s'agira de pouvoir aller vers une **caractérisation plus fine de ces différentes formes de la protéine Ole e1** notamment en utilisant des marquages spécifiques des parties N-terminales ou C-terminales qui sont spécifiques de chaque forme.

En parallèle, il sera également intéressant d'appliquer cette méthodologie à **d'autres variétés d'oliviers**. L'objectif ultime étant de pouvoir trouver des **traitements plus efficaces** contre les **allergies aux pollens**.

III. Conclusion

De manière générale, la **caractérisation complète d'une N-glycoprotéine** nécessite de combiner plusieurs approches qui sont l'**analyse MS des N-glycoprotéines** entières, l'**analyse MS et MS/MS des N-glycanes libres** et l'**analyse MS, MS/MS et MS³ de N-glycopeptides**. L'ensemble de ces données va permettre de déterminer le **nombre de sites de N-glycosylation**, la **structure** du ou des **N-glycanes** sur ces différents sites et la **nature du ou des sites N-glycosylés** (position de l'asparagine dans la séquence en acides aminés de la protéine).

Toutefois, les N-glycoprotéines sont souvent **peu abondantes** par rapport aux autres protéines et présentent une **forte hétérogénéité** dans la nature de leur N-glycosylation. De ce fait, pour améliorer leur détection et leur caractérisation par spectrométrie de masse, il est souvent nécessaire d'utiliser de nombreuses **étapes de purification**, d'**enrichissement** et de **marquage chimiques** qui peuvent entraîner la **perte d'échantillon**.

De même, le **choix des enzymes** permettant de **décrocher les sucres** et des **enzymes** permettant de **digérer** les N-glycoprotéines afin d'obtenir des N-glycopeptides ne contenant chacun qu'un site de N-glycosylation est également très important et il n'est pas rare de constater que pour aller vers une caractérisation fine des N-glycoprotéines il faut combiner l'utilisation de plusieurs enzymes.

Toutes ces stratégies sont bien sûr plus faciles à utiliser dans le cas de la caractérisation d'une protéine N-glycosylée contenant uniquement un site de N-glycosylation occupé. En revanche dans le **cas de mélanges complexes de N-glycoprotéines** contenant **plusieurs sites de N-glycosylation** occupés par **différents N-glycanes**, l'utilisation de ces stratégies va rapidement devenir compliqué notamment lorsqu'il s'agira de déterminer quel site de N-glycosylation est occupé par quels N-glycanes.

De ce fait, aujourd'hui, les nouveaux besoins résident dans la mise au point de méthodologies permettant de déterminer de manière plus aisée quels sites de N-glycosylation sont occupés par combien et quels N-glycanes.

Ainsi, avec l'acquisition au laboratoire d'un nouveau système de chromatographie liquide, la **nanoLC-Chip** et la commercialisation de chip contenant de la **phase carbone graphite** (Agilent Technologies), nous avons travaillé sur l'analyse nanoLC-Chip-MS/MS de mélange de peptides et de N-glycopeptides sans passer par ces nombreuses étapes de purification et d'enrichissement des N-glycopeptides.

L'utilisation de ce système en **couplage avec une trappe à ions HCTultra** (Bruker Daltonics) a su montrer tout son intérêt pour l'identification et la caractérisation de N-glycoprotéines.

En effet, les **trois études** réalisés en utilisant ce type de système pour l'analyse nanoLC-Chip-MS/MS de N-glycopeptides nous ont permis de :

- Déterminer la **structure** de l'ensemble des **N-glycanes majoritaires** et **minoritaire** présents sur le site de N-glycosylation d'un **anticorps monoclonal** mais également d'étudier leur **abondance relative**.

- Identifier pour la **première fois un site de N-glycosylation** occupé par deux N-glycanes sur une **protéine de *Toxoplasma Gondii*** à partir de l'analyse de mélanges de peptides et de N-glycopeptides issus de la digestion des protéines présentes dans des bandes de **gels 1D**.
- Identifier les **différentes formes d'une protéine N-glycosylée**, de déterminer la **structure des N-glycanes** présents sur ces différentes formes et la **nature des sites de N-glycosylations** occupés à partir de l'analyse de mélanges de peptides et de N-glycopeptides issus de la digestion des protéines présentes dans des spots de **gels 2D colorés à l'argent**.

Même si dans les études réalisées ici, nous avons essentiellement travailler sur des protéines contenant un site de N-glycosylation occupé et que la digestion à la trypsine permettait de générer des peptides qui une fois N-glycosylés ont des masses détectables dans les gammes de m/z analysables par la trappe à ion, cette approche semble être très prometteuse pour l'étude de mélanges complexes de N-glycoprotéines contenant plusieurs sites de N-glycosylation.

En effet, en combinant la sensibilité et la résolution chromatographique des systèmes nanoLC-Chip avec les propriétés des colonnes de phase carbone graphite, il est possible de déterminer la nature de l'ensemble des N-glycanes présents sur un site et de déterminer la nature du site en digérant simplement les mélanges complexes de protéines et en les analysant par nanoLC-Chip-MS/MS.

BIBLIOGRAPHIE

Alche J.D., Castro A.J., Jimenez-Lopez J.C., Morales S., Zafra A., Hamman-Khalifa A.M. and Rodriguez-Garcia M.I. (2007). Differential characteristics of olive pollen from different cultivars: biological and clinical implications. *J Investig Allergol Clin Immunol* **17 Suppl 1**, 69-75.

Arnold J.N., Wormald M.R., Sim R.B., Rudd P.M. and Dwek R.A. (2007). The impact of glycosylation on the biological function and structure of human immunoglobulins. *Annu Rev Immunol* **25**, 21-50.

Batanero E., Villalba M. and Rodriguez R. (1994). Glycosylation site of the major allergen from olive tree pollen. Allergenic implications of the carbohydrate moiety. *Mol Immunol* **31**, 31-37.

Carruthers V.B. (2002). Host cell invasion by the opportunistic pathogen *Toxoplasma gondii*. *Acta Trop* **81**, 111-122.

Carter P.J. (2006). Potent antibody therapeutics by design. *Nat Rev Immunol* **6**, 343-357.

Castro A.J., de Dios Alche J., Cuevas J., Romero P.J., Alche V. and Rodriguez-Garcia M.I. (2003). Pollen from different olive tree cultivars contains varying amounts of the major allergen Ole e 1. *Int Arch Allergy Immunol* **131**, 164-173.

Davies M., Smith K.D., Harbin A.M. and Hounsell E.F. (1992). High-performance liquid chromatography of oligosaccharide alditols and glycopeptides on a graphitized carbon column. *J Chromatogr* **609**, 125-131.

Davies M.J., Smith K.D., Carruthers R.A., Chai W., Lawson A.M. and Hounsell E.F. (1993). Use of a porous graphitised carbon column for the high-performance liquid chromatography of oligosaccharides, alditols and glycopeptides with subsequent mass spectrometry analysis. *J Chromatogr* **646**, 317-326.

Dillon T.M., Bondarenko P.V., Rehder D.S., Pipes G.D., Kleemann G.R. and Ricci M.S. (2006). Optimization of a reversed-phase high-performance liquid chromatography/mass spectrometry method for characterizing recombinant antibody heterogeneity and stability. *J Chromatogr A* **1120**, 112-120.

Forgacs E. (2002). Retention characteristics and practical applications of carbon sorbents. *J Chromatogr A* **975**, 229-243.

Gadgil H.S., Pipes G.D., Dillon T.M., Treuheit M.J. and Bondarenko P.V. (2006). Improving mass accuracy of high performance liquid chromatography/electrospray ionization time-of-flight mass spectrometry of intact antibodies. *J Am Soc Mass Spectrom* **17**, 867-872.

Gaskins E., Gilk S., DeVore N., Mann T., Ward G. and Beckers C. (2004). Identification of the membrane receptor of a class XIV myosin in *Toxoplasma gondii*. *J Cell Biol* **165**, 383-393.

Gonzalez E.M., Villalba M., Lombardero M., Aalbers M., van Ree R. and Rodriguez R. (2002). Influence of the 3D-conformation, glycan component and microheterogeneity on the epitope structure of Ole e 1, the major olive allergen. Use of recombinant isoforms and specific monoclonal antibodies as immunological tools. *Mol Immunol* **39**, 93-101.

Gonzalez E.M., Villalba M., Quiralte J., Batanero E., Roncal F., Albar J.P. and Rodriguez R. (2006). Analysis of IgE and IgG B-cell immunodominant regions of Ole e 1, the main allergen from olive pollen. *Mol Immunol* **43**, 570-578.

Hamman-Khalifa A., Castro A.J., Jimenez-Lopez J.C., Rodriguez-Garcia M.I. and Alche J.D. (2008). Olive cultivar origin is a major cause of polymorphism for Ole e 1 pollen allergen. *BMC Plant Biol* **8**, 10.

- Hanai T.** (2003). Separation of polar compounds using carbon columns. *J Chromatogr A* **989**, 183-196.
- Hanai T.** (2004). Analysis of the mechanism of retention on graphitic carbon by a computational chemical method. *J Chromatogr A* **1030**, 13-16.
- Harvey D.J.** (2000a). Collision-induced fragmentation of underivatized N-linked carbohydrates ionized by electrospray. *J Mass Spectrom* **35**, 1178-1190.
- Harvey D.J.** (2000b). Electrospray mass spectrometry and fragmentation of N-linked carbohydrates derivatized at the reducing terminus. *J Am Soc Mass Spectrom* **11**, 900-915.
- Jefferis R.** (2005). Glycosylation of recombinant antibody therapeutics. *Biotechnol Prog* **21**, 11-16.
- Jefferis R.** (2006). Bioprocessing and Biopartnering, Touch Briefings.
- Kawasaki N., Ohta M., Hyuga S., Hashimoto O. and Hayakawa T.** (1999). Analysis of carbohydrate heterogeneity in a glycoprotein using liquid chromatography/mass spectrometry and liquid chromatography with tandem mass spectrometry. *Anal Biochem* **269**, 297-303.
- Keeley A. and Soldati D.** (2004). The glideosome: a molecular machine powering motility and host-cell invasion by Apicomplexa. *Trends Cell Biol* **14**, 528-532.
- Koizumi K.** (1996). High-performance liquid chromatographic separation of carbohydrates on graphitized carbon columns. *J Chromatogr A* **720**, 119-126.
- Koizumi K., Tanimoto T., Okada Y. and Matsuo M.** (1990). Isolation and characterization of three positional isomers of diglycosylcyclomaltoheptaose. *Carbohydr Res* **201**, 125-134.
- Lombardero M., Barbas J.A., Moscoso del Prado J. and Carreira J.** (1994). cDNA sequence analysis of the main olive allergen, Ole e I. *Clin Exp Allergy* **24**, 765-770.
- Mechref Y. and Novotny M.V.** (2002). Structural investigations of glycoconjugates at high sensitivity. *Chem Rev* **102**, 321-369.
- Meissner M., Schluter D. and Soldati D.** (2002). Role of *Toxoplasma gondii* myosin A in powering parasite gliding and host cell invasion. *Science* **298**, 837-840.
- Ninonuevo M., An H., Yin H., Killeen K., Grimm R., Ward R., German B. and Lebrilla C.** (2005). Nanoliquid chromatography-mass spectrometry of oligosaccharides employing graphitized carbon chromatography on microchip with a high-accuracy mass analyzer. *Electrophoresis* **26**, 3641-3649.
- Odenthal-Schnittler M., Tomavo S., Becker D., Dubremetz J.F. and Schwarz R.T.** (1993). Evidence for N-linked glycosylation in *Toxoplasma gondii*. *Biochem J* **291 (Pt 3)**, 713-721.
- Perkins P.D.** (2006). Glycoprotein Characterization: Using an HPLC-Chip System with Trap XCT Ultra MS for the Identification and Relative Distribution of N-linked Glycans. Agilent Technologies Application Note 5989-5157EN.
- Rodriguez R., Villalba M., Batanero E., Gonzalez E.M., Monsalve R.I., Huecas S., Tejera M.L. and Ledesma A.** (2002). Allergenic diversity of the olive pollen. *Allergy* **57 Suppl 71**, 6-16.
- Srebalus Barnes C.A. and Lim A.** (2007). Applications of mass spectrometry for the structural characterization of recombinant protein pharmaceuticals. *Mass Spectrom Rev* **26**, 370-388.
- van Ree R., Cabanes-Macheteau M., Akkerdaas J., Milazzo J.P., Loutelier-Bourhis C., Rayon C., Villalba M., Koppelman S., Aalberse R., Rodriguez R., Faye L. and Lerouge P.** (2000). Beta(1,2)-xylose and alpha(1,3)-fucose residues have a strong contribution in IgE binding to plant glycoallergens. *J Biol Chem* **275**, 11451-11458.

Villalba M., Batanero E., Lopez-Otin C., Sanchez L.M., Monsalve R.I., Gonzalez de la Pena M.A., Lahoz C. and Rodriguez R. (1993). The amino acid sequence of Ole e I, the major allergen from olive tree (*Olea europaea*) pollen. *Eur J Biochem* **216**, 863-869.

Villalba M., Batanero E., Monsalve R.I., Gonzalez de la Pena M.A., Lahoz C. and Rodriguez R. (1994). Cloning and expression of Ole e I, the major allergen from olive tree pollen. Polymorphism analysis and tissue specificity. *J Biol Chem* **269**, 15217-15222.

Villalba M., Lopez-Otin C., Martin-Orozco E., Monsalve R.I., Palomino P., Lahoz C. and Rodriguez R. (1990). Isolation of three allergenic fractions of the major allergen from *Olea europea* pollen and N-terminal amino acid sequence. *Biochem Biophys Res Commun* **172**, 523-528.

Wuhrer M., Deelder A.M. and Hokke C.H. (2005). Protein glycosylation analysis by liquid chromatography-mass spectrometry. *J Chromatogr B Analyt Technol Biomed Life Sci* **825**, 124-133.

Yin H. and Killeen K. (2007). The fundamental aspects and applications of Agilent HPLC-Chip. *J Sep Sci* **30**, 1427-1434.

Yin H., Killeen K., Brennen R., Sobek D., Werlich M. and van de Goor T. (2005). Microfluidic chip for peptide analysis with an integrated HPLC column, sample enrichment column, and nanoelectrospray tip. *Anal Chem* **77**, 527-533.

CHAPITRE 3

ETUDE DU PROTEOME DU CHEVEU HUMAIN

Cette étude a fait l'objet d'une collaboration avec L'ORÉAL, Paris, qui nous ont fourni les extraits de protéines du cheveu. L'ensemble des résultats ont donnés lieu à (1) un article publié en 2008 dans « Journal of Investigative Dermatology » (Annexe 4) et (2) un article en cours de rédaction qui présente les différents aspects protéomiques utilisés pour la caractérisation fine des protéines du cheveu.

I. Introduction

Depuis quelques années, l'étude des **protéines** du cheveu a pris un essor considérable notamment dans les domaines de la cosmétique [Dawber, 1996], [Dawber, 2002], [Popescu and Hocker, 2007].

En effet, les protéines sont les **composants majoritaires du cheveu** (60-95 %), les pourcentages restant étant des lipides, de l'eau, des pigments et des métaux dont les taux varient d'un type de cheveu à un autre [Wolfram, 2003]. Celles-ci subissent des **modifications au cours du temps** durant la **maturation du cheveu** en étant soumises à des **traitements naturels ou chimiques**. Ces modifications peuvent entraîner des **dégradations irréversibles du cheveu**.

Ainsi, pour éviter celles-ci, il est aujourd'hui nécessaire de **trouver des traitements efficaces**, qui, pour être mis au point, nécessitent d'une part de connaître la **composition** et la **nature exacte des protéines présentes dans le cheveu à maturation** ainsi que les **interactions** qui existent entre elles dans un **cheveu naturel** mais également dans un cheveu **ayant subi des traitements**.

De ce fait, nous avons utilisé l'analyse protéomique pour **caractériser l'ensemble des protéines** d'un **cheveu naturel** ce qui inclut l'identification de **modification post-traductionnelles**, la **localisation** des protéines et l'étude des **interactions** entre ces protéines.

I.1. Le cheveu

La structure du cheveu se divise en deux grands groupes, la **cuticule**, qui forme la **gaine externe** du cheveu et qui le protège des attaques extérieures et, le **cortex** qui est la **partie interne** du cheveu (figure 1).

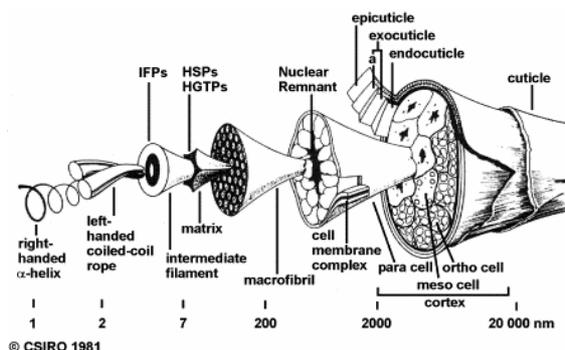


Figure 1 : Schéma d'un cheveu montrant le cortex et la cuticule [Plowman, 2003].

Le **cortex** se constitue de plusieurs cellules, les **cellules corticales**. Chacune de ces cellules contient **5 à 8 macrofibrilles** qui se composent elles-mêmes de **500 à 800 filaments intermédiaires** (IFs) qui sont entourés d'une **matrice protéique**.

L'ensemble forme un **complexe compacte très stable** par de nombreuses **liaisons inter- et intramoléculaires** de type disulfures, hydrogènes, ponts salins ou amides [Hearle, 2000] qui confère au cheveu sa rigidité, son élasticité et sa résistance aux attaques extérieures telles que les UV, la chaleur ou encore les traitements chimiques [Wolfram, 2003].

I.1.1. Les filaments intermédiaires (IFs)

Les **filaments intermédiaires** sont les composés clés dans la structure du cheveu. La formation des filaments intermédiaires découle directement de l'**association complexe de kératines**, appelées aussi les kératines des filaments intermédiaires (**KIFs**) en **polypeptides très hétérogènes**.

Les KIFs sont des protéines constituées de **chaînes polypeptidiques** qui possèdent une **tige centrale** de conformation en hélice alpha (le « rod domain ») et **deux extrémités N- et C-terminales** non-hélicales appelées les domaines « **tête** » (head) et « **queue** » (tail). Leur **conformation** en hélice alpha est due à des **liaisons de type hydrogène** entre les groupements chimiques polaires NH et CO de la liaison peptidique et à la **nature de leur séquence en acides aminés**. En effet, les séquences en acides aminés des kératines sont constituées de séquences répétitives de sept résidus, un heptade **abcdefg**, avec **a** et **d** qui sont impérativement des résidus hydrophobes (Phe, Ile, Leu, Tyr, Val, Met,

Ala) [Popescu and Hocker, 2007]. Leur nature et leur séquence en acides aminés ont été établies fin des années 1990 par Rogers et al. Aujourd'hui, on dénombre **15 KIFs** dans le cheveu qui se divisent en 2 familles selon leurs propriétés physico-chimiques :

- **Les kératines acides de Type I** : on compte **9 protéines** K31, K32, K33a, K33b, K34, K35, K36, K37 et K38. Celles-ci ont des pI acides (~ 5) et des PM 44-48 kDa [Langbein et al., 1999].
- **Les kératines neutres-basiques de Type II** : on compte **6 protéines** K81, K82, K83, K84, K85, K86. Celles-ci ont des pI neutres/basiques (> 6-7) et des PM 55-60 kDa [Langbein et al., 2001].

Pour chacune des deux familles, la **tige centrale** se constitue de **307 acides aminés** avec de **fortes homologies de séquences** (70% pour les KIFs K32, K35, K36 ; 90% pour les KIFs K31, K33a, K33b, K34, K37, K38 ; 70% pour les KIFs K82, K84, K85 ; 90% pour les KIFs K81, K83, K86) et des acides aminés majoritaires communs [Langbein and Schweizer, 2005], [Parry et al., 2006], [Plowman et al., 2002]. Concernant les domaines « **tête** » et « **queue** », riches en **cystéines**, leur taille et la nature des acides aminés ainsi que les séquences sont variables (tableau 1).

Protéines	Domaines de la protéine					
	Tête (N-terminal)		Tige centrale		Queue (C-terminal)	
	Nombre d'acides aminés	Acides aminés majoritaires	Nombre d'acides aminés	Acides aminés majoritaires	Nombre d'acides aminés	Acides aminés majoritaires
K81	106	R/C/G/P/S	307	A/E/L/K	92	C/G/S/N
K82	120	R/G/P/S/L/V	307	A/E/L/K	86	G/L/P/S/N
K83	111	R/C/G/P/S	307	A/E/L/K	75	C/G/S/N/L
K84	165	R/G/S/N	307	A/E/L/K	128	G/L/P/S/T/V
K85	123	R/C/G/P/S/N	307	A/E/L/K	77	R/C/G/S/N
K86	106	R/C/G/P/S	307	A/E/L/K	73	C/G/S/N

Protéines	Domaines de la protéine					
	Tête (N-terminal)		Tige centrale		Queue (C-terminal)	
	Nombre d'acides aminés	Acides aminés majoritaires	Nombre d'acides aminés	Acides aminés majoritaires	Nombre d'acides aminés	Acides aminés majoritaires
K31	56	N/C/P/S	307	A/E/L/Q	53	A/N/C/P/S/T
K32	96	A/C/G/L/P/S	307	A/E/L/Q	45	C/P/S/T/V
K33a	56	N/C/G/P/S	307	A/E/L/Q	41	A/N/C/G/P/S/T
K33b	56	N/C/G/P/S	307	A/E/L/Q	41	N/C/G/P/S/T
K34	56	N/C/G/P/S	307	A/E/L/Q	31	N/C/G/P/S/T
K35	67	A/C/G/L/P/S	307	A/E/L/Q	51	A/C/P/S
K36	93	A/C/G/L/S/T	307	A/E/L/Q	67	R/C/P/S/T/V
K37	104	A/C/G/L/P/S/T	307	A/E/L/Q	34	C/G/P/S/T
K38	104	A/C/G/L/P/S	307	A/E/L/Q	45	C/G/P/S/T

Tableau 1 : Illustration du nombre d'acides aminés et de la nature des acides aminés majoritaires dans les trois domaines (tête, tige centrale, queue) pour (A) les kératines basiques de type II et (B) les kératines acides de Type I (<http://www.expasy.org/tools/protparam.html>).

Leur caractère hydrophobe ainsi que leur géométrie en chaînes polypeptidiques leur permet de s'associer en dimères via les acides aminés hydrophobes. En effet, les résidus hydrophobes répétés vont se placer dans l'hélice alpha de telle sorte à pouvoir interagir avec une autre hélice alpha.

Ainsi, une kératine acide et une kératine basique vont former des dimères hélicoïdaux parallèles (le « coiled-coil dimer ») en s'enroulant l'une autour de l'autre au niveau de l'axe de la tige centrale. On définit alors plusieurs domaines dans cette tige centrale, les segments hélicoïdaux 1A, 1B, 2A et 2B qui sont connectés via des chaînes élastiques, les « linkers » L1, L12 et L2 (figure 2).

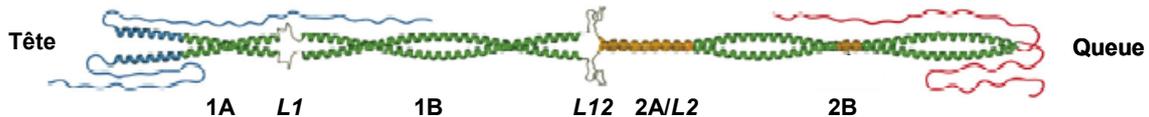


Figure 2 : Modèle atomique d'un dimère de structure hélicoïdale. Il est formé par l'appariement en parallèle à l'axe des tiges centrales de 2 chaînes de conformation en hélice alpha.

Ces dimères s'associent alors en tétramères via des liaisons inter- et intramoléculaires par chevauchement selon quatre modes possibles : (1) le mode A11 qui correspond au chevauchement antiparallèle de 2 segments 1B, (2) le mode A22 qui correspond au chevauchement antiparallèle de 2 segments 2B, (3) le mode A12 qui correspond au chevauchement antiparallèle complet des dimères et (4) le mode ACN qui correspond au chevauchement entre la partie N- et C-terminale (figure 3) [Parry et al., 2007].

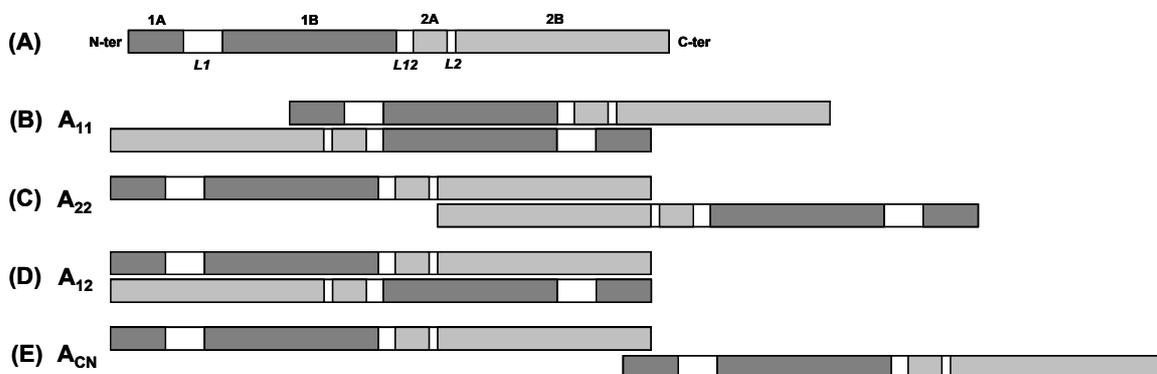


Figure 3: Illustration schématique (A) des domaines d'un dimère et mode d'association de ceux-ci en tétramères avec (B) le mode antiparallèle A₁₁, (C) le mode antiparallèle A₂₂, (D) le mode antiparallèle A₁₂ et (E) le mode parallèle A_{CN} [Parry et al., 2007].

Enfin, ces tétramères s'associent via les domaines tête et queue pour former des polymères très hétérogènes qu'on appelle des protofilaments, 8 protofilaments (ou 4 protofibrilles) donnant un filament intermédiaire [Jones, 2001], [Popescu and Hocker, 2007], [Hearle, 2007]. La figure 4 résume les différentes étapes de la formation des filaments intermédiaires ainsi que leur implication dans la structure du cheveu.

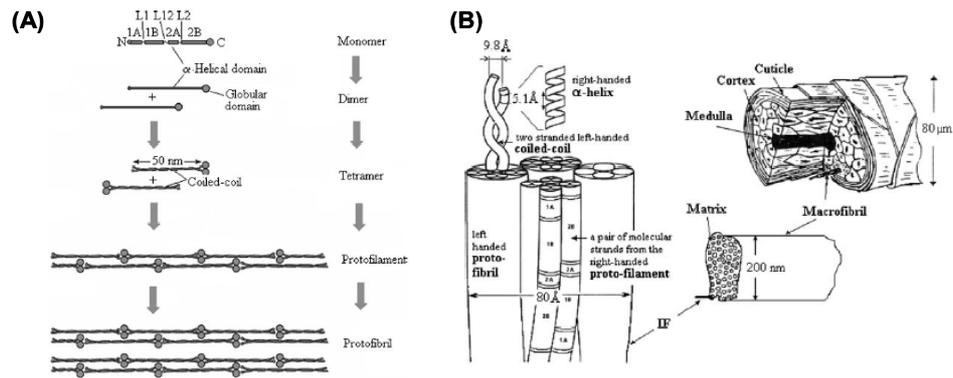


Figure 4 : Schéma illustrant les différentes étapes de la formation des IFs avec (A) la formation des protofibrilles à partir de l'association de monomères, les kératines (KIFs) et (B) l'association des protofibrilles pour donner les IFs présents dans le cortex [Popescu and Hocker, 2007].

Aujourd'hui, même si on sait que les kératines s'associent en dimères, tétramères puis en protofilaments, la formation de ces **polymères très hétérogènes** est encore mal comprise en ce qui concerne les **modes d'associations** et le **type de liaisons** qui entrent en jeu pour **consolider ce complexe**. On sait certes qu'il doit y avoir des liaisons inter- et intramoléculaires de type hydrogènes, disulfures, des liaisons amides entre glutamines et lysines ou encore des ponts salins mais **leur position exacte et leur nature exacte ne sont pas encore bien définies** et de nombreuses études ont été réalisées pour apporter de nouvelles réponses [Parry et al., 2007], [Smith and Parry, 2007], [Smith and Parry, 2008].

I.1.2. Les protéines de la matrice

Chaque filament intermédiaire est entouré d'une matrice protectrice constituée de protéines, les protéines associées aux kératines (KAPs). Leur nature et leur séquence en acides aminés ont été étudiées pour la première fois fin des années 1990 par Rogers et al. et aujourd'hui, on estime qu'il existe **plus de 80 KAPs**. Ce sont des protéines très **riches en cystéines, glycines et tyrosines** [Rogers et al., 2006].

De ce fait, selon leur nature, elles se divisent en **3 groupes** dans chacun desquels on retrouve **23 familles** qui se composent elles-mêmes de plusieurs protéines avec de **fortes homologies de séquences en acides aminés** :

- **High sulfur proteins (HSPs):** 11 familles (KAP 1 à 3, KAP 10 à 16, KAP 23). Celles-ci ont des PM de 10-15 kDa et un pourcentage de cystéines par protéine < 30.
- **Ultra high sulfur proteins (UHSPs):** 4 familles (KAP 4, KAP 5, KAP 9 et KAP 17). Celles-ci ont des PM de 15-50 kDa et un pourcentage de cystéines par protéine > 30.
- **High glycine tyrosine proteins (HGTPs):** 8 familles (KAP 6 à 8 et KAP 18 à 22). Celles-ci ont des PM de 6-9 kDa et un pourcentage de glycines et tyrosines variant de 35 à 60 %.

Néanmoins, leur rôle exacte et la manière dont elles interagissent entre elles ou avec les IFs est encore très mal défini et plusieurs modèles sont admis [Hearle, 2000]:

- **Le modèle F94** : En 1994, Feughelman suppose que les KAPs forment des amas globulaires hydrophobes avec une surface hydrophile, l'ensemble étant entouré d'eau, vu ici comme un polymère. Ainsi, l'ensemble formerait un échafaudage impénétrable constitué de deux systèmes polymériques.
- **Le modèle C/H** : Dans les années 1970, Chapman et Hearle supposent que les KAPs forment une sorte de caoutchouc isolant entre les filaments intermédiaires via des ponts disulfures inter- et intramoléculaires entre KAPs et avec les filaments intermédiaires. Les IFs possédant ces cystéines plutôt dans les domaines terminaux, ce modèle suppose que ces liaisons entre les IFs et les KAPs se font au niveau de la tête et de la queue.

I.1.3. Expression des protéines dans le follicule pileux

Depuis plusieurs années, Langbein, Rogers, Shimomura et al. se sont employés à montrer le taux d'expression des ARNm des kératines et des KAPs dans le follicule du cheveu [Langbein et al., 1999], [Langbein et al., 2001], [Langbein and Schweizer, 2005], [Rogers et al., 2006]. Ainsi aujourd'hui on connaît la nature des protéines majoritairement présentes dans le cortex et la cuticule du follicule pileux (figure 5).

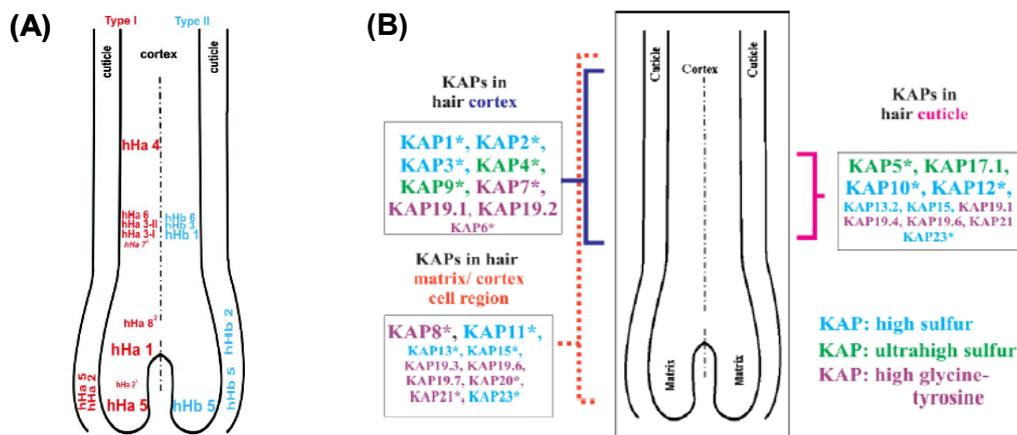


Figure 5 : Schéma représentant le taux d'expression de (A) les kératines et (B) les KAPs dans le follicule pileux selon Langbein, Rogers, Shimomura et al. [Langbein et al., 1999], [Langbein et al., 2001], [Langbein and Schweizer, 2005], [Rogers et al., 2006].

Dans la partie matrice du cortex à la base du follicule pileux, les kératines K31, K35 et K85 sont majoritairement exprimées. Puis au cours de la pousse du cheveu, dans le cortex, on trouve K33a, K33b et K36 pour les acides et K86, K81 et K83 pour les basiques. Encore plus tard dans la maturation, on trouve K34 fortement exprimée. Enfin concernant K37 et K38 elles sont faiblement exprimées dans le cortex. Dans la cuticule, on ne trouve que 4 kératines exprimées à la base du follicule, K32, K35, K82 et K85.

Concernant les KAPs, ils ont essentiellement fait des mesures globales des taux d'expression d'ARNm par familles de KAPs. Ainsi on constate qu'on retrouve majoritairement dans le cortex les familles 1, 2, 3, 4, 9, 7, 6 et dans la cuticule les familles 5, 10, 12, 15, 21, 23. Enfin à la base du follicule, dans la matrice du cortex, on trouve les familles riches en soufre 11, 13, 15, 23 et les familles riches en glycines et tyrosines 8, 20, 21, 23. Les familles très riches en soufres semblent être plutôt exprimées pendant la maturation.

Ces informations sont primordiales puisqu'elles permettent de suggérer que lors de la formation des filaments intermédiaires à la base du follicule, les kératines K31, K35 et K85 vont plutôt interagir avec des KAPs riches en soufre et riches en glycine et tyrosine et qu'au cours de la maturation, on a formation des KAPs très riches en soufres qui vont interagir avec les kératines en augmentant le nombre de ponts disulfures pour consolider ce complexe moléculaire compacte.

Néanmoins, ce ne sont que des hypothèses et pour l'instant aucune expérience d'expression des ARNm n'a été réalisée sur des cheveux matures.

I.2. L'analyse protéomique des protéines du cheveu à maturation : un défi

Les protéines du cheveu possèdent des propriétés physico-chimiques qui sont peu compatibles avec des approches classiques en analyse protéomique.

En effet, dans un premiers temps les **kératines sont difficiles à extraire et à solubiliser** dans des tampons compatibles avec l'électrophorèse sur gel ou la chromatographie liquide. Cela est dû à leur nature en acides aminés très hydrophobes, leur structure et aux interactions protéines-protéines ou intra protéiques. De plus les KIFs étant les protéines majoritaires dans le cheveu, la **détection des protéines minoritaires n'est pas aisée**.

Concernant les **KAPs**, celles-ci sont d'une part très riches en cystéines (> 30%) (présence de nombreux ponts disulfures intra- et intermoléculaires entre kératines et KAPs) et d'autre part elles contiennent très **peu de lysines et d'arginines** qui correspondent aux sites de coupures de la trypsine.

De ce fait, très peu de peptides analysables par spectrométrie de masse sont générés. De plus ces peptides sont souvent communs à une famille de KAPs. Enfin, les KAPs ayant des **pI basiques et des poids moléculaires peu élevés**, leur détection sur gels 2D est difficile.

Finalement, les protéines du cheveu présentent de **fortes homologies de séquences**, ce qui rend l'interprétation des données de masse difficiles [Plowman et al., 2002], [Parry et al., 2006].

Actuellement, très peu d'études protéomiques des protéines du cheveu sont décrites dans la littérature. Les premières études réalisées sur ce type de protéines étaient basées sur la séparation par gels d'électrophorèse et ont été réalisées dans les années 1980 sur les protéines de la laine [Marshall and Blagrove, 1979], [Plowman, 2003].

Par la suite, des études similaires ont été réalisées sur les protéines du cheveu humain. Les premières kératines identifiées le furent par électrophorèse bidimensionnelle et analyse Western-blot en utilisant des anticorps spécifiques pour chaque membre des familles de kératines (Type I et Type II). Ces gels 2D pauvres en nombre de spots, montraient un profil très similaire à celui observé pour les protéines de la laine : (1) une longue traînée dans la zone de PM 62 kDa et les gammes de pI 4-7

correspondant aux KIFs de Type II et (2) un chapelet dans les pI acides correspondant aux KIFs de Type I (figure 6).

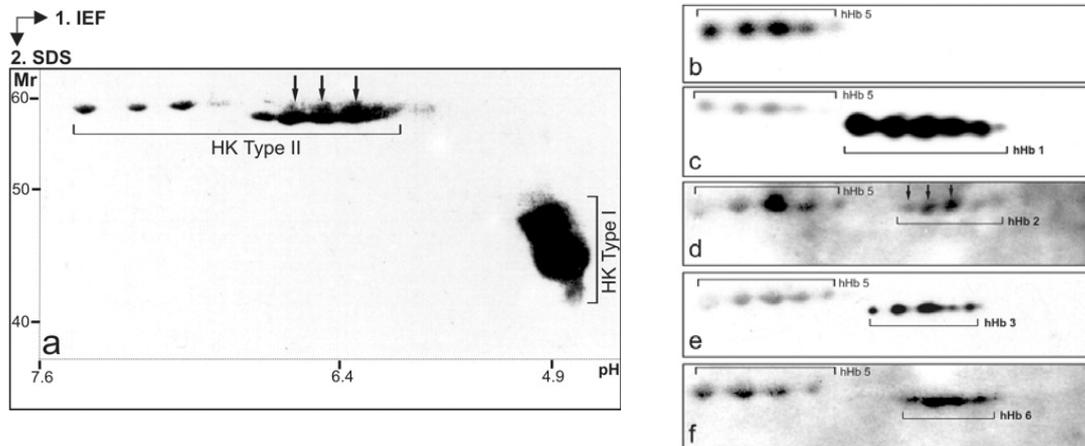


Figure 6 : Kératines identifiées à partir d'un gel d'électrophorèse bidimensionnel et d'analyses Western-Blot avec (a) le profil complet du gel marqué au bleu de coomassie, et les résultats obtenus par Western-Blot sur ce même gel 2D avec des anticorps dirigés contre (b)K85, (c) K85 et K81, (d) K85 et K82, (e) K85 et K83, (f) K85 et K86 [Langbein et al., 1999], [Langbein et al., 2001].

Néanmoins, même si ces études ont permis de montrer la forte expression de certaines KIFs de Type I (K31, K33a, K33b, K34 et K35) et de Type II (K81, K82, K83, K85 et K86), aucune analyse n'a été faite par spectrométrie de masse afin d'identifier les différentes formes de ces protéines [Langbein et al., 1999], [Langbein et al., 2001]. En effet, si l'on se réfère à la figure 6b, on constate que la kératine K85 (hHb5) est identifiée dans un chapelet de spots alignés. Ceci suppose que nous avons plusieurs formes de cette protéine qui entraînent des différences dans la migration selon le pI.

Plus récemment, une alternative fut proposée utilisant la chromatographie liquide bidimensionnelle en séparant un mélange complexe de peptides issus de la digestion à la trypsine de l'extrait total des protéines du cheveu. Les peptides sont séparés sur une colonne échangeuse de cations (SCX) puis sur une colonne de phase inverse C18. Cette approche avantageuse permet de s'affranchir des problèmes de solubilisation mais également des effets de suppression de signal dus au nombre élevé de peptides générés. Cette étude a permis d'identifier un certain nombre de KIFs, de KAPs et de nombreuses autres protéines qui sont impliquées dans la formation et l'organisation des kératines. De plus elle a également permis de mettre en évidence la présence de modifications post-traductionnelles de type méthylation, di-méthylation et tri-méthylation sur les protéines du cheveu [Lee et al., 2006].

I.3. Contexte de l'étude et choix de la stratégie

A l'heure actuelle même si la nature des protéines présentes dans le follicule du cheveu sont connues (séquences en acides aminés, taux d'expression), peu d'informations sont disponibles

concernant leurs **modes d'interaction** et la **manière dont elles évoluent** au cours de la maturation du cheveu ceci impliquant leur présence ou non et quels types de modifications elles subissent.

De ce fait, afin de mieux comprendre cette machinerie moléculaire très complexe, nous avons utilisé l'analyse protéomique pour la caractérisation fine de l'ensemble des protéines de la tige capillaire ce qui inclut l'identification des **protéines majoritaires et minoritaires**, l'identification de **PTMs**, l'étude des **interactions** possibles entre ces protéines et la **localisation** de ces protéines en travaillant sur des **extraits spécifiques du cortex** et sur des **extraits totaux** (cortex et cuticule).

Avant d'aborder la caractérisation des protéines du cheveu par spectrométrie de masse, nous avons optimisé en amont les méthodes d'extraction et de séparation de ces protéines pour les rendre compatibles avec l'analyse protéomique.

De ce fait, nous avons travaillé sur différents extraits de cheveux naturels qui ont été analysés selon deux techniques séparatives complémentaires couplées à la spectrométrie de masse :

- **Gels d'électrophorèse bidimensionnels** suivie d'analyses nanoLC-MS/MS.
- **Chromatographie liquide bidimensionnelle** suivie d'analyses par spectrométrie de masse.

Pour des raisons de confidentialité, nous ne décrivons ici que les résultats qui sont en cours de publication.

II. Méthodologie

II.1. Extraction des protéines

Les extractions des protéines ont été réalisées par les collaborateurs.

II.1.1. Extraction des protéines du cheveu (cortex et cuticule)

L'ensemble des protéines du cheveu ont été extraites suivant le protocole décrit par Lee et al. [Lee et al., 2006].

II.1.2. Extraction des protéines du cortex

Après avoir délipidé les cheveux avec de l'éthanol et du cyclohexane, les protéines ont été extraites dans 200mM Tris-HCl, 200mM DTT, 1,1% de Triton X-100 et 2% CHAPS durant 16 H à 37 °C. L'extrait de protéines a ensuite été alkylé avec une solution à 1M iodoacetamide et 3M Tris-HCl, pH 8,4 durant 10 minutes dans le noir à température ambiante puis dialysé dans des cassettes de dialyse de 3500 MWCO (Pierce, Rockford, USA) dans de l'eau pendant 48 H avant d'être congelé.

II.2. Etude des protéines du cortex par gels 2D suivis d'analyses nanoLC-Chip-MS/MS

II.2.1. Séparation sur gels 2D

Les protéines sont reprises dans un tampon de réhydratation composé de 7M urée, 2M thiourée, 2% CHAPS, 0,5% d'ampholytes pH 3-11 et d'une solution de réhydratation DeStreak (GE Healthcare, Uppsala, Suède) puis incubées avec des strips IPG pH 3-11 (GE Healthcare) dans une cellule de focalisation isoélectrique PROTEAN (BioRad) durant 16 h. L'isoélectrofocalisation est réalisée par paliers de tensions croissantes jusqu'à atteindre 14400 VH. Les strips sont ensuite rééquilibrés dans 6M urée, 2% SDS et 0,5M DTT puis déposés sur des gels NuPAGE à 10% Bis-Tris (Invitrogen, Cergy Pontoise, France). L'électrophorèse est réalisée à 200 V pendant 40 minutes dans un système de mini-cellules Novex (Invitrogen). Tous les gels sont colorés au bleu de Coomassie (Invitrogen).

II.2.2. Digestion des protéines

Les spots d'intérêt sont excisés manuellement, les protéines sont réduites et alkylées puis digérées à la trypsine. Les étapes de réduction, alkylation sont effectuées par le robot de digestion MassPREP, qui est un système automatisé de manipulation de liquide conçu par Packard Instrument - PerkinElmer (MultiPROBE II) et adapté à l'analyse protéomique pour la digestion de protéines « in gel » par Micromass – Waters. Les spots de gel sont lavés deux fois dans 100µL d'un mélange ACN/25mM NH₄HCO₃ (1/1) pendant 10 minutes. Les résidus cystéines sont réduits dans 50 µL d'une solution de 10mM DTT dans 25mM NH₄HCO₃ à 57 °C durant 1h puis alkylés par ajout de 50µL d'une solution à 55mM d'iodoacétamide dans 25mM NH₄HCO₃ pendant 20 minutes à température ambiante. Après avoir déshydraté les spots de gel en ajoutant de l'acétonitrile, les protéines sont digérées dans le gel par ajout de 15 µL d'une solution de trypsine dans 25mM NH₄HCO₃ (c= 0,0125 µg/µL). La réaction de digestion se fait à température ambiante pendant une nuit. Les peptides issus de la digestion sont extraits du morceau de gel à l'aide d'une solution H₂O/ACN (35/60) + 5% HCOOH pendant 1 heure à température ambiante puis l'excès d'ACN est éliminé par évaporation au speedvac.

II.2.3. Identification des protéines par nanoLC-Chip-MS/MS

Les mélanges de peptides ont été analysés par nanoLC-Chip-MS/MS sur le système nanoHPLC-Chip série 1100 de Agilent (Agilent Technologies, Palo Alto, USA) couplé à une trappe à ions HCTultra (Bruker Daltonics, Brême, Allemagne). Les peptides sont séparés sur une colonne de phase inverse C18 (Zorbax 300SB-C18, 75µm x 43mm, 5µm de diamètre interne) à un débit de 300 nL/min par un gradient d'acétonitrile allant de 8 à 40% du solvant B en 7 minutes avec A : eau + 2% acétonitrile + 0,1% d'acide formique et B : acétonitrile + 2% eau + 0,1% d'acide formique. L'acquisition des spectres de masse se fait en mode positif et la tension appliquée à l'entrée du capillaire a été optimisée à -1850 V. Pour l'acquisition des spectres en mode MS et MS/MS le système travaille en automatique. Pour chaque spectre MS, les trois ions les plus abondants et préférentiellement doublement chargés sont sélectionnés pour être isolés et fragmentés. Afin de gagner en sensibilité, le balayage des spectres MS/MS se fait en mode « ultrascan resolution » à une vitesse de 26000 *m/z* par seconde. Pour chaque spectre MS/MS, on moyenne 6 balayages ce qui permet d'améliorer la qualité des spectres. L'ensemble

du système est piloté par le logiciel ChemStation (Agilent Technologies) et EsquireControl (Bruker Daltonics). Les performances du système nanoLC-Chip-MS/MS sont vérifiées en analysant 200 fmol d'un mélange standard de peptides issus de la digestion à la trypsine d'albumine de sérum bovin.

Les données de masse collectées durant ces analyses sont traitées, converties en fichiers *.mgf, puis interprétées par stratégie PFF en comparant les listes de masses calculées aux listes de masses théoriques présentes dans la banque SwissProt via le moteur de recherche MASCOT™ (Matrix Science, Londres, Grande-Bretagne). Nous avons utilisé les critères de recherche suivants : (1) 1 site de coupure maximum manqué par peptide, (2) les peptides peuvent avoir une proline comme acide aminé N-terminal, (3) nous autorisons un écart de 0,3 Da sur les masses des ions parents et des ions fragments, (4) nous pouvons avoir les modifications suivantes : carbamidométhylation d'une cystéine (+ 57 Da), oxydation d'une méthionine (+ 16 Da), méthylation (+ 14 Da) des histidines et des arginines, et les lysines peuvent être mono-, di- (+ 28) et tri-méthylées (+ 42 Da). Pour les protéines identifiées avec un peptide et les peptides identifiés avec une PTM, les spectres MS/MS sont inspectés manuellement et validés dans le cas où nous avons des séries successives d'ions fragments de type b et y interprétables qui permettent d'attribuer avec certitude la séquence à une protéine. De plus, les KIFs ayant de fortes homologues de séquences, les séquences en acides aminés des peptides identifiés avec une PTM sont vérifiées qu'elles sont uniques en comparant les séquences des KIFs en utilisant un logiciel d'alignement de séquence (MultAlin, INRA, Toulouse, France).

II.3. Etude des deux extraits de protéines du cheveu par chromatographie liquide bidimensionnelle (SCX/RP/MS/MS des peptides)

II.3.1. Digestion des deux extraits de protéines (extrait total et extrait du cortex)

Des aliquots des deux extraits sont incubés dans 2% SDS, 20mM DTT et 50mM Na₂HPO₄ pendant 1 h à 67 °C afin de réduire toutes les cystéines. Celles-ci sont ensuite alkylées dans 40mM iodoacetamide pendant 1 h à température ambiante. Les protéines sont ensuite précipitées à l'éthanol et rincées 2 fois dans 70% d'éthanol afin d'éliminer les excès de SDS, de DTT, de Na₂HPO₄ et d'iodoacetamide. Elles sont alors reprises dans 100mM NH₄HCO₃ et 2M urée puis digérées pendant une nuit à 37 °C par ajout de trypsine dans un rapport 1 pour 20 en masse de protéines.

II.3.2. Première dimension : chromatographie par échange de cations (SCX)

Les deux extraits de peptides sont fractionnés sur un système LC Waters 625 (Waters, Milford, MA, USA). Les peptides sont séparés sur une colonne PolySULFOETHYLA™ (PolyLC INC, Columbia, MD, USA) (100mm x 2.1mm, 5µm, 300 Å) à un débit de 200 µL/min par un gradient en trois étapes : 0% de B pendant 15 minutes, 0% à 25% de solvant B en 30 minutes puis de 25% à 100% de B en 20 minutes avec A : 5mM KH₂PO₄, 25 % ACN, pH 3 et B : 5mM KH₂PO₄, 25 % ACN, 350mM KCl, pH 3. Les fractions de peptides sont collectées toutes les deux minutes de la 1^{ère} à la 70^{ème} minute, concentrées par évaporation sous vide puis dessalées en utilisant des cônes de pipettes ZipTip_{C18} (Millipore, Bedford, MA, USA) avant d'être analysées par nanoLC-MS/MS.

II.3.3. Seconde dimension : analyses nanoLC-Chip-MS/MS sur phase inverse C18

Les fractions résultantes des deux extraits sont analysées par nanoLC-Chip-MS/MS. Le système LC, le spectromètre de masse ainsi que les conditions utilisées sont les mêmes que celles décrites ci-dessus pour l'analyse des extraits de peptides issus des spots de gels 2D. Afin de vérifier les résultats, ces mêmes fractions ont également été analysées par nanoLC-MS/MS sur un système nanoHPLC Ultimate 3000 (Dionex) couplé à un spectromètre de masse LTQ-FT-ICR (Thermo) en utilisant une colonne de phase inverse C18 (75µm x 150mm). Les peptides sont élués à un débit de 300 nL/min en utilisant un gradient d'acétonitrile allant de 4% à 50% de solvant B en 30 minutes avec A : eau + 2% acétonitrile + 0,1% d'acide formique et B : 80% d'acétonitrile + eau + 0,1% d'acide formique.

II.3.4. Interprétation des données de masse

Les données de masse collectées durant ces analyses sont traitées, converties en fichiers *.mgf. Pour chaque extrait on a en tout 35 fichiers *.mgf. Tous les fichiers correspondant à un des deux extraits sont additionnés en utilisant un logiciel créé au laboratoire (concatene.exe) afin d'obtenir un fichier *.mgf unique. Les données de masse présentes dans celui-ci sont alors interprétées par stratégie PFF en comparant les listes de masses calculées aux listes de masses théoriques présentes dans la banque de l'humain de SwissProt via le moteur de recherche MASCOT™ (Matrix Science, Londres, Grande-Bretagne). Nous avons utilisé les critères de recherche suivants : (1) 1 site de coupure maximum manqué par peptide, (2) les peptides peuvent avoir une proline comme acide aminé N-terminal, (3) nous autorisons un écart de 0,3 Da sur les masses des ions parents et des ions fragments, (4) nous pouvons avoir les modifications suivantes : carbamidométhylation et tri-oxydation d'une cystéine, oxydation d'une méthionine, méthylation des histidines et des arginines, et les lysines peuvent être mono-, di- et tri-méthylées. Pour les peptides identifiés avec une PTM, les spectres MS/MS sont inspectés manuellement et validés dans le cas où nous avons des séries successives d'ions fragments de type b et y interprétables. De plus, les KIFs ayant de fortes homologues de séquences, chaque séquence en acides aminés des peptides identifiés avec une PTM sont vérifiées qu'elles sont uniques en comparant les séquences des KIFs en utilisant un logiciel d'alignement de séquence (MultAlin, INRA, Toulouse, France). Pour les protéines identifiées avec un peptide, les spectres MS/MS sont inspectés manuellement et validés dans le cas où nous avons des séries successives d'ions fragments de type b et y interprétables qui permettent d'attribuer avec certitude la séquence à une protéine.

III. Résultats

III.1. Protéines identifiées dans les extraits du cortex après séparation sur gel 2D des protéines suivi d'analyses nanoLC-MS/MS

La figure 7 montre le profil du gel 2D de l'extrait des protéines du cortex obtenu après avoir optimisé les paramètres de séparation. Celui-ci montre un profil similaire à celui décrit par Langbein et al. lors des études menées pour détecter les protéines par marquage avec des anticorps [Langbein et al., 1999], [Langbein et al., 2001] : un chapelet de spots dans la gamme des 62 kDa étendue dans la zone des pI 4 à 7 qui correspond aux KIFs de Type II et un groupe de spots dans les pI acides qui correspond aux KIFs de Type I.

Cependant, à la différence des études menées par Langbein et al (1999 et 2001), le gel 2D révèle dans la **zone inférieure à 42 kDa** un nombre important de spots (85 en tout) qui d'après leur zone de migration pourraient correspondre aux protéines de la matrice, les KAPs. Ces **85 spots** ont été découpés et analysés par nanoLC-Chip-MS/MS.

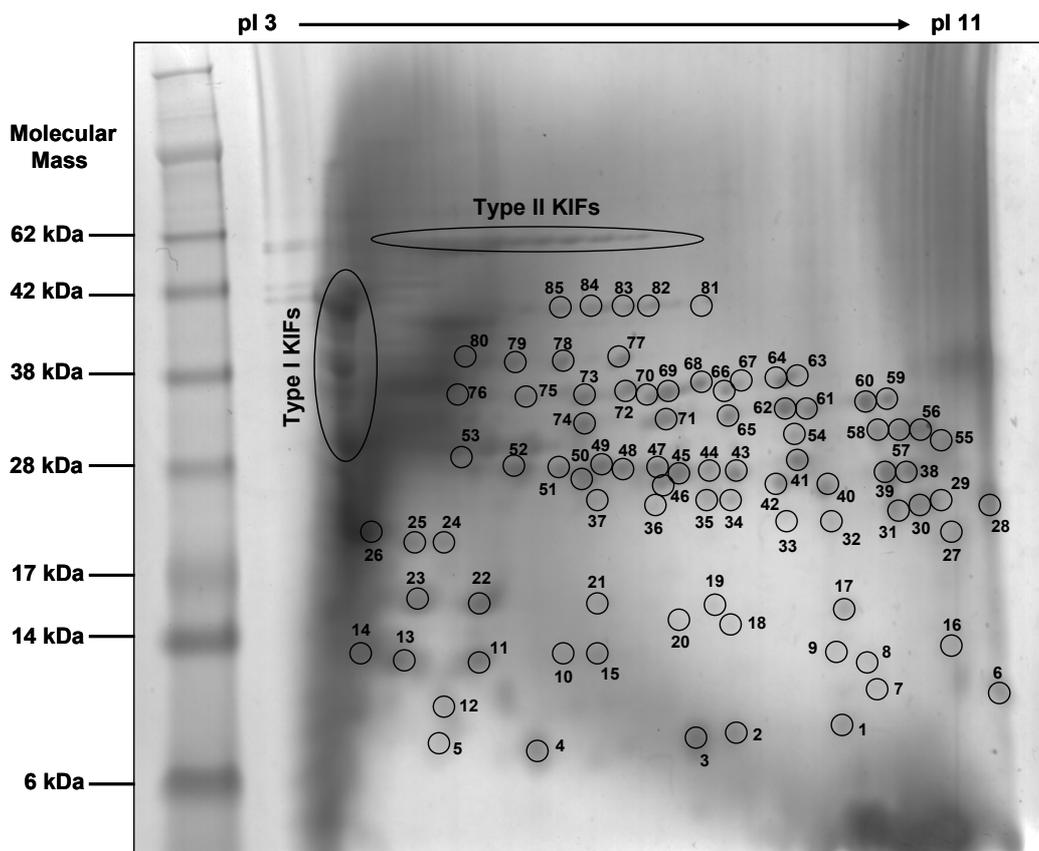


Figure 7 : Profil du gel d'électrophorèse bidimensionnelle coloré au bleu de Coomassie de l'extrait des protéines du cortex. Les spots d'intérêt qui ont été excisés pour des analyses nanoLC-MS/MS sont entourés et annotés.

Après interprétation des données de masse par recherche dans la banque protéique SwissProt via le moteur de recherche MASCOT™, nous avons pu identifier **26 protéines** en tout qui peuvent se répartir en trois groupes distincts, les **KIFs**, les **KAPs** et les protéines **S100** qui fixent le calcium. L'ensemble de ces protéines sont identifiés **dans plusieurs spots** répartis **sur l'ensemble du gel 2D** et souvent **plusieurs protéines** sont identifiées **dans un même spot** (tableau 2).

Comme attendu, les protéines majoritairement identifiées sont les KIFs : **4 de Type II** (K81, K83, K85 et K86) et **6 de Type I** (K31, K32, K33a, K33b, K34 et K35). Parmi celles-ci les plus fréquemment identifiées sont K81, K85, K86 et K31.

Concernant les KAPs, nous avons pu identifier **7 KAPs** (1.5, 2.4, 3.1, 3.2, 3.3, 11.1 et 13.1) avec un ou deux peptides spécifiques et **2 familles de KAPs** (4 et 9) avec dans la plupart des cas un peptide commun à une famille. Le fait que nous n'ayons pas identifié plus de KAPs n'est pas étonnant.

En effet, les KAPs sont des protéines basiques de PM peu élevées qui sont peu abondantes par rapport aux KIFs (on estime qu'elles ne représentent que 20% de la totalité des protéines). Ceci rend leur migration ainsi que leur détection sur un gel 2D difficile. De plus elles sont riches en cystéines et présentent peu de sites de coupure à la trypsine. Enfin, les 23 familles de KAPs (environ 80 KAPs en tout) présentent de fortes homologies de séquences.

Néanmoins un résultat intéressant est que nous identifions des **KIFs** et des **KAPs** sur l'ensemble du gel 2D **dans des mêmes spots** qui souvent se situent dans des **zones inférieures** (pour les KIFs) ou **supérieures** (pour les KAPs) à leur **PM** et dans des **gammes de pH ne correspondant pas à leurs pI**. Le fait d'observer des KIFs sur l'ensemble du gel 2D n'est pas dû à des problèmes de migration mais semble indiquer que nous sommes en présence de **formes dégradées des KIFs** portant des **PTMs**, le tout pouvant engendrer ces décalages de pI et de PM. En effet, Lee et al. ont mis en évidence la présence de PTMs de type méthylation sur les KIFs. De ce fait nous avons recherché des PTMs.

Les résultats des recherches MASCOT nous ont permis d'identifier des PTMs sur les KIFs de **Type II** (K81, K83, K85 et K86) avec **6 PTMs** identifiées dans des spots présents dans la région 28-42 kDa et **3 PTMs** dans des spots des régions 14-28 kDa (tableau 3).

Proteins	Sequences of modified peptides	Identified in spots numbers
K81/ K86	K.LAELEGALQK ^{dm} AK.Q	23
K81/ K86	R.GLTGGFGSH ^m SVC ^{am} GGFR.A	29/30/31/32/41
K81/ K83/ K86	R.K tm SDLEANVEALIQEIDFLR.R	76/80
K81/ K85/ K86	R.FLEQQNK ^{dm} LLETK.W	56/58/85
K85	R.NFSSC ^{am} SAVAPK ^{dm} TGNR.C	55/56/57/58/59/60/61/62/64/69/70/71/73/76/78/80
K85	R.NFSSC ^{am} SAVAPK tm TGNR.C	56/57/62/71/76
K85	K.LLETK tm WQFYQNQR.C	58/73
K85	K.K tm YEEVALR.A	62
K85	R.LTAEIENAK ^{dm} C ^{am} QR.A	83/84/85

Tableau 3 : Liste des peptides identifiés portant des modifications post-traductionnelles avec les protéines dans lesquelles on retrouve ces séquences en acides aminés et les numéros des spots dans lesquels on identifie ces peptides. Pour chaque peptide, l'acide aminé modifié est souligné et le type de modification est indiqué en exposant par une lettre avec m (méthylation), dm (di-méthylation), tm (tri-méthylation), et am (carbamidométhylation).

Cinq de ces PTMs ont pu être attribuées à K85, les 4 autres sont présentes sur des peptides communs à plusieurs KIFs de Type II. Concernant les KIFs de Type I, aucune PTM n'a été identifiée. Ceci suggère qu'elles ne subissent pas de PTMs et expliquerait le fait que les KIFs de Type I ne sont pas observées sur l'ensemble du gel 2D comme c'est le cas pour les Type II. Toutefois, l'identification de PTMs à partir de gels 2D n'étant pas chose facile (peptides minoritaires masquées par les peptides majoritaires), il n'est pas exclu que les KIFs de Type I soient modifiées.

Néanmoins, les quelques peptides identifiés ici avec des PTMs de type méthylation n'expliquent pas ces décalages de pI et cette hétérogénéité sur le gel 2D. Par exemple K85 est identifiée dans la zone de PM 28 kDa à un pI de 10 alors que théoriquement elle a un PM de 56 kDa et un pI de 6,3.

La présence de ces protéines dans des **zones inférieures à leurs PM théoriques** pourrait s'expliquer par la présence de **formes dégradées des KIFs** générant ainsi des protéines de tailles plus petites correspondant aux différents domaines N-terminaux, C-terminaux ou à la tige centrale. Toutefois les données de spectrométrie de masse ne nous ont pas permis de déterminer avec certitudes quelles formes dégradées nous avons dans chaque spot. En effet, dans la plupart des cas, nous identifions plusieurs KIFs dans un même spot avec souvent des peptides communs à plusieurs kératines et des recouvrements sur l'ensemble de la protéine.

Par exemple, pour le spot 59 (PM 33-38 kDa, pI 10), nous avons identifié trois kératines de Type II K81 (PM 55 kDa, pI 5,5), K85 (PM 56 kDa, pI 6,3) et K86 (PM 53 kDa, pI 5,5). K86 est identifiée avec 8 peptides dont seulement deux sont spécifiques à K86 et appartiennent à la partie C-terminale. K85 est identifiée avec 13 peptides dont 7 sont spécifiques, 5 correspondent à la partie N-terminale et 2 correspondent à la partie C-terminale. K81 est identifiée avec 7 peptides dont deux sont spécifiques et appartiennent à la partie C-terminale. De ce fait, dans ce spot, il n'est pas exclu que nous ayons des formes dégradées de K81 et K86 contenant le domaine C-terminal et plusieurs formes de K85 contenant chacune un des trois domaines.

Concernant les **décalages de pI**, ceux-ci peuvent être expliqués par la présence dans ces spots de **protéines plus basiques** qui pourraient former des **liaisons intermoléculaires avec les KIFs**. En effet, nous identifions dans des mêmes spots des KIFs et des KAPs. Par exemple, la KAP 2.4 de PM 13 kDa et de pI 8,3 est identifiée dans plusieurs spots de la région des 30 kDa, notamment dans le spot 59. Elle pourrait se lier aux formes dégradées de K85, K81 et K86 qui voient déjà leur pI modifié par la dégradation et entraîner un décalage dans les gammes de pI basiques. Actuellement on sait que les KIFs et les KAPs forment des liaisons intermoléculaires de type ponts disulfures, néanmoins, dans notre cas les ponts ont été réduits et alkylés avant de séparer les protéines. De ce fait, une question se pose, quels types d'interactions existe-t-il entre ces protéines ?

Une hypothèse réside dans des liaisons engendrées par des enzymes de la famille des transglutaminases (TGase). Les transglutaminases sont des enzymes dépendantes du calcium qui engendrent la formation de **liaisons intermoléculaires de type N^ε-(γ-glutamine) lysine** entre protéines [Folk and Finlayson, 1977]. En effet, des études ont montré la présence de la TGase 1 et de la TGase 5 dans la gaine interne du follicule pileux et de la **transglutaminase E (TGase 3)** dans la tige pileuse du cheveu [Yoneda et al., 1998], [Thibaut et al., 2005]. De plus cette dernière a été identifiée par spectrométrie de masse dans des extraits totaux de cheveu [Lee et al., 2006]. En étant

spécifiquement exprimée dans la tige pileuse, cette enzyme semble jouer un rôle important dans les étapes de kératinisation de la tige et donc dans la formation des filaments intermédiaires.

Afin de mettre en évidence ce type de liaison, **deux expériences complémentaires** ont été réalisées en collaboration avec l'équipe de BA Bernard, L'Oréal. Dans un premier temps, plusieurs spots du gel 2D dans la zone des 20-38 kDa et des pI 10-11 ont été excisés et les protéines digérées avec une enzyme aspécifique, la pronase. Des études ont montré que cette enzyme, connue pour dégrader une protéine jusqu'à ne plus avoir qu'un mélange d'acides aminés, laisse la liaison intermoléculaire glutamine-lysine intacte [Kim and Yim, 1999]. Le **mélange d'acides aminés** a ensuite été analysé et nous a permis de montrer que nous avons un épaulement dans le chromatogramme des acides aminés qui pourrait correspondre à cette liaison isopeptidique (figure 8). En parallèle, des **expériences *in vitro*** ont été réalisées entre une kératine et un peptide synthétique de la KAP 3.1 en présence de TGase 3 montrant la possibilité d'avoir ce type de liaisons entre KIFs et KAPs dans le cheveu.

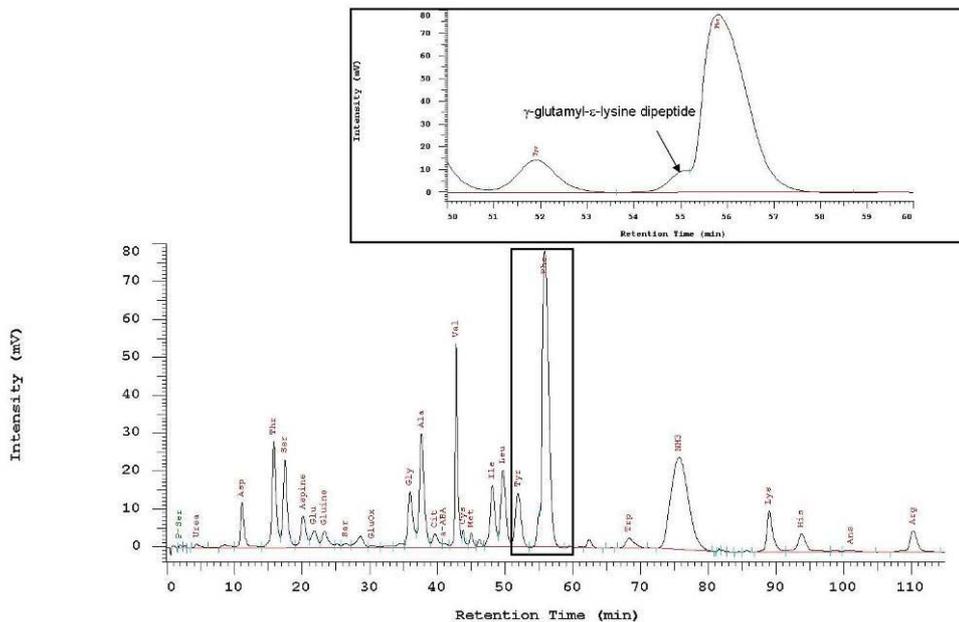


Figure 8 : Profil chromatographique de l'analyse des acides aminés résultant de la digestion à la pronase de plusieurs spots du gel 2D correspondant aux gammes de PM 20-38 kDa et de pI 10-11 et zoom dans la zone des temps de rétention 50-60 minutes où on observe un épaulement qui pourrait correspondre à la liaison isopeptidique.

L'ensemble de cette étude a permis de montrer la possibilité de la présence d'un second réseau covalent dans le cheveu basé sur la formation de liaisons glutamine-lysine engendrées par la TGase 3 (le premier étant basé sur les ponts disulfures). Ainsi il a pu être suggéré que ce type de liaisons entre les IFs et les KAPs engendrées par la TGase 3 participe directement dans la formation progressive de l'échafaudage durant la pousse du cheveu. Ces résultats ont fait l'objet d'une publication en 2008 dans « Journal of Investigative Dermatology » (annexe 4).

Enfin cette théorie peut également être renforcée par le fait que nous identifions dans plusieurs spots de gel 2D les protéines **S100-A8** et **S100-A9** dont la présence avait déjà été montrée dans le follicule pileux du cheveu [Schmidt et al., 2001]. Ces petites protéines (10-13 kDa) appartiennent à la famille des protéines S100 qui fixent le calcium et qui sont impliquées dans de nombreux processus biologiques tels que le développement et la différenciation cellulaire, la régulation du cycle d'une cellule, la transcription et l'activité des récepteurs de transduction du signal. La transglutaminase étant une enzyme qui a besoin du calcium pour être activée, ces deux protéines pourraient lui fournir celui-ci [Eckert et al., 2005].

En effet, des études ont montré que cette famille de protéines était un substrat naturel de la transglutaminase et plus particulièrement pour S100-A7, S100-A10 et S100-A11 [Ruse et al., 2001], [Robinson and Eckert, 1998]. De plus, une de ces protéines, la S100-A3, est exprimée dans la cuticule du cheveu humain [Kizawa et al., 1996] et semble être impliquée dans la structure stable du cheveu ainsi que dans les phénomènes engendrant la dégradation du cheveu. En effet, l'expression de S100-A3 dans la cuticule diminue après que le cheveu ait été lavé ou traité chimiquement [Inoue et al., 2000].

Enfin, ces deux protéines sont très riches en lysines suggérant qu'elles pourraient former des liaisons intermoléculaires avec les glutamines des kératines et des KAPs. Elles pourraient donc d'une part jouer un rôle de transporteur de calcium pour activer la TGase 3 et d'autre part jouer un rôle de stabilisateur des filaments intermédiaires en formant des liaisons intermoléculaires avec les KIFs et les KAPs.

En résumé, la **séparation des protéines du cortex sur gels 2D** suivi de leur **analyse par nanoLC-Chip-MS/MS** a permis d'apporter de nouvelles informations concernant la structure du cheveu :

- Les résultats obtenus suggèrent que les **KIFs existent sous différentes formes dégradées** ne contenant que les domaines tête, queue, tige centrale individuellement ou tête/tige centrale et queue/tige centrale entraînant de fortes modifications de leur PM et de leur pI.
- Les **KIFs de Type II semblent porter plus de PTMs que les Types I**. En effet, les Types II sont observées sur l'ensemble du gel 2D alors que les Type I sont plutôt observées dans les zones de pI 4-6 et 8-11. les PTMs n'ont été identifiées que sur des peptides des KIFs de Type II. Les KIFs de Type II pourraient donc être plus exposées vers la surface des IFs.
- Les décalages de PM et de pI laissent penser que les **KIFs pourraient former des liaisons intermoléculaires de type glutamine-lysine avec les KAPs** puisque celles-ci sont **identifiées dans des mêmes spots** dans des gammes de PM et de pI qui ne correspondent pas à leurs gammes théoriques. Par exemple, les KIFs de Type I acides détectées dans des pI basiques pourraient former des liaisons avec des KAPs plus basiques comme la KAP 2.4. De même les KIFs de Type II détectées dans des pI très acides pourraient former des liaisons avec des KAPs plus acides comme la KAP 3.1 par exemple. De plus la présence des familles de KAP 2 et 3 dans la plupart des spots (25 et 22 spots respectivement) renforce cette hypothèse.

Enfin, nous avons identifiées deux protéines S100-A8 et S100-A9 de la famille des **protéines S100** qui fixent le calcium. Celles-ci pourraient **jouer le rôle de fournisseur de calcium à la**

transglutaminase qui en a besoin pour être activée et entraîner la formation de liaisons intermoléculaires de type glutamine-lysine. De plus ces deux protéines étant **riches en lysines**, elles pourraient également **former des liaisons intermoléculaires avec les glutamines des KIFs et des KAPs**.

III.2. Identification des protéines par analyse LC/nanoLC-Chip-MS/MS de mélanges complexes de peptides.

La séparation des protéines sur gels d'électrophorèse 2D est limitante pour la détection des protéines minoritaires et des protéines aux PM et pI extrêmes. De ce fait, nous avons complété cette première étude en utilisant une approche de chromatographie liquide bidimensionnelle en offline, l'analyse LC/nanoLC-Chip-MS/MS de mélanges complexes de peptides issus de la digestion des extraits de protéines.

Le but de cette étude était d'une part de pouvoir identifier un nombre supérieur de protéines et d'autre part d'identifier plus de modifications post-traductionnelles. L'objectif final étant de pouvoir expliquer la forte hétérogénéité des protéines du cheveu observée sur le gel 2D.

De plus nous avons réalisé cette étude sur deux types d'extraits, un extrait du cortex et un extrait total (cortex et cuticule) d'un cheveu naturel. L'objectif étant de donner une image globale de la localisation des protéines dans les différentes zones d'un cheveu mature.

III.2.1. Extrait total (cortex et cuticule)

a) Les protéines identifiées

Après avoir purifié les mélanges complexes de peptides issus de la digestion à la trypsine des extraits protéiques par chromatographie liquide sur une colonne échangeuse de cations, les 35 fractions résultantes ont été analysées par nanoLC-Chip-MS/MS sur une chip phase inverse C18. Nous avons pu identifier 53 protéines au total (tableau 4).

Les protéines majoritairement identifiées sont les KIFs de Type I et II. Nous avons pu en identifier 14 des 15 qui sont exprimés dans le cheveu, **5 KIFs de Type II** (K81, K82, K83, K85 et K86) et **9 KIFs de Type I** (K31, K32, K33a, K33b, K34, K35, K36, K37 et K38). Parmi celles-ci, les recouvrements de séquences les plus élevés ont été obtenus pour K81, K85, K86, K31, K33a et K33b suggérant qu'il pourrait s'agir des KIFs les plus représentés dans le cheveu mature.

Concernant les **KAPs**, **21 ont été identifiées en tout dont 11 avec des peptides spécifiques** (KAP 1.1, KAP 2.4, KAP 24.1, KAP 11.1, KAP 3.1, KAP 3.2, KAP 3.3, KAP 13.2, KAP 19.5, KAP 4.3, KAP 4.9) et 10 avec des peptides communs à une famille de KAPs (KAP 4.7/4.12/4.15 avec 4 peptides, KAP 4.5/4.8 avec 2 peptides, KAP 9.2/9.9 ou 9.5 avec 2 peptides, KAP 10.1/10.3/10.7 avec 3 peptides).

De plus, nous avons pu identifier **24 autres protéines mineures** avec 1 à 4 peptides identifiés par protéine. Celles-ci peuvent être classées en différents groupes selon leur rôle biologique : (1) les protéines desmosomales (desmoglein-4, plakoglobin, plakophilin, desmoplakin) qui jouent un rôle crucial dans l'adhésion des kératinocytes aux cellules voisines, (2) des protéines de structures (histones H2A, H2B, H4), (3) des protéines qui fixent le calcium (S100-A3, calmodulin-like protein 3), (4) des protéines de stabilisation (heat shock cognate 71kDa), (5) des protéines impliquées dans la signalisation (14-3-3 protein, leucine rich repeat containing protein 15), (6) des enzymes protéolytiques (lysozyme g-like, bleomicin, hydrolases), (7) un peptide antimicrobien (dermicin) connu pour protéger la surface épithéliale, (8) des protéines qui fixent le sélénium qui sont impliquées dans le transport des protéines intra-Golgi, (9) la sialdase-2 qui hydrolyse les composés sialilés, (10) des glycoprotéines, (11) des protéines présentes dans les cellules épithéliales (serpin B5, Syntaxin-2), (12) la nesprine-1 qui est impliquées dans la maintenances de l'organisation nucléaire et la structure (Lee et al., 2007). Enfin, nous avons pu identifier la TGase 3 qui est impliquée dans la formation de liaisons intermoléculaires glutamine-lysine comme suggéré dans la partie concernant les gels 2D.

AC	Protéins	Sequence coverage (%)	Number of unique peptides	AC	Protéins	Sequence coverage (%)	Number of unique peptides		
Q43790	Keratin Type II K86	71	45	P27482	Calmodulin-like protein 3	19	2		
P78386	Keratin Type II K85	75	47	Q86SJ6	Desmoglein-4 precursor	4	3		
Q14533	Keratin Type II K81	63	44	P33778	Histone H2B type 1-B	19	2		
P78385	Keratin Type II K83	54	40	P88576	Histone H2B type 1-D				
Q9NSB4	Keratin Type II K82	31	18	Q93079	Histone H2B type 1-H				
Q15323	Keratin Type I K31	62	31	P06899	Histone H2B type 1-J				
O76009	Keratin Type I K33a	72	33	O60614	Histone H2B type 1-K				
Q14525	Keratin Type I K33b	72	30	Q99880	Histone H2B type 1-L				
O76011	Keratin Type I K34	59	25	Q99879	Histone H2B type 1-M				
Q92764	Keratin Type I K35	36	17	Q99877	Histone H2B type 1-N				
O76013	Keratin Type I K36	19	6	P23227	Histone H2B type 1-O				
Q14532	Keratin Type I K32	31	11	Q16778	Histone H2B type 2-E				
O76014	Keratin Type I K37	19	7	Q50NW6	Histone H2B type 2-F				
O76015	Keratin Type I K38	19	7	Q8N267	Histone H2B type 3-B				
Q9BYR9	KAP 2.4	42	6	P57053	Histone H2B type F-S				
Q8IUC1	KAP 11.1	24	3	P62807	Histone H2B type 1-C/E/F/G/I				
Q9BYR8	KAP 3.1	24	2	P31947	14-3-3 protein sigma	9	2		
Q9BYR7	KAP 3.2	17	1	P14923	Junction plakoglobin	5	3		
Q9BYQ4	KAP 9.2	19	2	P04908	H2A type 1-B	26	3		
Q9BYP9	KAP 9.5 or 9.9	21	2	Q93077	Histone H2A type 1-C				
Q9BYR6	KAP 3.3	17	1	P20671	Histone H2A type 1-D				
Q3LI83	KAP 24.1	5	1	P28001	Histone H2A type 1-E				
Q3LI72	KAP 19.5	22	1	Q96KK5	Histone H2A type 1-H				
Q9BYR0	KAP 4.7	21	4	Q99878	Histone H2A type 1-J				
Q9BYR4	KAP 4.3	17	1	P0C0S8	Histone H2A type 1				
Q9BYQ6	KAP 4.9	22	2	Q6F113	Histone H2A type 2-A				
Q07627	KAP 1.1	6	1	Q16777	Histone H2A type 2-C				
Q62LG2	KAP 13.2	7	1	Q7L7L0	Histone H2A type 3				
Q9BYQ5	KAP 4.6 or 4.15	23	4	P01606	Dermicin precursor			10	1
Q9BYR2	KAP 4.5	10	2	P33764	Protein S100-A3			12	1
Q9BYQ9	KAP 4.8	16	2	P62806	Histone H4	27	4		
P60331	KAP 10.1	8	3	Q13228	Selenium-binding protein 1	2	1		
P60369	KAP 10.3	10	3	P11142	Heat shock cognate 71 kDa protein	2	1		
P60409	KAP 10.7	6	3	Q9Y3R4	Sialidase-2	3	1		
Q9BQ66	KAP 4.12	22	4	Q13867	Bleomycin hydrolase	3	1		
				P36952	Serpin B5 precursor	3	1		
				P63244	Guanine nucleotide-binding protein subunit beta 2-like 1	4	1		
				Q86SG7	Lysozyme g-like protein 2 precursor	6	1		
				Q13836	Plakophilin-1	2	1		
				Q8TDC3	BR serine/threonine-protein kinase 1	1	1		
				Q8TF66	Leucine-rich repeat-containing protein 15 precursor	2	1		
				Q08188	Protein-glutamine gamma-glutamyltransferase E precursor	2	1		
				P11279	Lysosome-associated membrane glycoprotein 1 precursor	2	1		
				P13473	Lysosome-associated membrane glycoprotein 2 precursor	2	1		
				Q8NFP1	Nesprin-1	2	1		

Tableau 4 : Liste des protéines identifiées par analyse LC/nanoLC-Chip-MS/MS du mélange complexe de peptides issus de la digestion de l'extrait total de cheveu avec numéro d'accension, nom de la protéine, pourcentage de recouvrement et nombre de peptides uniques identifiés.

L'ensemble de ces protéines minoritaires avaient déjà été identifiées par Lee et al. par LC 2D, et nos résultats ne font que confirmer la présence de ces protéines dans le cheveu.

b) Les modifications post-traductionnelles des kératines

Les recherches dans la banque « humain » disponible dans SwissProt se sont faites via le moteur de recherche MASCOT™ en autorisant les différentes modifications suivantes: mono-, di- et tri-méthylation des lysines, méthylation des histidines et des arginines et tri-oxydation des cystéines en acides cystéiques. Les résultats ont clairement montré que tous les peptides identifiés avec des PTMs appartiennent à des KIFs.

Même si la différence de masse générée par certaines modifications comme la tri-oxydation des cystéines (+ 47,985 Da) est spécifique et ne conduit à aucune ambiguïté, ceci n'était pas le cas pour les modifications de type méthylation.

En effet, les masses des arginines et des lysines modifiées peuvent être très proches. Par exemple, la masse d'une arginine méthylée ne varie que de 0,023 Da de la masse d'une lysine acétylée ou tri-méthylée. De même, la masse d'une lysine di-méthylée ne varie que de 0,006 Da de la masse d'une arginine.

De plus, les KIFs de Type I et de Type II présentent de fortes homologues de séquences avec souvent des substitutions d'acides aminés tels que lysine/arginine, leucine/isoleucine, glutamine/lysine qui peuvent poser problèmes lors de l'interprétation des données de masse. Nous citerons ici quelques exemples d'homologies de séquences dans la famille des KIFs de Type II :

- Substitution lysine /arginine et leucine/isoleucine : AKQDMACLLKdm et AKQDMACLIR (delta de masse de 0,006 Da).
- Substitution glutamine/lysine : FLEQQNK et FLEQKNK (delta de masse de 0,036 Da).

De même, certains spectres MS/MS pour lesquels nous n'avons pas les séries complètes des ions fragments b ou y peuvent également poser problème. C'est le cas par exemple de l'ion doublement chargé de rapport m/z 674,841. Il pouvait être attribué au peptide ATAENEFVALKdmK avec une di-méthylation de la lysine ou au peptide ATAENEFVVLKK non modifié.

En effet, la masse de l'ion fragment correspondant à ALKdmK ne diffère que de 0,023 Da de la masse de l'ion fragment correspondant à VLKK. De ce fait, sans la présence des ions fragments ATAENEFVA ou ATAENEFVV, il est impossible de trancher.

De ce fait, tous les spectres MS/MS ont été interprétés manuellement et comparés aux séquences pour chaque groupe des KIFs en utilisant un logiciel d'alignement de séquence.

De plus afin de confirmer toutes les identifications faites à partir d'une trappe à ions, tous les échantillons ont été analysés chez les collaborateurs sur un LTQ-FTICR pour faire de la mesure de masse précise. Ainsi, nous avons pu lever toutes les ambiguïtés concernant les peptides identifiés avec une PTM. Ces résultats ne seront pas présentés dans ce travail.

Nous avons pu identifier trois types de modifications : la **tri-oxydation des cystéines** (12 peptides), la **di-méthylation des lysines** (6 peptides) et la **méthylation des histidines** (2 peptides). La majorité des modifications ont été observées **sur les KIFs de Type II** avec un nombre plus élevé de

cystéines tri-oxydées que **sur les KIFs de Type I** (tableau 5). Certains peptides identifiés avec des PTMs sont spécifiques d'une protéine comme c'est le cas pour K85 et K86, mais de manière générale, les peptides identifiés avec des modifications sont communs à plusieurs protéines.

Total hair extract	
Proteins containing these sequences	Modified peptides identified
K81/ K83/ K86	K.AKQDMA <u>C</u> ^{tox} LIR.E
K81/ K83/ K85/ K86	K.K ^{dm} DVDC ^{am} AYLR.K
K81/ K83/ K85/ K86	K.LAELEGALQK ^{dm} AK.Q
K81/ K86	R.AFSC ^{am} ISAC ^{tox} GPRPGR.C
K81/ K83/ K85/ K86	R. <u>C</u> ^{tox} KLAELEGALQK.A
K81/ K85/ K86	R.DLNM ^{ox} DC ^{tox} IIEIK.A
K81/ K83/ K85/ K86	R.EAEC ^{ox} VEADSGR.L
K81/ K83/ K85/ K86	R.FAAFIDK ^{dm} VR.F
K81/ K83/ K85/ K86	R.FLEQQNK ^{dm} LLETK.L
K86	R.GGVVC ^{am} GDLC ^{tox} ASTTAPVSTR.V
K81/ K83/ K86	R.GLTGGFGSH ^m SV ^C ^{am} GGFR.A
K81/ K83/ K86	R.GLTGGFGSHSV ^C ^{tox} GGFR.A
K81/ K83/ K85/ K86	R.TK ^{dm} EEINELNR.M
K86	R.VLQSH ^m ISDTSVVVK.L
K85	R.GGVSC ^{ox} GGLSYSTTPGR.Q
K85	R.SLC ^{am} NLGSC ^{tox} GPR.I
K31/ K33a/ K33b/ K34/ K35	R.ARLE <u>C</u> ^{tox} EINTYR.S
K31/ K34/ K36	R.ILDELTL <u>C</u> ^{tox} K.S
K33a/ K33b	R.ILDELTL <u>C</u> ^{tox} R.S
K31/ K33a/ K33b/ K34	R.LASYLEK ^{dm} VR.Q
K31/ K33a/ K33b/ K34/ K35	R.LEC ^{tox} EINTYR.S

Tableau 5 : Liste des peptides identifiés avec des modifications post-traductionnelles dans un extrait total de cheveu ainsi que les protéines dans lesquelles on retrouve ces séquences en acides aminés. Pour chaque peptide, l'acide aminé modifié est souligné et le type de modification est indiqué en exposant par une lettre avec m (méthylation), dm (diméthylation), tox (tri-oxydation), ox (oxydation) et am (carbamidométhylation).

De plus, certains peptides ont été identifiés avec et sans modification ou encore avec des modifications différentes. Par exemple, le peptide GLTGGFGSHSVCGGFR est commun à K81, K83 et K86. Celui-ci a été identifié sous deux formes, une forme avec une histidine méthylé et une forme avec une cystéine tri-oxydé.

Ceci indique que nous sommes en présence de différentes formes modifiées ou non d'une protéine ou encore qu'une protéine peut porter une modification tandis qu'une autre protéine peut porter l'autre modification. Néanmoins, nous n'avons pas pu dire quelle protéine exactement porte quelle modification.

Toutefois, nous avons pu constater que les kératines basiques étaient plus modifiées que les kératines acides suggérant que celles-ci pourraient être plus exposées vers les surfaces des filaments intermédiaires dans le cortex et que ces protéines pourraient être surexprimées dans la cuticule du cheveu à maturation. En effet, les tri-oxydations des cystéines en acides cystéiques se produisent lors

d'expositions aux UV. Cette supposition est en accord avec les expériences d'expression d'ARNm des kératines dans le follicule pileux qui avaient montré que K85 était exprimée dans la cuticule (Langbein et al., 2001).

III.2.2. Extrait du cortex

a) Les protéines identifiées dans le cortex

La tableau 6 énumère les protéines identifiées dans les extraits du cortex qui font partie des KIFs et des KAPs.

AC	Proteins	Coverage (%)	Number of unique peptides
O43790	Keratin Type II K86	75	43
P78386	Keratin Type II K85	76	47
Q14533	Keratin Type II K81	64	41
P78385	Keratin Type II K83	53	35
Q15323	Keratin Type I K31	69	30
O76009	Keratin Type I K33a	74	31
Q14525	Keratin Type I K33b	73	29
O76011	Keratin Type I K34	57	24
Q9NSB4	Keratin Type II K82	11	6
Q92764	Keratin Type I K35	43	16
Q9BYR9	KAP 2.4	34	4
Q8IUC1	KAP 11.1	24	3
Q9BYR8	KAP 3.1	24	2
Q9BYR7	KAP 3.2	17	1
Q9BYQ4	KAP 9.2	23	3
Q9BYP9	KAP 9.5 or 9.9	25	3
Q9BYR6	KAP 3.3	17	1
Q3LI83	KAP 24.1	5	1
Q3LI72	KAP 19.5	22	1
Q9BYR0	KAP 4.7	16	1
Q9BYR4	KAP 4.3	35	3
Q9BYQ8	KAP 4.9	31	4
Q07627	KAP 1.1	6	1
Q52LG2	KAP 13.2	7	1
Q9BYQ5	KAP 4.6 or 4.15	31	5
Q9BYR3	KAP 4.4 or 4.13	15	2
Q9BYR5	KAP 4.2	22	2
Q9BYQ7	KAP 4.1 or 4.10	10	1

Tableau 6 : Liste des protéines identifiées par analyse LC/nanoLC-Chip-MS/MS du mélange complexe de peptides issus de la digestion à la trypsine de l'extrait du cortex du cheveu avec numéro d'accension, nom de la protéine, pourcentage de recouvrement et nombre de peptides uniques identifiés.

Les protéines majoritaires sont les KIFs avec en tout 10 KIFs identifiées : **5 KIFs de Type II** (K81, K82, K83, K85 et K86) et **5 KIFs de Type I** (K31, K33a, K33b, K34 et K35). Les recouvrements de séquence les plus élevés ont été obtenus pour K81, K85, K86, K31, K33a et K33b suggérant que ces KIFs sont les plus abondants dans le cortex.

Concernant les **KAPs**, nous avons pu en identifier **16 avec des peptides spécifiques** (KAP 1.1, KAP 2.4, KAP 24.1, KAP 11.1, KAP 3.1, KAP 3.2, KAP 3.3, KAP 13.2, KAP 19.5, KAP 4.3, KAP 4.9, KAP 4.6 ou 4.15, KAP 4.7, KAP 4.2, KAP 4.4 ou 4.13 et KAP 4.1 ou 4.10).

b) Les modifications post-traductionnelles des kératines du cortex

Ayant déjà discuté auparavant des problèmes rencontrés lors de l'interprétation des données de masses, nous ne décrivons ici que brièvement les résultats des identifications des protéines avec leurs modifications post-traductionnelles.

Comme dans le cas de l'extrait total de cheveu, nous avons pu identifier les 3 modifications post-traductionnelles suivantes: la **tri-oxydation des cystéines en acides cystéiques** (7 peptides), la **di-méthylation des lysines** (5 peptides) et la **méthylation des histidines** (1 peptide). De plus nous avons pu mettre en évidence une autre modification post-traductionnelle, la **tri-méthylation des lysines** avec un peptide identifié (tableau 7).

Hair cortical extract	
Proteins containing these sequences	Modified peptides identified
K81/ K83/ K85/ K86	K.LAELEGALQK ^{dm} AK.Q
K81/ K83/ K85/ K86	R.C ^{tox} KLAELEGALQK.A
K81/ K85/ K86	R.DLNM ^{ox} DC ^{tox} IIEIK.A
K81/ K83/ K85/ K86	R.EAEC ^{tox} VEADSGR.L
K81/ K83/ K85/ K86	R.FAAFIDK ^{dm} VR.F
K81/ K83/ K85/ K86	R.FLEQQNK ^{dm} LLETK.W
K81/ K83/ K85/ K86	R.FLEQQNK tm LLETK.W
K81/ K83/ K85/ K86	R.TK ^{dm} EEINELNR.M
K81/ K83/ K86	R.GLTGGFGSH ^m SVC ^{am} GGFR.A
K81/ K83/ K86	R.GLTGGFGSHSVC ^{tox} GGFR.A
K31/ K33a/ K33b/ K34/ K35	R.ARLEC ^{tox} EINTYR.S
K31/ K33a/ K33b/ K34	R.LEC ^{tox} EINTYR.S
K31/ K34	R.ILDELTL ^{tox} K.S
K31/ K33a/ K33b/ K34	R.LASYLEK ^{dm} VR.Q

Tableau 7 : Liste des peptides identifiés avec des modifications post-traductionnelles dans un extrait du cortex du cheveu ainsi que les protéines dans lesquelles on retrouve ces séquences en acides aminés. Pour chaque peptide, l'acide aminé modifié est souligné et le type de modification est indiqué en exposant par une lettre avec m (méthylation), dm (di-méthylation), tm (tri-méthylation), tox (tri-oxydation), ox (oxydation) et am (carbamidométhylation).

De même, on constate que les protéines portant le plus de modifications sont les KIFs de Type II. Tous les peptides identifiés avec des modifications sont communs au moins à deux protéines d'une même famille et certains peptides sont identifiés avec différentes modifications comme c'est le cas pour les peptides FLEQQNKLETK et GLTGGFGSHSVCGGFR.

III.2.3. Comparaison des résultats extrait total vs extrait du cortex

Dans un premier temps on constate que pour les KIFs identifiées avec les pourcentages de recouvrement les plus élevées, ceux-ci ne varient que de $\pm 7\%$ pour une même protéine identifiée dans les deux extraits. De même en termes de peptides identifiés, ils sont identiques à un ou deux peptides près (figure 9).

(A) K86 identifiée dans extrait total du cheveu (71% de recouvrement de séquence)

1 MTCGSYCGGR **AFSCISACGP RPRGCCITAA PYRGISCYRG LTGGFGSHSV**
 51 **CGGFRAGSCG RSFGYRSGGV CGPSPPCITT VSVNESLLTP LNLEIDPNAQ**
 101 **CVKQEEKEQI KSLNSRFAAF IDKVRFLEQQ NKLLETKLQF YQNRECCQSN**
 151 **LEPLFEGYIE TLRREAECVE ADGRLASEL NHVQEVLEGY KKKYEEEVSL**
 201 **RATAENEFVA LKKDVDCAYL RKSLEANVE ALIQEIDFLR RLYEEEEIRVL**
 251 **QSHISDTSVV VKLDNSRDNL MDCIIAEIKA QYDDIVTRSR AEAESWYRSK**
 301 **CEEMKATVIR HGETLRR**TKE** EINELNRMIQ RLTAEVENAK CQNSKLEAAV**
 351 **AQSEQQGEAA LSDARCKLAE LEGALQKAKQ DMACLIREYQ EVMNSKLGDL**
 401 **IEIATYRRLI EGEEQRLCEG VGSVNVCVSS SRGGVVCGLD CASTTAPVVS**
 451 **TRVSSVPSNS NVVVGTTNAC APSAR**VGVC GSKRC

(B) K86 identifiée dans extrait du cortex du cheveu (75% de recouvrement de séquence)

1 MTCGSYCGGR **AFSCISACGP RPRGCCITAA PYRGISCYRG LTGGFGSHSV**
 51 **CGGFRAGSCG RSFGYRSGGV CGPSPPCITT VSVNESLLTP LNLEIDPNAQ**
 101 **CVKQEEKEQI KSLNSRFAAF IDKVRFLEQQ NKLLETKLQF YQNRECCQSN**
 151 **LEPLFEGYIE TLRREAECVE ADGRLASEL NHVQEVLEGY KKKYEEEVSL**
 201 **RATAENEFVA LKKDVDCAYL RKSLEANVE ALIQEIDFLR RLYEEEEIRVL**
 251 **QSHISDTSVV VKLDNSRDNL MDCIIAEIKA QYDDIVTRSR AEAESWYRSK**
 301 **CEEMKATVIR HGETLRR**TKE** EINELNRMIQ RLTAEVENAK CQNSKLEAAV**
 351 **AQSEQQGEAA LSDARCKLAE LEGALQKAKQ DMACLIREYQ EVMNSKLGDL**
 401 **IEIATYRRLI EGEEQRLCEG VGSVNVCVSS SRGGVVCGLD CASTTAPVVS**
 451 **TRVSSVPSNS NVVVGTTNAC APSAR**VGVC GSKRC

Figure 9 : Comparaison des recouvrements de séquences de la kératine basique de Type II K86 identifiée dans (A) l'extrait total de cheveu et (B) l'extrait du cortex du cheveu. Les peptides identifiés via le moteur de recherche MASCOT™ sont en gras et soulignés.

Par exemple, les kératines basiques de Type II K81, K83, K85 et K86 ont été identifiées dans les deux extraits avec des recouvrements de séquences proches à $\pm 4\%$. Seule K82 est identifiée avec un recouvrement plus élevée dans l'extrait total de cheveu (31% contre 11%). Cela suggère que

cette protéine mineure est **plus abondante dans la cuticule** que dans le cortex. De même, dans aucun des deux extraits, **K84 n'est identifiée**. Elle semblerait donc absente dans un cheveu mature.

Quant aux kératines acides de Type I, nous avons pu constater que **K32, K36, K37 et K38 n'ont pas été identifiées dans les extraits du cortex** et qu'elles ne sont identifiées dans les extraits totaux qu'avec des recouvrements faibles même si elles ont des sites de coupure de la trypsine permettant de générer des peptides de tailles classiques pour être détectés. De ce fait, cela suggère qu'elles sont absentes dans le cortex et qu'elles ne sont que faiblement représentées dans la cuticule du cheveu mature.

Concernant les **modifications post-traductionnelles** uniquement identifiées sur les **KIFs de Type I et II**, nous avons pu constater que dans les deux extraits, ce sont les kératines basiques de **Type II qui ont le plus grand nombre d'acides aminés modifiés** (8 à 14 modifications contre 1 à 3 pour les kératines de Type I). De plus, parmi ces modifications il y a plus de cystéines tri-oxydées sur les kératines de Type II. Cela laisse à penser que les **kératines basiques** sont plus présentes **vers la surface du cheveu**. Le tableau 8 résume les résultats obtenus concernant les kératines identifiées dans les deux extraits.

AC	Proteins	Total hair extract			Hair cortical extract		
		Coverage (%)	Number of unique peptides	Number of PTMs	Coverage (%)	Number of unique peptides	Number of PTMs
O43790	Keratin Type II K86	71	45	14	75	43	12
P78386	Keratin Type II K85	75	47	10	76	47	8
Q14533	Keratin Type II K81	63	44	11	64	41	12
P78385	Keratin Type II K83	54	40	10	53	35	11
Q9NSB4	Keratin Type II K82	31	18	3	11	6	3
Q15323	Keratin Type I K31	62	31	0	69	30	0
O76009	Keratin Type I K33a	72	33	3	74	31	3
Q14525	Keratin Type I K33-b	72	30	3	73	29	2
O76011	Keratin Type I K34	59	25	3	57	24	2
Q92764	Keratin Type I K35	36	17	1	43	16	1
O76013	Keratin Type I K36	19	6	1	0	0	0
Q14532	Keratin Type I K32	31	11	0	0	0	0
O76014	Keratin Type I K37	19	7	0	0	0	0
O76015	Keratin Type I K38	19	7	0	0	0	0

Tableau 8 : Comparaison des résultats d'identification des kératines dans l'extrait total de cheveu et l'extrait du cortex avec le numéro d'accension, le nom de la protéine, le pourcentage de recouvrement de séquence, le nombre de peptides uniques et le nombre de modifications post-traductionnelles.

Si l'on compare les **KAPs** identifiées dans les deux extraits, il est difficile de se prononcer sur leur présence ou leur absence dans un des deux extraits ainsi que sur leur abondance. En effet, les KAPs ne génèrent que très peu de peptides identifiables et de plus elles sont minoritaires dans les

extraits. Toutefois, nous pouvons constater d'une part qu'elles ne sont pas modifiées et que d'autre part nous avons identifié en tout 24 KAPs (tableau 9).

AC	Proteins	Total hair extract		Hair cortical extract	
		Coverage (%)	Number of unique peptides	Coverage (%)	Number of unique peptides
Q9BYR9	KAP 2.4	42	6	34	4
Q8IUC1	KAP 11.1	24	3	24	3
Q9BYR8	KAP 3.1	24	2	24	2
Q9BYR7	KAP 3.2	17	1	17	1
Q9BYQ4	KAP 9.2	19	2	23	3
Q9BYP9	KAP 9.5 or 9.9	21	2	25	3
Q9BYR6	KAP 3.3	17	1	17	1
Q3LI83	KAP 24.1	5	1	5	1
Q3LI72	KAP 19.5	22	1	22	1
Q9BYR0	KAP 4.7	21	4	16	1
Q9BYR4	KAP 4.3	17	1	35	3
Q9BYQ8	KAP 4.9	22	2	31	4
Q07627	KAP 1.1	6	1	6	1
Q52LG2	KAP 13.2	7	1	7	1
Q9BYQ5	KAP 4.6 or 4.15	23	4	31	5
Q9BYR2	KAP 4.5	10	2	0	0
Q9BYQ9	KAP 4.8	16	2	0	0
P60331	KAP 10.1	8	3	0	0
P60369	KAP 10.3	10	3	0	0
P60409	KAP 10.7	6	3	0	0
Q9BQ66	KAP 4.12	22	4	0	0
Q9BYR3	KAP 4.4 or 4.13	0	0	15	2
Q9BYR5	KAP 4.2	0	0	22	2
Q9BYQ7	KAP 4.1 or 4.10	0	0	10	1

Tableau 9 : Comparaison des résultats d'identification des KAPs dans l'extrait total de cheveu et l'extrait du cortex avec le numéro d'accèsion, le nom de la protéine, le pourcentage de recouvrement de séquence et le nombre de peptides uniques.

Vingt-deux d'entre elles appartiennent au groupe riche en soufre (**HSPs**) et très riche en soufre (**UHSPs**), une appartient au groupe riche en glycine et tyrosine (KAP 19.5) et la KAP 24.1 mise en évidence par Rogers et al. en 2007 [Rogers et al., 2007] qui semble être spécifiquement exprimée dans la cuticule du follicule pileux. Dans notre cas nous avons pu identifier cette protéine dans les deux extraits suggérant ainsi que dans un cheveu mature cette protéine est présente dans le cortex.

Enfin, nous n'identifions aucun peptide des KAPs de la famille 10 dans les extraits du cortex. Cela suggère que cette famille de KAPs est plutôt présente dans la cuticule. Ce résultat est en accord avec les expériences de Rogers qui a montré que la famille de KAP 10 est uniquement exprimée dans la cuticule du follicule pileux [Rogers et al., 2004].

IV. Discussion

Nous avons décrit ici **deux approches complémentaires** pour une étude détaillée des protéines présentes dans un cheveu mature. Dans un **premier temps**, nous avons effectué une étude des **protéines du cortex** en utilisant l'approche basée sur la séparation des protéines **sur gel 2D** suivie de l'analyse **nanoLC-Chip-MS/MS** des peptides issus de **85 spots**. Dans un **second temps**, afin de compléter cette première étude, nous avons utilisé une approche basée sur la **séparation de mélanges complexes de peptides** par **LC 2D couplée à la MS**. Pour cette seconde étude, nous avons travaillé sur un **extrait total de cheveu** (cortex et cuticule) et un **extrait du cortex**. Chacune de ces deux approches nous a apporté de nouvelles informations primordiales pour une meilleure compréhension de cette machinerie complexe qu'est le cheveu.

En effet, après avoir optimisé plusieurs paramètres pour obtenir la meilleure séparation sur gel 2D, cette première approche nous a permis de montrer la **forte hétérogénéité dans la nature des kératines** présentes dans le cheveu. Nous avons pu identifier plusieurs kératines sur l'ensemble du gel 2D **dans des gammes de PM et de pI qui ne correspondent pas aux leurs**. Par exemple, les kératines de Type II K85 et K86 qui ont respectivement des PM de 56 kDa et 53 kDa et des pI de 6,3 et 5,6 ont été identifiées dans plusieurs spots dans la gamme des PM 30 kDa et des pI 9-10. Ces différences dans les zones de migration s'expliquent par le fait que nous sommes en présence de **formes dégradées** avec des PTMs de type **méthylation, di-méthylation et tri-méthylation**.

L'approche gel 2D nous a également permis de montrer que nous avons la **présence de liaisons inter- et intramoléculaires de type glutamine-lysine dans le cheveu**. En effet, nous avons pu identifier **dans des mêmes spots des kératines et des KAPs**. Or les KAPs étant des petites protéines basiques, elles ne devraient pas être identifiées dans les gammes de PM 30 kDa entre les pI 4 à 11. Les extraits de protéines ayant été traités pour réduire et alkyliser les ponts disulfures, fortement présents dans le cheveu, ces différences de migration pourraient s'expliquer par la présence de liaisons intermoléculaires entre kératines et KAPs d'un nouveau type. La **présence de TGase 3** dans le cheveu laisse supposer que ces **liaisons sont de type N^ε-(γ-glutamyl) lysine**.

En effet, la TGase est une enzyme qui, après avoir été activée par la présence de calcium, catalyse la formation de liaisons isopeptidiques entre le groupe ϵ -NH₂ d'une lysine et le groupe γ -carboxamide d'une glutamine pour former des agrégats macromoléculaires insolubles [Folk and Finlayson, 1977]. Cette hypothèse a pu être confirmée d'une part par l'analyse des acides aminés issus de la digestion aspécifiques des protéines présentes dans plusieurs spots du gel 2D et d'autre part par des expériences *in vitro* de la réaction de la TGase 3 avec la kératine acide de Type I K33a et un peptide synthétique de la KAP 3.1 (annexe 4).

De plus, cette hypothèse peut être renforcée si l'on se réfère aux séquences en acides aminés de ces deux types de protéines. En effet, comme nous avons pu le voir dans la partie introductive, les kératines acides et basiques contiennent trois domaines, la tête (N-terminal) et la queue (C terminal) toutes deux riches en cystéines et la tige centrale pauvre en cystéines. Ainsi, nous pouvons dans un premier temps en déduire que les **liaisons disulfures** qui stabilisent les filaments intermédiaires sont plutôt **présentes aux extrémités des kératines**.

De même, si on s'intéresse à la composition en acide aminé de la tige centrale, on constate qu'il y a quatre acides aminés qui sont les plus représentés : l'acide glutamique (E), la leucine (L), l'alanine (A) pour les deux types de kératines. De plus pour chacun des deux types de kératines, un cinquième acide aminé vient s'ajouter aux quatre premiers : la glutamine pour les KIFs acides de Type I et la lysine pour les KIFs basiques de Type II. Cette complémentarité particulière (glutamine/lysine) entre les deux types de kératines renforce l'hypothèse de **liaisons glutamine-lysine entre KIFs** qui joueraient un rôle de stabilisation dans l'association de celles-ci en filaments intermédiaires.

Enfin, à ces associations intermoléculaires entre KIFs, viennent également s'ajouter des associations avec les KAPs qui forment la matrice protectrice des filaments intermédiaires. En effet, deux familles de KAPs ont majoritairement été identifiées : des KAPs appartenant à la famille des HSPs qui sont riches en glutamines et en lysines, et des KAPs de la famille des UHSPs qui sont très riches en soufre. De ce fait, on peut supposer que les **UHSPs** sont plutôt impliquées dans la formation de **ponts disulfures en tête et queue des kératines** et que les **HSPs** vont préférentiellement former des **liaisons intermoléculaire** de type N^{ϵ} -(γ -glutamyl) lysine dans la **tige centrale des kératines** (figure 10).

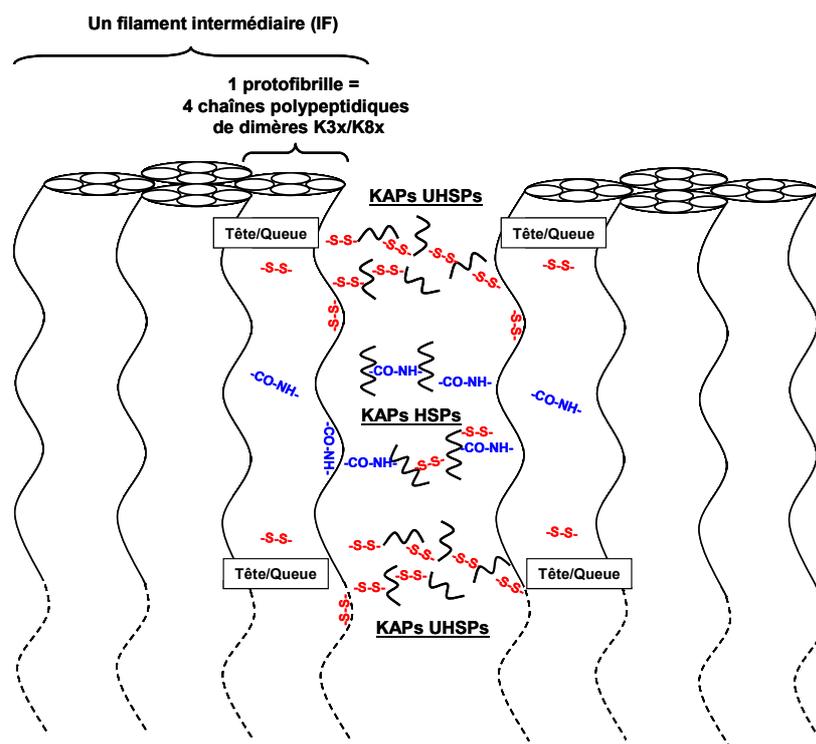


Figure 10 : Représentation schématique d'un modèle structurel montrant quels types d'interactions peuvent être mis en jeu dans les filaments intermédiaires et entre ceux-ci et la matrice protectrice qui l'entoure.

Enfin, l'identification de **deux protéines de la famille de S100** qui transportent le calcium, permet également de renforcer cette hypothèse. En effet, celle-ci pourraient **fournir le calcium à la TGase 3**, activant ainsi cette enzyme pour qu'elle engendre des liaisons isopeptidiques entre kératines

et KAPs. Ces protéines étant également **riches en lysines**, elles pourraient **s'associer aux kératines de Type I ou aux KAPs de la famille des HSPs**.

Cependant, à l'heure actuelle il ne nous a pas encore été possible de déterminer quelles sont les associations exactes de protéines qu'il existe dans le cheveu, car en effet, nous pouvons avoir de nombreuses combinaisons tels que kératine/kératine, kératine/KAP, KAP/KAP, kératine/S100, KAP/S100 et bien d'autres.

Néanmoins, l'analyse de mélanges complexes de protéines par l'approche gel 2D suivi d'analyses nanoLC-Chip-MS/MS étant limité pour l'identification de modifications post-traductionnelles et la détection de protéines minoritaires ou de protéines aux propriétés physico-chimiques extrêmes (PM bas et pI extrêmes), nous avons complété cette première étude par une approche de **séparation de mélanges complexes de peptides par LC 2D couplée à la MS**. Celle-ci a été utilisée pour étudier différenciellement le contenu protéique d'un extrait total de cheveu et d'un extrait du cortex. L'ensemble nous a permis d'aller vers une **caractérisation fine des protéines du cheveu** puisque nous avons pu **identifier les KIFs avec des recouvrements de séquences très élevées** ainsi qu'un certain nombre de **PTMs** et de **localiser certaines protéines dans un cheveu à maturation**.

Globalement, les protéines identifiées dans cette seconde étude peuvent se regrouper en deux grandes catégories : des **protéines de structures** (les KIFs et les KAPs) et des **protéines** identifiées par la présence de protéines **dégradées témoignant d'une ancienne activité cellulaire**. Ces dernières, identifiées avec des recouvrements très faibles, sont considérées comme n'ayant qu'un impact minime dans la qualité de la fibre. Ainsi, nous n'avons pas cherché à explorer cette catégorie. En revanche nous avons cherché à être le plus complet possible dans la caractérisation des KIFs et des KAPs.

Ainsi nous avons pu identifier toutes les KIFs de Type I et de Type II, mise à part la K84, avec des recouvrements de séquences très élevées. En comparant ceux-ci sur plusieurs expériences, nous avons pu rapidement constater que les **séquences peptidiques identifiées** et la **nature des recouvrements** des protéines étaient **reproductibles d'une expérience à une autre pour les KIFs**. Cette première observation est intéressante et pourrait nous permettre de **confirmer l'hypothèse de liaisons intermoléculaires de type glutamine-lysine**. En effet, on constate que **les peptides non identifiés** par spectrométrie de masse ont souvent **des glutamines et des lysines dans leurs séquences** en acides aminés (par exemple MIQR, CQNSK, CQLGDR, SKCEEMKATVIR). Même si certains de ces peptides ont des PM trop faibles pour qu'ils soient détectables dans les gammes de m/z analysables par la trappe à ion, rien ne prouve qu'ils ne soient pas **impliqués dans ce type de liaisons**.

De plus, nous avons pu constater que les **recouvrements de séquences** obtenus pour les KIFs acides de **Type I** sont **plus faibles** que ceux des KIFs basiques de Type II. Cela laisse supposer que les **Type I pourraient être plus impliquées dans les liaisons intermoléculaires**. Cette hypothèse selon laquelle les KIFs acides de Type I forment les liaisons intermoléculaires peut également être confortée si l'on regarde de plus près les identifications de PTMs.

En effet, nous avons pu mettre en évidence l'existence de **PTMs de type méthylation, di-méthylation et tri-méthylation** sur les **lysines** et les **histidines** des **KIFs**. Ces modifications semblent être **plus présentes** sur les KIFs basiques de **Type II** (8 à 14 PTMs) que sur les KIFs acides de Type I

En effet, dans un premier temps, **K84** n'a été identifiée dans aucune des expériences. Cette protéine semble donc ne **pas** être **exprimée dans le cheveu mature**, ce qui est en accord avec les études menés par Langbein et Rogers [Langbein et al., 1999], [Langbein et al., 2001]. De même, **K82** a été identifiée avec des pourcentages de recouvrement plus faibles dans le cortex suggérant que celle-ci est **plus exprimée dans la cuticule**. Cette supposition est d'ailleurs en accord avec l'expérience réalisée en séparant les protéines du cortex sur gel 2D puisque cette protéine n'a été identifiée dans aucun spot. Les protéines **K32, K36, K37 et K38** n'ont été identifiées que dans les extraits totaux, cela suggère qu'elles ne sont **exprimées que dans la cuticule**. Concernant **K32** celle-ci a été identifiée dans l'expérience gel 2D mais toutefois que dans 3 spots et avec un peptide qui de plus est commun à d'autres kératines. De ce fait, dans ce cas là, nous pouvons dire que la LC 2D nous permet de trancher et donc de dire que cette protéine n'est **peut-être pas présente dans le cortex** d'un cheveu à maturation.

Toutefois, dans certains cas, notamment en ce qui concerne les KIFs identifiés dans les deux extraits, même si nous pouvons affirmer que celles-ci sont présentes dans le cortex, nous ne pouvons pas affirmer qu'elles sont ou non présentes dans la cuticule du cheveu, d'autant plus que les expériences de marquage de ces KIFs dans le follicule pileux réalisées par Rogers et al. ont montré que seul 4 KIFs sont exprimées dans sa cuticule, la K82, K85, K32 et K35 [Langbein et al., 1999], [Langbein et al., 2001].

Concernant les KAPs, il est difficile de suggérer leur localisation dans le cheveu. En effet, de manière générale ces protéines sont faiblement représentées par rapport aux KIFs dans le cheveu. De plus, leurs séquences en acides aminés très particulières (nombre élevé de cystéines et peu de site de coupures) ne permettent que de générer très peu de peptides de digestion identifiables par spectrométrie de masse. Enfin, chaque famille de KAPs présente de fortes homologues de séquences et les peptides générés après digestion sont souvent communs à une famille. Ainsi, les KAPs sont identifiées avec très peu de peptides, donc des recouvrements de séquences très faibles et souvent nous identifions des familles de KAPs et non des KAPs individuelles.

Toutefois, l'approche LC 2D a permis de montrer comme dans le cas de l'approche gel 2D que les **KAPs majoritairement présentes dans le cheveu** appartiennent aux groupes des **UHSPs** et des **HSPs**, indiquant que ce sont ces deux types de protéines qui **entrent en jeu dans la formation de liaisons avec les filaments intermédiaires**. D'autre part aucune PTM n'a été identifiée sur les KAPs, ce qui suggère qu'elles sont très résistantes aux attaques extérieures.

V. Conclusion

L'objectif de cette étude menée en collaboration avec L'ORÉAL, était de mettre **au point des méthodes analytiques** pour **l'identification de l'ensemble des protéines de la tige capillaire**, et cela afin de mieux **comprendre la structure du cheveu** en termes de **localisation des protéines, modifications des protéines et interactions protéines-protéines**.

Le cheveu formant un complexe filamenteux très stable par des liaisons inter- et intramoléculaires, l'étude des protéines de celui-ci reste difficile. Ainsi, il a fallu adapter les méthodes d'extraction et de solubilisation aux méthodes de séparations tout en restant compatible avec l'analyse par spectrométrie de masse.

En utilisant **deux approches complémentaires**, la **séparation des protéines sur gel d'électrophorèse 2D** et la **séparation des peptides par LC 2D**, nous avons pu mettre en évidence que:

- Les familles de **KAPs majoritairement présentes** dans le cheveu appartiennent aux groupes **des UHSPs et des HSPs**.
- Les **KIFs** semblent **se dégrader** durant la pousse du cheveu.
- Les **KIFs** portent de **nombreuses PTMs** de type **méthylation** des histidines, **di-, tri-méthylation des lysines** et **tri-oxydation des cystéines en acides cystéiques**. De plus, les KIFs de **Type II sont plus modifiées**.
- Les **KAPs** et les **KIFs** forment des **liaisons inter- et intramoléculaires** de type « **glutamine-lysine** » engendrées par la présence de transglutaminase **TGase 3** dans le cheveu.

En combinant tous ces résultats obtenus, nous avons pu apporter de nouvelles informations primordiales pour une meilleure compréhension de cette machinerie complexe qu'est le cheveu en termes de protéines identifiées dans un cheveu mature, des dégradations et modifications qu'elles subissent, de leurs interactions et de leur localisation dans le cheveu (cortex ou cuticule). En effet, ceci nous a permis de **suggérer un modèle structurale du cheveu** avec les KAPs **UHSPs** plus impliquées dans la formation de **ponts disulfures** et les KAPs **HSPs** plus impliquées dans les **liaisons intermoléculaires glutamine-lysine**. De plus nous avons pu montrer que les KIFs de **Type II** sont **plus soumises aux dégradations** suggérant qu'elles sont plus exposées dans les IFs.

En **perspective**, nous envisageons de **mettre en évidence** la présence de cette **liaison intermoléculaires de type glutamine-lysine dans un couple de protéine** et par la suite de **situer cette liaison sur ces couples de kératines**. Cette deuxième étape sera longue et difficile puisqu'à l'heure actuelle, nous ne savons pas si nous avons une ou plusieurs liaisons entre un ou plusieurs couples kératines/KAPs et de ce fait comment ces isopeptides complexes se comporteront en spectrométrie de masse (taille, ionisation, fragmentation, spécifique ou non d'une protéine, ...).

Enfin, l'utilisation de ces deux stratégies de séparation ainsi que l'analyse par spectrométrie de masse pourra être utilisée pour mettre en évidence l'influence des différents traitements sur la nature des protéines dans un cheveu à maturation ainsi que d'étudier la qualité d'un cheveu selon son ethnie.

BIBLIOGRAPHIE

- Dawber R.** (1996). Hair: its structure and response to cosmetic preparations. *Clin Dermatol* **14**, 105-112.
- Dawber R.** (2002). Cosmetic and medical causes of hair weathering. *J Cosmet Dermatol* **1**, 196-201.
- Eckert R.L., Sturniolo M.T., Broome A.M., Ruse M. and Rorke E.A.** (2005). Transglutaminase function in epidermis. *J Invest Dermatol* **124**, 481-492.
- Folk J.E. and Finlayson J.S.** (1977). The epsilon-(gamma-glutamyl)lysine crosslink and the catalytic role of transglutaminases. *Adv Protein Chem* **31**, 1-133.
- Hearle J.W.** (2000). A critical review of the structural mechanics of wool and hair fibres. *Int J Biol Macromol* **27**, 123-138.
- Hearle J.W.** (2007). Proteins fibers: structural mechanics and future opportunities. *J. Mater. Sci.* **42**, 8010-8019.
- Inoue T., Sasaki I., Yamaguchi M. and Kizawa K.** (2000). Elution of S100A3 from hair fiber: new model for hair damage emphasizing the loss of S100A3 from cuticle. *J. Cosmet. Sci.* **51**, 15-25.
- Jones L.N.** (2001). Hair structure anatomy and comparative anatomy. *Clin Dermatol* **19**, 95-103.
- Kim H.J. and Yim H.** (1999). Investigation of Transglutaminase-Induced Peptide Cross-Linking by Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry. *Bull. Korean Chem. Soc.* **20**, 1299-1302.
- Kizawa K., Uchiwa H. and Murakami U.** (1996). Highly-expressed S100A3, a calcium-binding protein, in human hair cuticle. *Biochim Biophys Acta* **1312**, 94-98.
- Langbein L., Rogers M.A., Winter H., Praetzel S., Beckhaus U., Rackwitz H.R. and Schweizer J.** (1999). The catalog of human hair keratins. I. Expression of the nine type I members in the hair follicle. *J Biol Chem* **274**, 19874-19884.
- Langbein L., Rogers M.A., Winter H., Praetzel S. and Schweizer J.** (2001). The catalog of human hair keratins. II. Expression of the six type II members in the hair follicle and the combined catalog of human type I and II keratins. *J Biol Chem* **276**, 35123-35132.
- Langbein L. and Schweizer J.** (2005). Keratins of the human hair follicle. *Int Rev Cytol* **243**, 1-78.
- Lee Y.J., Rice R.H. and Lee Y.M.** (2006). Proteome analysis of human hair shaft: from protein identification to posttranslational modification. *Mol Cell Proteomics* **5**, 789-800.
- Marshall R.C. and Blagrove R.J.** (1979). Successful isoelectric focusing of wool low-sulfur proteins. *J Chromatogr* **172**, 351-356.
- Parry D.A., Smith T.A., Rogers M.A. and Schweizer J.** (2006). Human hair keratin-associated proteins: sequence regularities and structural implications. *J Struct Biol* **155**, 361-369.
- Parry D.A., Strelkov S.V., Burkhard P., Aebi U. and Herrmann H.** (2007). Towards a molecular description of intermediate filament structure and assembly. *Exp Cell Res* **313**, 2204-2216.
- Plowman J.E.** (2003). Proteomic database of wool components. *J Chromatogr B Analyt Technol Biomed Life Sci* **787**, 63-76.

Plowman J.E., Bryson W.G., Flanagan L.M. and Jordan T.W. (2002). Problems associated with the identification of proteins in homologous families: the wool keratin family as a case study. *Anal Biochem* **300**, 221-229.

Popescu C. and Hocker H. (2007). Hair--the most sophisticated biological composite material. *Chem Soc Rev* **36**, 1282-1291.

Robinson N.A. and Eckert R.L. (1998). Identification of transglutaminase-reactive residues in S100A11. *J Biol Chem* **273**, 2721-2728.

Rogers M.A., Langbein L., Praetzel-Wunder S., Winter H. and Schweizer J. (2006). Human hair keratin-associated proteins (KAPs). *Int Rev Cytol* **251**, 209-263.

Rogers M.A., Langbein L., Winter H., Beckmann I., Praetzel S. and Schweizer J. (2004). Hair keratin associated proteins: characterization of a second high sulfur KAP gene domain on human chromosome 21. *J Invest Dermatol* **122**, 147-158.

Rogers M.A., Winter H., Langbein L., Wollschlager A., Praetzel-Wunder S., Jave-Suarez L.F. and Schweizer J. (2007). Characterization of human KAP24.1, a cuticular hair keratin-associated protein with unusual amino-acid composition and repeat structure. *J Invest Dermatol* **127**, 1197-1204.

Ruse M., Lambert A., Robinson N., Ryan D., Shon K.J. and Eckert R.L. (2001). S100A7, S100A10, and S100A11 are transglutaminase substrates. *Biochemistry* **40**, 3167-3173.

Schmidt M., Gillitzer R., Toksoy A., Brocker E.B., Rapp U.R., Paus R., Roth J., Ludwig S. and Goebeler M. (2001). Selective expression of calcium-binding proteins S100a8 and S100a9 at distinct sites of hair follicles. *J Invest Dermatol* **117**, 748-750.

Smith T.A. and Parry D.A. (2007). Sequence analyses of Type I and Type II chains in human hair and epithelial keratin intermediate filaments: promiscuous obligate heterodimers, Type II template for molecule formation and a rationale for heterodimer formation. *J Struct Biol* **158**, 344-357.

Smith T.A. and Parry D.A. (2008). Three-dimensional modelling of interchain sequence similarities and differences in the coiled-coil segments of keratin intermediate filament heterodimers highlight features important in assembly. *J Struct Biol* **162**, 139-151.

Thibaut S., Candi E., Pietroni V., Melino G., Schmidt R. and Bernard B.A. (2005). Transglutaminase 5 expression in human hair follicle. *J Invest Dermatol* **125**, 581-585.

Wolfram L.J. (2003). Human hair: a unique physicochemical composite. *J Am Acad Dermatol* **48**, S106-114.

Yoneda K., Akiyama M., Morita K., Shimizu H., Imamura S. and Kim S.Y. (1998). Expression of transglutaminase 1 in human hair follicles, sebaceous glands and sweat glands. *Br J Dermatol* **138**, 37-44.

ANNEXE 3

TABLEAU COMPLET DES PROTEINES IDENTIFIEES PAR APPROCHE GEL 2D SUIVI D'ANALYSES NANOLC-MS/MS

Spots	AC	Protein names	Mass (Da)	pI	Number of unique peptides	Coverage (%)	Mascot ions score	Error (ppm)	Case of proteins identified with one unique peptide Peptide sequences	Sequences of modified peptides
1	P78386	Keratin type II K85	55766	6,28	2	3	80	17		
2	Q14533	Keratin type II K81	54936	5,48	4	9	122	37		
	O43790	Keratin type II K86	53466	5,56	4	9	122	37		
	P78386	Keratin type II K85	55766	6,28	3	7	94	47		
3	Q14533	Keratin type II K81	54936	5,48	4	5	151	42		
	P78385	Keratin type II K83	54179	5,34	4	5	151	42		
	O43790	Keratin type II K86	53466	5,56	4	5	151	42		
	P78386	Keratin type II K85	55766	6,28	4	5	130	47		
4	Q14533	Keratin type II K81	54936	5,48	1	1	46	15		
5	O43790	Keratin type II K86	53466	5,56	5	10	180	32		
6	O43790	Keratin type II K86	53466	5,56	6	18	465	25		
	P78386	Keratin type II K85	55766	6,28	5	15	310	54		
	Q9BYR2	Keratin-associated protein 4.5	19900	8,24	1	5	57	29	R.PTCamVISTCamPR.P	
	Q9BYQ5	Keratin-associated protein 4.6	21807	8,43	1	4	57	29	R.PTCamVISTCamPR.P	
	Q9BYR0	Keratin-associated protein 4.7	22472	8,35	1	4	57	29	R.PTCamVISTCamPR.P	
	Q9BYQ9	Keratin-associated protein 4.8	20704	8,23	1	4	57	29	R.PTCamVISTCamPR.P	
	Q9BQ66	Keratin-associated protein 4.12	21390	8,37	1	4	57	29	R.PTCamVISTCamPR.P	
7	P78386	Keratin type II K85	55766	6,28	3	10	151	35		
	O43790	Keratin type II K86	53466	5,56	2	6	90	32		
	Q15323	Keratin, type I K31	47202	4,87	1	2	70	84	R.CamGPCamNSFVR.-	
8	P78386	Keratin type II K85	55766	6,28	4	11	193	55		
	O43790	Keratin type II K86	53466	5,56	4	10	200	35		
	Q14533	Keratin type II K81	54936	5,48	3	6	181	29		
9	P78386	Keratin type II K85	55766	6,28	6	17	257	53		
	O43790	Keratin type II K86	53466	5,56	4	9	181	37		
	Q14533	Keratin type II K81	54936	5,48	4	8	175	35		
10	Q9BYR8	Keratin-associated protein 3.1	10532	5,99	1	17	93	59	R.SCamSVPTGPATTFCamSFDK.S	

11	P78386	Keratin type II K85	55766	6,28	2	4	95	60		
	O43790	Keratin type II K86	53466	5,56	2	6	87	63		
	Q9BYR8	Keratin-associated protein 3.1	10532	5,99	1	17	83	59	R.SCamSVPTGPATTFCamSFDK.S	
12	P78386	Keratin type II K85	55766	6,28	10	21	511	34		
	O43790	Keratin type II K86	53466	5,56	9	18	444	31		
13	O43790	Keratin type II K86	53466	5,56	15	31	529	41		
	Q14533	Keratin type II K81	54936	5,48	15	30	517	39		
	P78386	Keratin type II K85	55766	6,28	13	24	396	42		
	Q15323	Keratin, type I K31	47202	4,87	8	27	354	49		
	Q9BYR8	Keratin-associated protein 3.1	10532	5,99	1	17	86	59	R.SCamSVPTGPATTFCamSFDK.S	
14	P78386	Keratin type II K85	55766	6,28	19	29	599	47		
	O43790	Keratin type II K86	53466	5,56	19	32	594	49		
	O76009	Keratin, type I K33a	45906	4,82	12	28	558	44		
	Q15323	Keratin, type I K31	47202	4,87	11	33	476	45		
	Q14525	Keratin, type I K33b	46185	4,81	8	20	278	35		
	O76011	Keratin, type I K34	44661	4,74	8	19	266	39		
	Q92764	Keratin, type I K35	47556	4,75	4	8	155	30		
	Q9BYR8	Keratin-associated protein 3.1	10532	5,99	1	17	120	59	R.SCamSVPTGPATTFCamSFDK.S	
15	P78386	Keratin type II K85	55766	6,28	3	10	100	37		
	Q9BYR8	Keratin-associated protein 3.1	10532	5,99	1	17	87	59	R.SCamSVPTGPATTFCamSFDK.S	
16	Q14533	Keratin type II K81	54936	5,48	3	6	105	72		
	O43790	Keratin type II K86	53466	5,56	3	6	105	72		
	P78386	Keratin type II K85	55766	6,28	2	4	99	50		
	Q9BYS1	Keratin-associated protein 1.5	17997	6,59	2	10	66	153		
17	O43790	Keratin type II K86	53466	5,56	9	22	721	39		
	P78386	Keratin type II K85	55766	6,28	9	21	680	35		
18	P78386	Keratin type II K85	55766	6,28	4	8	125	64		
19	P78386	Keratin type II K85	55766	6,28	9	23	384	47		
	O43790	Keratin type II K86	53466	5,56	6	16	292	49		
20	O43790	Keratin type II K86	53466	5,56	10	23	665	47		

	P78386	Keratin type II K85	55766	6,28	7	14	384	52		
21	P78386	Keratin type II K85	55766	6,28	7	15	297	54		
22	O43790	Keratin type II K86	53466	5,56	9	19	530	43		
	P78386	Keratin type II K85	55766	6,28	11	23	523	48		
23	O43790	Keratin type II K86	53466	5,56	17	36	1880	48		K.LAELEGALQKdmAK.Q
	P78386	Keratin type II K85	55766	6,28	17	27	1243	50		
	Q14533	Keratin type II K81	54936	5,48	14	24	1179	47		K.LAELEGALQKdmAK.Q
	Q14525	Keratin, type I K33b	46185	4,81	5	12	155	41		
	Q15323	Keratin, type I K31	47202	4,87	5	12	155	41		
24	O43790	Keratin type II K86	53466	5,56	13	31	579	71		
	P78386	Keratin type II K85	55766	6,28	12	30	551	42		
	Q14533	Keratin type II K81	54936	5,48	12	28	538	71		
	Q15323	Keratin, type I K31	47202	4,87	9	22	340	59		
	O76011	Keratin, type I K34	44661	4,74	6	15	246	49		
25	O43790	Keratin type II K86	53466	5,56	17	39	869	38		
	Q14533	Keratin type II K81	54936	5,48	15	30	790	38		
	P78386	Keratin type II K85	55766	6,28	12	27	619	41		
	Q14525	Keratin, type I K33b	46185	4,81	11	26	307	35		
	Q15323	Keratin, type I K31	47202	4,87	9	21	301	39		
	O76011	Keratin, type I K34	44661	4,74	8	19	283	40		
26	O43790	Keratin type II K86	53466	5,56	23	51	977	53		
	Q15323	Keratin, type I K31	47202	4,87	14	36	910	56		
	Q14533	Keratin type II K81	54936	5,48	21	43	792	47		
	O76009	Keratin, type I K33a	45906	4,82	14	40	681	50		
	Q14525	Keratin, type I K33b	46185	4,81	13	35	665	57		
	O76011	Keratin, type I K34	44661	4,74	11	31	589	60		
	P78385	Keratin type II K83	54179	5,34	17	33	578	45		
	P78386	Keratin type II K85	55766	6,28	15	30	546	45		
27	Q14533	Keratin type II K81	54936	5,48	9	17	320	57		
	O43790	Keratin type II K86	53466	5,56	9	18	319	61		

	P78386	Keratin type II K85	55766	6,28	5	10	138	42		
	Q15323	Keratin, type I K31	47202	4,87	1	2	54	5	R.DNAELENLIR.E	
	O76009	Keratin, type I K33a	45906	4,82	1	2	54	5	R.DNAELENLIR.E	
	Q14525	Keratin, type I K33b	46185	4,81	1	2	54	5	R.DNAELENLIR.E	
28	P78386	Keratin type II K85	55766	6,28	2	4	96	48		
	Q9BYR9	Keratin-associated protein 2.4	13469	8,32	1	5	49	6	R.PVTCamVPR.C	
29	Q14533	Keratin type II K81	54936	5,48	6	17	284	43		R.GLTGGFGSHmSVCamGGFR.A
	O43790	Keratin type II K86	53466	5,56	6	17	284	43		R.GLTGGFGSHmSVCamGGFR.A
	Q15323	Keratin, type I K31	47202	4,87	1	2	77	5	R.DNAELENLIR.E	
	O76009	Keratin, type I K33a	45906	4,82	1	2	77	5	R.DNAELENLIR.E	
	Q14525	Keratin, type I K33b	46185	4,81	1	2	77	5	R.DNAELENLIR.E	
	P78386	Keratin type II K85	55766	6,28	3	5	54	48		
30	Q14533	Keratin type II K81	54936	5,48	9	20	678	55		R.GLTGGFGSHmSVCamGGFR.A
	O43790	Keratin type II K86	53466	5,56	9	20	678	55		R.GLTGGFGSHmSVCamGGFR.A
	P78386	Keratin type II K85	55766	6,28	6	15	260	55		
	Q9BYR8	Keratin-associated protein 3.1	10532	5,99	1	17	83	59	R.SCamSVPTGPATTFCamSFDK.S	
	Q92764	Keratin, type I K35	47556	4,75	1	2	56	29	R.SFSACamSVGLGR.S	
31	Q14533	Keratin type II K81	54936	5,48	8	18	434	58		R.GLTGGFGSHmSVCamGGFR.A
	P78386	Keratin type II K85	55766	6,28	4	7	100	55		
32	O43790	Keratin type II K86	53466	5,56	9	23	558	44		R.GLTGGFGSHmSVCamGGFR.A
	P78386	Keratin type II K85	55766	6,28	4	7	134	45		
33	Q14533	Keratin type II K81	54936	5,48	4	8	179	85		
	O43790	Keratin type II K86	53466	5,56	4	8	179	85		
34	P78386	Keratin type II K85	55766	6,28	12	25	500	49		
	O43790	Keratin type II K86	53466	5,56	10	23	403	50		
	P06702	S100-A9	13234	5,71	1	11	111	3	K.QLSFEEFIMLMAR.L	
35	O43790	Keratin type II K86	53466	5,56	9	24	388	41		
	P06702	S100-A9	13234	5,71	1	11	91	3	K.QLSFEEFIMLMAR.L	
36	P78386	Keratin type II K85	55766	6,28	8	21	282	32		
	O43790	Keratin type II K86	53466	5,56	7	18	236	55		

	P06702	S100-A9	13234	5,71	1	11	111	3	K.QLSFEEFIMLMAR.L	
37	Q14533	Keratin type II K81	54936	5,48	7	16	358	59		
	O43790	Keratin type II K86	53466	5,56	7	16	358	59		
	P06702	S100-A9	13234	5,71	1	11	176	3	K.QLSFEEFIMLMAR.L	
	P78386	Keratin type II K85	55766	6,28	3	8	168	69		
38	Q14533	Keratin type II K81	54936	5,48	7	14	216	58		
	O43790	Keratin type II K86	53466	5,56	7	14	216	58		
	P78386	Keratin type II K85	55766	6,28	5	11	175	56		
	P06702	S100-A9	13234	5,71	1	11	111	3	K.QLSFEEFIMLMAR.L	
	Q8IUUC0	Keratin-associated protein 13.1	18308	8,54	1	11	50	43	R.SLGYGGCamGFPSLGYGVGFCamR.P	
39	Q14533	Keratin type II K81	54936	5,48	7	15	421	59		
	O43790	Keratin type II K86	53466	5,56	7	15	421	59		
	P06702	S100-A9	13234	5,71	1	11	110	3	K.QLSFEEFIMLMAR.L	
40	Q14533	Keratin type II K81	54936	5,48	9	20	432	66		
	O43790	Keratin type II K86	53466	5,56	9	20	432	66		
	P06702	S100-A9	13234	5,71	1	11	98	3	K.QLSFEEFIMLMAR.L	
41	O43790	Keratin type II K86	53466	5,56	14	35	782	47		R.GLTGGFGSHmSVCamGGFR.A
	P78386	Keratin type II K85	55766	6,28	8	23	391	52		
42	Q14533	Keratin type II K81	54936	5,48	8	18	469	52		
	O43790	Keratin type II K86	53466	5,56	8	18	469	52		
	P06702	S100-A9	13234	5,71	1	11	177	3	K.QLSFEEFIMLMAR.L	
43	Q14533	Keratin type II K81	53466	5,56	11	21	493	50		
	O43790	Keratin type II K86	53466	5,56	11	21	493	50		
	P78386	Keratin type II K85	55766	6,28	7	16	348	81		
44	Q14533	Keratin type II K81	54936	5,48	10	21	468	44		
	O43790	Keratin type II K86	53466	5,56	10	21	468	44		
	P06702	S100-A9	13234	5,71	1	11	111	3	K.QLSFEEFIMLMAR.L	
45	O43790	Keratin type II K86	53466	5,56	13	31	653	47		
	P78386	Keratin type II K85	55766	6,28	6	17	372	28		
	P06702	S100-A9	13234	5,71	1	11	92	3	K.QLSFEEFIMLMAR.L	

46	O43790	Keratin type II K86	53466	5,56	12	31	561	57		
	P78386	Keratin type II K85	55766	6,28	10	27	502	52		
	P06702	S100-A9	13234	5,71	1	11	106	3	K.QLSFEEFIMLMAR.L	
47	P78386	Keratin type II K85	55766	6,28	15	32	782	39		
	O43790	Keratin type II K86	53466	5,56	13	26	523	41		
	P06702	S100-A9	13234	5,71	1	11	104	3	K.QLSFEEFIMLMAR.L	
48	P78386	Keratin type II K85	55766	6,28	10	24	594	57		
	O43790	Keratin type II K86	53466	5,56	12	35	522	62		
	P06702	S100-A9	13234	5,71	1	11	104	3	K.QLSFEEFIMLMAR.L	
49	O43790	Keratin type II K86	53466	5,56	17	42	964	39		
	P78386	Keratin type II K85	55766	6,28	10	24	611	39		
	P06702	S100-A9	13234	5,71	1	11	103	3	K.QLSFEEFIMLMAR.L	
50	O43790	Keratin type II K86	53466	5,56	14	31	914	56		
	P78386	Keratin type II K85	55766	6,28	11	26	637	48		
51	P78386	Keratin type II K85	55766	6,28	10	23	683	31		
	O43790	Keratin type II K86	53466	5,56	10	26	554	36		
52	O43790	Keratin type II K86	53466	5,56	21	49	1594	45		
	P78386	Keratin type II K85	55766	6,28	14	34	1209	39		
53	O43790	Keratin type II K86	53466	5,56	25	59	1630	49		
	Q14533	Keratin type II K81	54936	5,48	22	46	1229	47		
	P78386	Keratin type II K85	55766	6,28	15	33	1111	41		
54	P78386	Keratin type II K85	55766	6,28	6	14	264	52		
	O43790	Keratin type II K86	53466	5,56	6	15	253	42		
55	P78386	Keratin type II K85	55766	6,28	7	17	337	34		R.NFSSCamSAVAPKdmTGNR.C
	O43790	Keratin type II K86	53466	5,56	6	16	297	32		
	Q14533	Keratin type II K81	54936	5,48	5	10	167	39		
	Q15323	Keratin, type I K31	47202	4,87	3	8	138	23		
	Q9BYR9	Keratin-associated protein 2.4	13469	8,32	2	24	121	82		
	P05109	S100-A8	10828	6,51	1	22	104	45	K.ELDINTDGAVNFQEFILILVIK.M	
	Q9BYR8	Keratin-associated protein 3.1	10532	5,99	1	17	89	59	R.SCamSVPTGPATTFCamSFDK.S	

59	O43790	Keratin type II K86	53466	5,56	8	27	462	34		
	P78386	Keratin type II K85	55766	6,28	13	32	436	37		R.NFSSCamSAVAPKdmTGNR.C
	Q14533	Keratin type II K81	54936	5,48	7	21	322	38		
	Q9BYR9	Keratin-associated protein 2.4	13469	8,32	1	11	111	63	R.PITCamCamPSSCamTAVVCamR.P	
	Q8IUC1	Keratin-associated protein 11.1	17074	8,32	1	9	80	59	R.QTTCamISNPCamSTTYSR.P	
	Q9BYR6	Keratin-associated protein 3.3	10357	5,40	1	17	44	48	R.GCamSVPTGPATTICamSSDK.S	
	Q9BYR8	Keratin-associated protein 3.1	10532	5,99	1	17	42	59	R.SCamSVPTGPATTFCamSFDK.S	
60	P78386	Keratin type II K85	55766	6,28	11	25	459	37		R.NFSSCamSAVAPKdmTGNR.C
	O43790	Keratin type II K86	53466	5,56	7	23	419	33		
	Q9BYR9	Keratin-associated protein 2.4	13469	8,32	2	24	108	118		
	P05109	S100-A8	10828	6,51	1	22	90	45	K.ELDINTDGAVNFQEFLLILVIK.M	
	Q92764	Keratin, type I K35	47556	4,75	2	5	55	21		
	Q9BYR8	Keratin-associated protein 3.1	10532	5,99	1	17	36	59	R.SCamSVPTGPATTFCamSFDK.S	
61	P78386	Keratin type II K85	55766	6,28	10	24	412	37		R.NFSSCamSAVAPKdmTGNR.C
	O43790	Keratin type II K86	53466	5,56	8	26	389	52		
	Q14533	Keratin type II K81	54936	5,48	6	16	260	59		
	Q9BYR9	Keratin-associated protein 2.4	13469	8,32	2	24	116	82		
	P05109	S100-A8	10828	6,51	1	22	86	45	K.ELDINTDGAVNFQEFLLILVIK.M	
	Q9BYR6	Keratin-associated protein 3.3	10357	5,40	1	17	76	48	R.GCamSVPTGPATTICamSSDK.S	
	Q8IUC1	Keratin-associated protein 11.1	17074	8,32	1	9	74	59	R.QTTCamISNPCamSTTYSR.P	
	Q9BYQ4	Keratin-associated protein 9.2	18247	8,16	1	10	56	132	R.TTCamFQPTCamVSSCamCamQPSCamCam.-	
	Q9BYP9	Keratin-associated protein 9.9	16253	8,22	1	11	56	132	R.TTCamFQPTCamVSSCamCamQPSCamCam.-	
Q15323	Keratin, type I K31	47202	4,87	1	3	53	21	R.LNVEVDAAPTVDLNR.V		
62										K.KtmYEEEVAlR.A
	P78386	Keratin type II K85	55766	6,28	12	24	466	41		R.NFSSCamSAVAPKdmTGNR.C
										R.NFSSCamSAVAPKtmTGNR.C
	O43790	Keratin type II K86	53466	5,56	8	19	315	64		
	P05109	S100-A8	10828	6,51	1	22	116	45	K.ELDINTDGAVNFQEFLLILVIK.M	
	Q15323	Keratin, type I K31	47202	4,87	1	2	61	5	R.DNAELENLIR.E	
	O76009	Keratin, type I K33a	45906	4,82	1	2	61	5	R.DNAELENLIR.E	

	Q14525	Keratin, type I K33b	46185	4,81	1	2	61	5	R.DNAELENLIR.E	
	Q92764	Keratin, type I K35	47556	4,75	1	2	50	26	R.QLVESDINGLR.R	
63	P78386	Keratin type II K85	55766	6,28	7	15	324	39		
	O43790	Keratin type II K86	53466	5,56	4	15	278	32		
	Q15323	Keratin, type I K31	47202	4,87	3	8	116	30		
	P05109	S100-A8	10828	6,51	1	22	96	45	K.ELDINTDGAVNFQEFLILVIK.M	
	Q9BYR9	Keratin-associated protein 2.4	13469	8,32	1	11	74	63	R.PITCamCamPSSCamTAVVCamR.P	
	Q9BYR6	Keratin-associated protein 3.3	10357	5,40	1	17	69	48	R.GCamSVPTGPATTICamSSDK.S	
	Q9BYR7	Keratin-associated protein 3.2	10400	5,40	1	17	47	49	R.SCamSVPTGPATTICamSSDK.S	
64	P78386	Keratin type II K85	55766	6,28	14	30	606	32		
	O43790	Keratin type II K86	53466	5,56	7	23	392	34		
	Q14533	Keratin type II K81	54936	5,48	6	17	273	37		
	Q15323	Keratin, type I K31	47202	4,87	2	7	115	30		
	P05109	S100-A8	10828	6,51	1	22	87	45	K.ELDINTDGAVNFQEFLILVIK.M	
	Q92764	Keratin, type I K35	47556	4,75	1	2	51	26	R.QLVESDINGLR.R	
	Q9BYR9	Keratin-associated protein 2.4	13469	8,32	1	11	49	63	R.PITCamCamPSSCamTAVVCamR.P	
Q9BYR6	Keratin-associated protein 3.3	10357	5,40	1	17	41	48	R.GCamSVPTGPATTICamSSDK.S		
65	P78386	Keratin type II K85	55766	6,28	3	10	147	35		
	P06702	S100-A9	13234	5,71	3	37	143	18		
	Q14533	Keratin type II K81	54936	5,48	2	11	108	29		
	P78385	Keratin type II K83	54179	5,34	2	11	108	29		
	O43790	Keratin type II K86	53466	5,56	2	11	108	29		
	P05109	S100-A8	10828	6,51	2	23	77	25		
66	O43790	Keratin type II K86	53466	5,56	5	20	275	32		
	P78386	Keratin type II K85	55766	6,28	8	19	262	43		
	Q14533	Keratin type II K81	54936	5,48	4	14	239	35		
	Q15323	Keratin, type I K31	47202	4,87	3	7	157	27		
	Q9BYR8	Keratin-associated protein 3.1	10532	5,99	1	17	56	101	R.SCamSVPTGPATTFCamSFDK.S	
	Q9BYR9	Keratin-associated protein 2.4	13469	8,32	2	25	51	69		
Q9BYR6	Keratin-associated protein 3.3	10357	5,40	1	17	41	107	R.GCamSVPTGPATTICamSSDK.S		
67	P78386	Keratin type II K85	55766	6,28	12	30	502	36		
	O43790	Keratin type II K86	53466	5,56	9	31	441	29		

	Q9BYR6	Keratin-associated protein 3.3	10357	5,40	1	17	86	48	R.GCamSVPTGPATTICamSSDK.S	
	P05109	S100-A8	10828	6,51	1	22	83	45	K.ELDINTDGAVNFQEFLILVIK.M	
68	P78386	Keratin type II K85	55766	6,28	7	18	476	30		
	O43790	Keratin type II K86	53466	5,56	3	12	188	29		
	Q9BYR9	Keratin-associated protein 2.4	13469	8,32	1	11	93	63	R.PITCamCamPSSCamTAVVCamR.P	
69	P78386	Keratin type II K85	55766	6,28	13	27	569	34		R.NFSSCamSAVAPKdmTGNR.C
	O43790	Keratin type II K86	53466	5,56	10	30	525	29		
	Q14533	Keratin type II K81	54936	5,48	8	23	357	33		
	Q15323	Keratin, type I K31	47202	4,87	2	5	128	33		
	Q9BYR9	Keratin-associated protein 2.4	13469	8,32	2	24	109	82		
	Q9BYQ4	Keratin-associated protein 9.2	18247	8,16	1	10	96	132	R.TTCamFQPTCamVSSCamCamQPSCamCam.-	
	Q9BYP9	Keratin-associated protein 9.9	16253	8,22	1	11	96	132	R.TTCamFQPTCamVSSCamCamQPSCamCam.-	
	Q9BYR8	Keratin-associated protein 3.1	10532	5,99	1	17	82	59	R.SCamSVPTGPATTFCamSFDK.S	
	Q9BYR6	Keratin-associated protein 3.3	10357	5,40	1	17	59	48	R.GCamSVPTGPATTICamSSDK.S	
70	P78386	Keratin type II K85	55766	6,28	11	25	649	49		R.NFSSCamSAVAPKdmTGNR.C
	O43790	Keratin type II K86	53466	5,56	7	21	347	32		
	Q14533	Keratin type II K81	54936	5,48	6	16	268	36		
	Q15323	Keratin, type I K31	47202	4,87	3	9	139	25		
	P05109	S100-A8	10828	6,51	1	22	78	45	K.ELDINTDGAVNFQEFLILVIK.M	
	Q9BYQ4	Keratin-associated protein 9.2	18247	8,16	1	10	73	132	R.TTCamFQPTCamVSSCamCamQPSCamCam.-	
	Q9BYP9	Keratin-associated protein 9.9	16253	8,22	1	11	73	132	R.TTCamFQPTCamVSSCamCamQPSCamCam.-	
	Q9BYR8	Keratin-associated protein 3.1	10532	5,99	1	17	56	59	R.SCamSVPTGPATTFCamSFDK.S	
	Q9BYR6	Keratin-associated protein 3.3	10357	5,40	1	17	54	48	R.GCamSVPTGPATTICamSSDK.S	
71	P78386	Keratin type II K85	55766	6,28	11	25	631	30		R.NFSSCamSAVAPKdmTGNR.C R.NFSSCamSAVAPKtmTGNR.C
	O43790	Keratin type II K86	53466	5,56	8	26	456	33		
	Q14533	Keratin type II K81	54936	5,48	7	19	387	36		
	Q9BYR9	Keratin-associated protein 2.4	13469	8,32	1	11	107	63	R.PITCamCamPSSCamTAVVCamR.P	
	Q15323	Keratin, type I K31	47202	4,87	1	3	91	35	R.DSLENTL TESEAR.Y	
	Q14532	Keratin, type I K32	50286	4,81	1	2	91	35	R.DSLENTL TESEAR.Y	
	O76009	Keratin, type I K33a	45906	4,82	1	3	91	35	R.DSLENTL TESEAR.Y	
	Q9BYR6	Keratin-associated protein 3.3	10357	5,40	1	17	63	48	R.GCamSVPTGPATTICamSSDK.S	

72	P78386	Keratin type II K85	55766	6,28	13	27	735	30		R.NFSSCamSAVAPKdmTGNR.C
	O43790	Keratin type II K86	53466	5,56	10	30	521	33		
	Q14533	Keratin type II K81	54936	5,48	9	23	366	36		
	Q9BYR9	Keratin-associated protein 2.4	13469	8,32	1	11	115	63	R.PITCamCamPSSCamTAVVCamR.P	
	P05109	S100-A8	10828	6,51	1	22	80	45	K.ELDINTDGAVNFQEFLLILVIK.M	
	Q15323	Keratin, type I K31	47202	4,87	1	3	77	35	R.DSLENTLSEAR.Y	
	Q14532	Keratin, type I K32	50286	4,81	1	2	77	35	R.DSLENTLSEAR.Y	
	O76009	Keratin, type I K33a	45906	4,82	1	3	77	35	R.DSLENTLSEAR.Y	
	Q9BYR6	Keratin-associated protein 3.3	10357	5,40	1	17	68	48	R.GCamSVPTGPATTICamSSDK.S	
73	P78386	Keratin type II K85	55766	6,28	12	26	671	31		R.NFSSCamSAVAPKdmTGNR.C K.LLETKtmWQFYQNR.C
	O43790	Keratin type II K86	53466	5,56	5	20	401	31		
	Q14533	Keratin type II K81	54936	5,48	3	10	223	39		
	P78385	Keratin type II K83	54179	5,34	3	10	223	39		
	Q9BYR9	Keratin-associated protein 2.4	13469	8,32	1	11	104	63	R.PITCamCamPSSCamTAVVCamR.P	
	Q9BYR6	Keratin-associated protein 3.3	10357	5,40	1	17	44	48	R.GCamSVPTGPATTICamSSDK.S	
74	P78386	Keratin type II K85	55766	6,28	8	20	393	28		
	O43790	Keratin type II K86	53466	5,56	5	17	346	23		
	Q15323	Keratin, type I K31	47202	4,87	1	3	66	35	R.DSLENTLSEAR.Y	
	Q14532	Keratin, type I K32	50286	4,81	1	2	66	35	R.DSLENTLSEAR.Y	
	O76009	Keratin, type I K33a	45906	4,82	1	3	66	35	R.DSLENTLSEAR.Y	
	Q9BYR6	Keratin-associated protein 3.3	10357	5,40	1	17	64	48	R.GCamSVPTGPATTICamSSDK.S	
75	Q14533	Keratin type II K81	54936	5,48	13	32	656	32		
	P78386	Keratin type II K85	55766	6,28	12	25	592	30		
	Q15323	Keratin, type I K31	47202	4,87	5	18	257	24		
	Q9BYR9	Keratin-associated protein 2.4	13469	8,32	2	25	76	45		
	Q9BYR6	Keratin-associated protein 3.3	10357	5,40	1	17	76	48	R.GCamSVPTGPATTICamSSDK.S	
	Q92764	Keratin, type I K35	47556	4,75	1	2	57	26	R.QLVESDINGLR.R	
	Q9BYR7	Keratin-associated protein 3.2	10400	5,40	1	17	46	49	R.SCamSVPTGPATTICamSSDK.S	
	Q9BYR8	Keratin-associated protein 3.1	10532	5,99	1	17	33	59	R.SCamSVPTGPATTICamSFDK.S	
76	O43790	Keratin type II K86	53466	5,56	15	39	757	36		R.KtmSDLEANVEALIQEIDFLR.R
	Q14533	Keratin type II K81	54936	5,48	14	32	753	38		R.KtmSDLEANVEALIQEIDFLR.R

	P78386	Keratin type II K85	55766	6,28	11	26	661	28		R.NFSSCamSAVAPKdmTGNR.C
										R.NFSSCamSAVAPKtmTGNR.C
	P78385	Keratin type II K83	54179	5,34	10	23	610	41		R.KtmSDLEANVEALIQEIDFLR.R
	Q15323	Keratin, type I K31	47202	4,87	3	7	146	27		
	Q9BYR9	Keratin-associated protein 2.4	13469	8,32	2	25	124	45		
	Q92764	Keratin, type I K35	47556	4,75	1	2	82	26	R.QLVESDINGLR.R	
	Q9BYR8	Keratin-associated protein 3.1	10532	5,99	1	17	65	59	R.SCamSVPTGPATTFCamSFDK.S	
Q9BYR6	Keratin-associated protein 3.3	10357	5,40	1	17	61	48	R.GCamSVPTGPATTICamSSDK.S		
77	P78386	Keratin type II K85	55766	6,28	13	29	706	41		
	O43790	Keratin type II K86	53466	5,56	10	36	595	33		
	Q14533	Keratin type II K81	54936	5,48	9	29	539	37		
	Q15323	Keratin, type I K31	47202	4,87	4	12	191	28		
	Q9BYR9	Keratin-associated protein 2.4	13469	8,32	1	11	110	63	R.PITCamCamPSSCamTAVVCamR.P	
	Q9BYR8	Keratin-associated protein 3.1	10532	5,99	1	17	76	59	R.SCamSVPTGPATTFCamSFDK.S	
	Q9BYR6	Keratin-associated protein 3.3	10357	5,40	1	17	69	48	R.GCamSVPTGPATTICamSSDK.S	
	P05109	S100-A8	10828	6,51	1	22	56	45	K.ELDINTDGAVNFQEFLILVIK.M	
	Q92764	Keratin, type I K35	47556	4,75	1	2	52	26	R.QLVESDINGLR.R	
	Q9BYR2	Keratin-associated protein 4.5	19900	8,24	1	10	45	51	R.VSCamHTTCamYRPTCamVISTCamPR.P	
	Q9BYQ5	Keratin-associated protein 4.6	21807	8,43	1	9	45	51	R.VSCamHTTCamYRPTCamVISTCamPR.P	
	Q9BYR0	Keratin-associated protein 4.7	22472	8,35	1	9	45	51	R.VSCamHTTCamYRPTCamVISTCamPR.P	
	Q9BYQ9	Keratin-associated protein 4.8	20704	8,23	1	9	45	51	R.VSCamHTTCamYRPTCamVISTCamPR.P	
Q9BQ66	Keratin-associated protein 4.12	21390	8,37	1	9	45	51	R.VSCamHTTCamYRPTCamVISTCamPR.P		
78	P78386	Keratin type II K85	55766	6,28	15	32	806	36		R.NFSSCamSAVAPKdmTGNR.C
	O43790	Keratin type II K86	53466	5,56	11	36	692	31		
	Q14533	Keratin type II K81	54936	5,48	10	28	556	34		
	Q14525	Keratin, type I K33b	46185	4,81	3	12	125	20		
	Q15323	Keratin, type I K31	47202	4,87	3	12	125	20		
	Q9BYR9	Keratin-associated protein 2.4	13469	8,32	2	25	109	45		
	P05109	S100-A8	10828	6,51	1	22	80	45	K.ELDINTDGAVNFQEFLILVIK.M	
	Q9BYR8	Keratin-associated protein 3.1	10532	5,99	1	17	77	59	R.SCamSVPTGPATTFCamSFDK.S	
	Q9BYR6	Keratin-associated protein 3.3	10357	5,40	1	17	63	48	R.GCamSVPTGPATTICamSSDK.S	
	Q92764	Keratin, type I K35	47556	4,75	1	2	50	26	R.QLVESDINGLR.R	

	Q9BYR7	Keratin-associated protein 3.2	10400	5,40	1	17	49	49	R.SCamSVPTGPATTICamSSDK.S	
79	O43790	Keratin type II K86	53466	5,56	12	31	750	35		
	P78386	Keratin type II K85	55766	6,28	13	28	672	41		
	Q14533	Keratin type II K81	54936	5,48	11	25	627	32		
	Q15323	Keratin, type I K31	47202	4,87	4	14	161	20		
	Q9BYR9	Keratin-associated protein 2.4	13469	8,32	3	37	127	67		
	Q9BYR8	Keratin-associated protein 3.1	10532	5,99	1	17	89	59	R.SCamSVPTGPATTFCamSFDK.S	
	Q9BYR6	Keratin-associated protein 3.3	10357	5,40	1	17	65	48	R.GCamSVPTGPATTICamSSDK.S	
	Q92764	Keratin, type I K35	47556	4,75	1	2	48	26	R.QLVESDINGLR.R	
80	O43790	Keratin type II K86	53466	5,56	14	43	854	36		R.KtmSDLEANVEALIQEIDFLR.R
	P78386	Keratin type II K85	55766	6,28	14	31	776	33		R.NFSSCamSAVAPKdmTGNR.C
	Q14533	Keratin type II K81	54936	5,48	13	36	690	35		R.KtmSDLEANVEALIQEIDFLR.R
	P78385	Keratin type II K83	54179	5,34	10	28	590	37		R.KtmSDLEANVEALIQEIDFLR.R
	Q15323	Keratin, type I K31	47202	4,87	6	20	281	25		
	O76011	Keratin, type I K34	44661	4,74	3	11	143	26		
	Q9BYR9	Keratin-associated protein 2.4	13469	8,32	2	24	124	82		
	Q9BYR7	Keratin-associated protein 3.2	10400	5,40	1	17	82	49	R.SCamSVPTGPATTICamSSDK.S	
	Q9BYR8	Keratin-associated protein 3.1	10532	5,99	1	17	79	59	R.SCamSVPTGPATTFCamSFDK.S	
	Q9BYR6	Keratin-associated protein 3.3	10357	5,40	1	17	74	48	R.GCamSVPTGPATTICamSSDK.S	
	Q92764	Keratin, type I K35	47556	4,75	1	2	62	26	R.QLVESDINGLR.R	
81	P78386	Keratin type II K85	55766	6,28	10	26	506	48		
	O43790	Keratin type II K86	53466	5,56	11	32	456	36		
	Q14533	Keratin type II K81	54936	5,48	10	26	429	35		
	P78385	Keratin type II K83	54179	5,34	6	18	235	33		
	Q9BYR9	Keratin-associated protein 2.4	13469	8,32	2	24	119	88		
	Q9BYR8	Keratin-associated protein 3.1	10532	5,99	1	17	88	59	R.SCamSVPTGPATTFCamSFDK.S	
	Q9BYR6	Keratin-associated protein 3.3	10357	5,40	1	17	73	48	R.GCamSVPTGPATTICamSSDK.S	
	Q15323	Keratin, type I K31	47202	4,87	1	3	67	35	R.DSLENTL TESEAR.Y	
	Q14532	Keratin, type I K32	50286	4,81	1	2	67	35	R.DSLENTL TESEAR.Y	
	O76009	Keratin, type I K33a	45906	4,82	1	3	67	35	R.DSLENTL TESEAR.Y	
	Q9BYR2	Keratin-associated protein 4.5	19900	8,24	1	10	37	51	R.VSCamHTTCamYRPTCamVISTCamPR.P	
	Q9BYQ5	Keratin-associated protein 4.6	21807	8,43	1	9	37	51	R.VSCamHTTCamYRPTCamVISTCamPR.P	

	Q9BYR0	Keratin-associated protein 4.7	22472	8,35	1	9	37	51	R.VSCamHTTCamYRPTCamVISTCamPR.P	
	Q9BYQ9	Keratin-associated protein 4.8	20704	8,23	1	9	37	51	R.VSCamHTTCamYRPTCamVISTCamPR.P	
	Q9BQ66	Keratin-associated protein 4.12	21390	8,37	1	9	37	51	R.VSCamHTTCamYRPTCamVISTCamPR.P	
82	P78386	Keratin type II K85	55766	6,28	10	26	609	33		
	O43790	Keratin type II K86	53466	5,56	7	26	448	23		
	Q14533	Keratin type II K81	54936	5,48	6	19	322	30		
	P78385	Keratin type II K83	54179	5,34	5	17	234	34		
	Q15323	Keratin, type I K31	47202	4,87	3	11	208	24		
	Q9BYR9	Keratin-associated protein 2.4	13469	8,32	1	11	168	63	R.PITCamCamPSSCamTAVVCamR.P	
	Q9BYR8	Keratin-associated protein 3.1	10532	5,99	1	17	99	59	R.SCamSVPTGPATTFCamSFDK.S	
Q9BYR6	Keratin-associated protein 3.3	10357	5,40	1	17	60	48	R.GCamSVPTGPATTICamSSDK.S		
83	P78386	Keratin type II K85	55766	6,28	11	26	555	38		R.LTAEIENAKdmCamQR.A
	O43790	Keratin type II K86	53466	5,56	8	25	407	34		
	Q14533	Keratin type II K81	54936	5,48	7	20	403	34		
	Q15323	Keratin, type I K31	47202	4,87	4	15	170	24		
	Q9BYR9	Keratin-associated protein 2.4	13469	8,32	1	11	119	63	R.PITCamCPSSCamTAVVCR.P	
	Q9BYR6	Keratin-associated protein 3.3	10357	5,40	1	17	70	48	R.GCamSVPTGPATTICamSSDK.S	
	Q9BYR7	Keratin-associated protein 3.2	10400	5,40	1	17	49	49	R.SCamSVPTGPATTICamSSDK.S	
84	P78386	Keratin type II K85	55766	6,28	12	28	776	38		R.LTAEIENAKdmCamQR.A
	O43790	Keratin type II K86	53466	5,56	14	38	684	36		
	Q14533	Keratin type II K81	54936	5,48	14	36	636	57		
	Q15323	Keratin, type I K31	47202	4,87	3	10	148	27		
	Q9BYR9	Keratin-associated protein 2.4	13469	8,32	2	25	117	45		
	Q9BYR8	Keratin-associated protein 3.1	10532	5,99	1	17	89	59	R.SCamSVPTGPATTFCamSFDK.S	
	Q9BYR6	Keratin-associated protein 3.3	10357	5,40	1	17	80	48	R.GCamSVPTGPATTICamSSDK.S	
	Q92764	Keratin, type I K35	47556	4,75	1	2	62	26	R.QLVESDINGLR.R	
	Q9BYR2	Keratin-associated protein 4.5	19900	8,24	1	10	39	51	R.VSCamHTTCamYRPTCamVISTCamPR.P	
	Q9BYQ5	Keratin-associated protein 4.6	21807	8,43	1	9	39	51	R.VSCamHTTCamYRPTCamVISTCamPR.P	
	Q9BYR0	Keratin-associated protein 4.7	22472	8,35	1	9	39	51	R.VSCamHTTCamYRPTCamVISTCamPR.P	
	Q9BYQ9	Keratin-associated protein 4.8	20704	8,23	1	9	39	51	R.VSCamHTTCamYRPTCamVISTCamPR.P	
Q9BQ66	Keratin-associated protein 4.12	21390	8,37	1	9	39	51	R.VSCamHTTCamYRPTCamVISTCamPR.P		
85	P78386	Keratin type II K85	55766	6,28	15	32	749	36		R.LTAEIENAKdmCamQR.A

									R.FLEQQN <u>Kdm</u> LLETK.W
O43790	Keratin type II K86	53466	5,56	12	35	694	36		R.FLEQQN <u>Kdm</u> LLETK.L
Q14533	Keratin type II K81	54936	5,48	11	28	677	37		R.FLEQQN <u>Kdm</u> LLETK.L
Q15323	Keratin, type I K31	47202	4,87	5	18	279	27		
Q9BYR9	Keratin-associated protein 2.4	13469	8,32	2	24	121	82		
Q9BYR8	Keratin-associated protein 3.1	10532	5,99	1	17	83	59	R.S <u>Cam</u> SVPTGPATTF <u>Cam</u> SFDK.S	
Q9BYR6	Keratin-associated protein 3.3	10357	5,40	1	17	81	48	R.G <u>Cam</u> SVPTGPATTI <u>Cam</u> SSDK.S	
Q92764	Keratin, type I K35	47556	4,75	1	2	67	26	R.QLVESDINGLR.R	

ANNEXE 4: Publication

Transglutaminase-3 Enzyme: A Putative Actor in Human Hair Shaft Scaffolding ?

Sébastien Thibaut, Nükhet Cavüsoglü, Emmanuelle De Becker, Franck Zerbib, Audrey Bednarczyk, Christine Schaeffer, Alain Van Dorselaer and Bruno A. Bernard , *Journal of Investigative Dermatology*, 2008.

Transglutaminase-3 enzyme: a putative actor in human hair shaft scaffolding?

Sébastien THIBAUT, Nükhet CAVUSOGLU, Emmanuelle DE BECKER, Franck ZERBIB,
Audrey BEDNARCZYK, Christine SCHAEFFER, Alain VAN DORSSELAER and Bruno A.
BERNARD

Journal of investigative dermatology, 2009, vol. 129, pages 449–459

Annexe 4 : pages 200-... :

La publication présentée ici dans la thèse est soumise à des droits détenus par un éditeur commercial.

Les utilisateurs de l'ULP peuvent consulter cette publication sur le site de l'éditeur :

<http://dx.doi.org/10.1038/jid.2008.231>

La version imprimée de cette thèse peut être consultée à la bibliothèque ou dans un autre établissement via une demande de prêt entre bibliothèques (PEB) auprès de nos services :

<http://www-sicd.u-strasbg.fr/services/peb/>

CONCLUSION GENERALE

CONCLUSION GENERALE

Aujourd'hui, on estime que pour un génome constitué d'environ 30 000 gènes, plus d'un million de protéines sont exprimées suite aux différentes étapes de transcription, maturation et traductions. Ceci reflète la forte diversité moléculaire que l'on peut rencontrer dans la structure des protéines. Et c'est cette diversité moléculaire qui va entraîner une variation dans le rôle biologique des protéines. C'est pourquoi, l'identification des protéines dans un milieu donné n'est plus suffisante et notre objectif a été de permettre une caractérisation plus complète de ces protéines pour donner des outils de compréhension de toute la finesse des régulations des mécanismes biologiques.

Durant mon travail de thèse, j'ai développé de **nouvelles méthodologies pour une caractérisation des plus détaillée des protéines** présentes dans divers extraits biologiques incluant **l'identification de différentes modifications post-traductionnelles (PTMs)**.

La N-glycosylation est une des PTMs les plus répandue chez les protéines et de nombreuses études ont montré ces dernières années que selon la structure de la N-glycosylation celle-ci avait un impact très fort sur la fonction des protéines dans un milieu donné. L'étude de l'hétérogénéité des N-glycoprotéines est donc aujourd'hui devenue primordiale.

Dans ce contexte, une première étude a été réalisée pour **la caractérisation fine des glycoformes d'une hormone de fécondité, la « Follicle Stimulating Hormone » (FSH)**. La FSH est une N-glycoprotéine dont le taux et la distribution des glycoformes varient durant le cycle menstruel de la femme. Des traitements à base d'injection de r-hFSH ont été élaborés et sont aujourd'hui utilisés pour combattre l'infertilité. Néanmoins ceux-ci ne prennent pas en compte la distribution des glycoformes.

Or pour présenter des traitements efficaces, il est nécessaire de connaître leur distribution fine en glycoformes. Cette étude avait donc pour objectif de décrire avec précision la **micro-hétérogénéité au niveau des deux sites** de N-glycosylations (**Asn⁵² et Asn⁷⁸**) ainsi que **l'hétérogénéité globale** des glycoformes de la **sous-unité alpha de la FSH**. Pour cela, nous avons développé des **stratégies complémentaires**:

- **L'analyse de la protéine entière** par ESI-MS afin d'étudier son hétérogénéité globale.
- **L'analyse par LC-MS des glycopeptides** issus de la digestion de la protéine afin d'étudier la micro-hétérogénéité sur chacun des deux sites de N-glycosylation de la FSH.

Ces deux méthodes complémentaires de caractérisation et de quantification relative des glycoformes de la FSH nous ont permis **d'identifier l'ensemble des N-glycannes sur chaque site de glycosylation** mais également de déterminer les **glycoformes majoritaires de la protéine**. Il nous a été possible de **montrer des variations** plus ou moins importantes dans la **composition** et **l'abondance** des différentes **glycoformes**.

Ultérieurement, ces deux méthodes complémentaires seront utilisées pour déterminer la nature et la distribution des glycoformes de lots de FSH recombinante purifiées selon différents protocoles et de pouvoir enfin **corrélér les variations de structures** avec les variations **d'activités biologiques durant le cycle menstruel**. Ainsi, il sera possible de **définir à quel moment** ces **fractions** doivent être **injectées durant le cycle menstruel** de la femme pour **aboutir à une nidation** ou **éviter toutes nidations supplémentaires** tout en maintenant une première nidation. Ceci permettrait **d'éviter des grossesses multiples** qui sont trop souvent observées lors de ces traitements hormonaux.

Nous avons donc apporté un outil analytique qui permet de mieux comprendre les mécanismes d'action de ces hormones.

Trois études supplémentaires ont également été réalisées pour **la caractérisation fine de N-glycoprotéines qui nous ont conduit à** développer de nouvelles stratégies basées sur l'analyse **nanoLC-Chip-MS/MS de mélanges complexes de peptides et de N-glycopeptides** en utilisant des **phases stationnaires de carbone graphite en complément de la phase inverse C18**. Nous avons appliqué cette nouvelle approche dans trois études qui nous ont permis de :

- Déterminer la **structure** des **4 N-glycannes majoritaires** et de **8 N-glycannes minoritaires** présents sur le site de N-glycosylation **Asn²⁹⁷** de l'**anticorps monoclonal A2CHM** mais également d'étudier leur **abondance relative**.
- Identifier pour la première fois **le site de N-glycosylation d'une protéine** de *Toxoplasma Gondii*, **l'Asn¹³⁶ de la GAP50** à partir de l'analyse d'extraits peptidiques issus de **bandes de gels 1D**.
- Identifier les **différentes formes de la protéine N-glycosylée Ole e1 (30 acides aminés mutés)** issue de différentes variétés de pollen d'olivier, de déterminer la structure des **14 N-glycannes** présents **sur ces différentes formes** et la **nature des sites de N-glycosylations occupés** à partir de l'analyse d'extraits peptidiques issus de spots de **gels 2D**.

Ces études illustrent l'importance et le rôle de la N-glycosylation dans la fonction biologique de ces différentes protéines et ouvrent de nouvelles perspectives dans le domaine thérapeutique. Une meilleure compréhension de la structure fine de ces glycoprotéines permettra à long terme d'améliorer ou encore trouver de nouveaux traitements pour combattre les différentes maladies tels que les cancers ou les allergies.

Enfin, durant ce travail de thèse j'ai également développé des méthodologies pour **la caractérisation fine des protéines du cheveu humain**. A l'heure actuelle même si la nature des protéines présentes dans le cheveu est connue, peu d'informations sont disponibles concernant leurs **modes d'interaction** et la **manière dont elles évoluent** au cours de la maturation du cheveu notamment en terme de modifications chimiques ou enzymatiques qu'elles subissent.

Les protéines du cheveu ont des **propriétés physico-chimiques peu compatibles** avec les approches classiquement utilisées en **analyse protéomique**. De ce fait, leur étude a nécessité la mise au point de méthodes d'extraction et de solubilisation adaptées aux techniques de séparation compatibles avec l'analyse par spectrométrie de masse. D'autre part, ces protéines présentent de **fortes homologies de séquences** qui rendent leur étude plus complexe par spectrométrie de masse.

Pour aller vers **une caractérisation fine de ces protéines**, il était donc nécessaire de combiner différentes approches. Ainsi, nous avons utilisé la stratégie de **séparation des protéines sur gels 2D** et la stratégie de **séparation par LC 2D de mélanges complexes de peptides issus de la digestion directe des extraits protéiques** :

L'approche « **gels 2D** » nous a permis de mettre en évidence la forte **hétérogénéité dans la structure des Kératines des Filaments Intermédiaires (KIFs)** (dégradations) et des nouvelles **liaisons inter- et intramoléculaires de type N^ε-(γ-glutamyl) lysine entre KIFs et les Protéines Associées aux Kératines (KAPs)**.

L'approche « **LC 2D** » nous a permis d'expliquer cette forte hétérogénéité observée sur le gel 2D en **identifiant des PTMs de type mono-, di-, tri-méthylation et tri-oxydation sur les KIFs**. De plus, des expériences réalisées sur différents extraits a permis de **suggérer globalement la localisation des KIFs dans le cheveu à maturation**.

Ces deux approches complémentaires nous ont permis de proposer **un nouveau modèle structural des filaments intermédiaires** qui met en jeu des interactions de type **ponts disulfures** et de type **N^ε-(γ-glutamyl) lysine entre KIFs et KAPs**. Nous avons pu ainsi apporter de nouvelles informations primordiales pour une **meilleure compréhension de cette machinerie complexe** qu'est le cheveu en termes de diversité structurale des **protéines identifiées** dans un cheveu à maturation, des **dégradations** et des **modifications** qu'elles subissent, de leurs **interactions** et de leur **localisation** dans le cheveu (cortex ou cuticule).

Ces deux méthodes complémentaires pourront être appliquées à des lots de cheveux soumis à des traitements chimiques afin d'étudier les dégradations induites par ces derniers. De même, il s'agira d'étudier plus en détail ce nouveau modèle structural des filaments intermédiaire en mettant en évidence d'une part la position exacte des différentes liaisons inter- et intramoléculaire et d'autre part de définir le rôle exacte des différentes PTMs identifiées sur les KIFs dans le cheveu.

Notre travail a donc permis d'avancer dans la compréhension de l'organisation moléculaire de la structure du cheveu.

En **conclusion**, au cours de ce travail de thèse, j'ai été amenée à **développer de nouvelles méthodologies pour une caractérisation structurale plus détaillée des protéines** présentes dans divers extraits biologiques incluant **l'identification de différentes PTMs** et plus particulièrement les **N-glycosylations** des asparagines, les **di- et tri-méthylations** des lysines, les **tri-oxydations** des cystéines et les **liaisons inter-chaînes peptidiques**. Ces optimisations méthodologiques nous ont permis de caractériser des **protéines** présentes en **faible quantité** dans des extraits protéiques complexes (de l'ordre du **subpicomolaire**).

Ces travaux, qui offrent de nouvelles perspectives dans le domaine des N-glycoprotéines à visée thérapeutiques mais également dans le domaine de la cosmétique, ont montré à quel point **selon les problématiques abordées, le choix de la stratégie analytique** utilisée est **primordial** et qu'il est souvent indispensable de **combiner plusieurs approches** pour obtenir une **image aussi fidèle que possible du contenu protéique** d'échantillons biologiques.

Notre travail a donc montré que l'étude des mécanismes biologiques passait nécessairement par une détermination totale, jusque dans les moindres détails, de la structure des biomolécules.

