

# THESE

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

par

**Agnès HOVASSE**

**Protéomique : maîtrise de l'instrumentation,  
applications et étude des glycosylations**

Soutenue le 29 septembre 2010 devant la commission d'examen :

Dr. Alain VAN DORSSELAER

Dr. Bernard MONSARRAT

Dr. Nicolas SOMMERER

Prof. Jacques HAIECH

Directeur de thèse

Rapporteur externe

Rapporteur externe

Examineur interne

*"Le savant n'est pas l'homme qui fournit les vraies réponses; c'est celui qui pose les vraies questions."*

Claude Lévi-Strauss

# Remerciements

Ce travail de thèse a été réalisé au sein du Laboratoire de Spectrométrie de Masse Bio-Organique de l'Institut Pluridisciplinaire Hubert Curien de Strasbourg (UMR 7178, CNRS-UDS)

Je tiens tout d'abord à remercier très sincèrement Alain Van Dorsselaer pour m'avoir permis d'effectuer cette thèse dans des conditions exceptionnelles, de m'avoir encouragé et fait confiance tout au long de ce périple, mais également Christine Schaeffer-Reiss pour son encadrement quotidien, son écoute et sa grande disponibilité.

Je remercie M. Jacques Haiech d'avoir bien voulu présider ce jury de thèse, ainsi que Messieurs Bernard Monsarrat et Nicolas Sommerer pour avoir accepté d'évaluer mon travail.

Je remercie également l'ensemble des personnes avec qui j'ai eu l'occasion de collaborer : Stan Tomavo, Emmanuel Bissé, Sandrine Jégou, Clara Cilindre et Andrea Mosca.

Un grand merci à tous ceux avec qui j'ai travaillé, discuté, rigolé... au sein du LSMBO :

Avec pour commencer, Sarah Cianferani-Sanglier (ma première manucure!), Véronique Trimbou (bonne chance pour l'avenir), Véronique Delval, Hélène Diemer et Jennifer Jung (merci ces moments de complicité entre filles) , l'indispensable Fabrice Varrier (l'ambiance ne serait vraiment pas la même sans toi, et nos ordi non plus d'ailleurs), la tout aussi indispensable Danièle Thierse, Fabrice Bertile (les américains pendant ce temps là...), Cyril Colas (souvent de bon conseil), Laurence Sabatier (un grand merci pour tes coups de pouces décisifs), Christine Carapito (j'ai beaucoup appris grâce à toi), l'indescriptible François Delalande, Alexandre Burel (la bio-info nous a changé la vie), Dominique Roecklin (pour ta grande contribution à l'aventure hémoglobine), Kévin Jeanpert, Jean-Michel Saliou et bien sur mon colloc' Jean-Marc Strub (merci pour ton support technique sans faille mais aussi pour toutes nos discussion "grutesques" et autres).

Merci aux anciens... qui sont ne le sont pas tant que ça, Sébastien (toutes ces heures à cogiter derrière les machines), Cédric (le président du club des éclopés), Guillaume (pour les pauses salon de thé), Laetitia (pour nos nombreuses séances de craquage et de remotivassions), mais aussi Audrey, Laurent, Dimitri et Chrystel.

Merci à ceux qui sont en train de finir, Daniel (c'est pas bien de fumer), Nicolas (dit le parisien), Thierry (tu verras, les trappes on finit par s'y attacher... après quelques années) et Bertrand; c'est la dernière ligne droite ce n'est pas le moment de craquer; et merci à la relève, Sarah, Magalie, Stéphanie, Elie et Amandine

Enfin un grand merci à tous ceux qui m'ont permis d'atteindre et de franchir cette étape.

Merci à Marion, à Benjamin et à mes parents pour m'avoir toujours soutenu et encouragé malgré mon parcours chaotique. Et pour finir, Richard, un simple merci n'est de loin pas suffisant pour exprimer tout ce que je te dois.

<b>INTRODUCTION GENERALE</b> .....	1
<b><u>INTRODUCTION BIBLIOGRAPHIQUE</u></b> .....	7
<b>Chapitre 1</b> : L'analyse des protéines N-glycosylées par spectrométrie de masse.....	9
<b>Chapitre 2</b> : Outils et stratégies pour l'analyse protéomique : application à l'analyse des glycoprotéines.....	29
<b><u>RESULTATS</u></b> .....	67
<b>PARTIE 1 : Développements</b>	
<b>Recherche de l'amélioration des méthodologies d'analyses LC-MS/MS pour l'identification et la caractérisation des protéines et des glycoprotéines</b> .....	69
<b>Chapitre 1</b> : Développements de nouvelles méthodes pour l'évaluation des performances d'un couplage nanoLC-MS/MS.....	71
<b>Chapitre 2</b> : Développements méthodologiques pour la détection et la caractérisation des protéines N-glycosylées.....	101
<b>PARTIE 2 : Applications</b>	
<b>Application des développements méthodologiques pour l'identification et la caractérisation des protéines et des glycoprotéines à différentes thématiques biologiques</b> .....	133
<b>Chapitre 3</b> : Identification et caractérisation de protéines et de glycoprotéines du parasite <i>Toxoplasma gondii</i> .....	135
<b>Chapitre 4</b> : Application de la méthodologie protéomique développée à l'étude des protéines des vins de Champagne.....	159
<b>Chapitre 5</b> : Quantification de HbA2 et étude des mutations de l'hémoglobine humaine.....	175
<b>CONCLUSION GENERALE</b> .....	211
<b>PARTIE EXPERIMENTALE</b> .....	217

<b>INTRODUCTION GENERALE</b> .....	1
------------------------------------	---

<b><u>INTRODUCTION BIBLIOGRAPHIQUE</u></b> .....	7
--	---

<b>Chapitre 1 : L'analyse des protéines N-glycosylées par spectrométrie de masse</b> .....	9
--	---

<b>1. La glycosylation une des modifications post-traductionnelles les plus courantes</b> .....	9
<b>1.1. Les modifications post-traductionnelles</b> .....	9
<b>1.2. Le cas des glycosylations</b> .....	11
1.2.1. Les glycosylations.....	11
1.2.2. Les N-glycosylations.....	12
<b>2. Stratégies pour la caractérisation des protéines N-glycosylées par spectrométrie de masse</b> .....	15
<b>2.1. Analyse des glycoprotéines entières</b> .....	16
2.1.1. De l'échantillon biologique complexe aux glycoprotéines : techniques de séparations et d'enrichissements.....	16
2.1.2. Analyse de glycoprotéines entières par spectrométrie de masse.....	17
<b>2.2. Des glycoprotéines aux glycopeptides</b> .....	18
2.2.1. Utilisation de différentes endoprotéases pour générer les glycopeptides.....	19
2.2.2. Méthodes de fractionnement et d'enrichissement des glycopeptides.....	19
<b>2.3. Analyse des peptides déglycosylés</b> .....	20
<b>2.4. Analyse des glycopeptides intacts</b> .....	22
2.4.1. Détection des glycopeptides grâce aux ions diagnostiques.....	22
2.4.2. L'analyse MS <sup>n</sup> des glycopeptides.....	24
<b>3. Conclusion</b> .....	28

<b>Chapitre 2 : Outils et stratégies pour l'analyse protéomique : application à l'analyse des glycoprotéines</b> .....	29
--	----

<b>1. La spectrométrie de masse : un outil central pour la caractérisation des protéines</b> .....	30
<b>1.1. Les sources d'ionisation</b> .....	30
1.1.1. La source électrospray (ESI).....	31
1.1.1.1. Processus d'ionisation-désorption ESI.....	31
1.1.1.2. La source nano-électrospray.....	31
1.1.1.3. L'ionisation des glycoprotéines et des glycopeptides.....	32
1.1.2. MALDI.....	33
<b>1.2. Les analyseurs</b> .....	34
1.2.1. Le quadripôle.....	34
1.2.2. La trappe ionique.....	35
1.2.3. Le Q-TOF : un analyseur en tandem.....	38
<b>1.3. La fragmentation peptidique</b> .....	39
1.3.1. Nomenclature des fragments peptidiques.....	39
1.3.2. La fragmentation induite par collision (CID) : le modèle du proton mobile.....	40
1.3.3. La dissociation par transfert d'électron (ETD).....	41
<b>2. Les stratégies protéomiques basées sur les techniques séparatives</b> .....	42
<b>2.1. Séparation des protéines</b> .....	43
2.1.1. L'électrophorèse sur gel bidimensionnel (Gel 2D).....	43
2.1.2. L'électrophorèse sur gel monodimensionnel (Gel 1D ou SDS-PAGE).....	43
2.1.3. Séparation des protéines par chromatographie liquide (LC).....	44
<b>2.2. Séparation des peptides</b> .....	44
2.2.1. La séparation des peptides par nano-chromatographie liquide (nanoLC).....	44
2.2.2. La chromatographie liquide multidimensionnelle : l'approche "shotgun".....	45

<b>3. Les stratégies d'identification en analyse protéomique par spectrométrie de masse</b> .....	47
<b>3.1. Les stratégies d'identification de protéines</b> .....	47
3.1.1. L'empreinte peptidique massique (Peptide Mass Fingerprint (PMF)) .....	47
3.1.2. Identification par LC-MS/MS .....	48
3.1.2.1. L'empreinte de fragmentation peptidique (Peptide Fragment Fingerprinting (PFF)) ....	49
3.1.2.2. Le séquençage de novo .....	50
<b>3.2. Les banques de données</b> .....	51
<b>3.3. Evaluation des résultats</b> .....	52
<b>4. Conclusion</b> .....	54

## **RESULTATS** .....

<b>PARTIE 1 : Développements</b>	
<b>Recherche de l'amélioration des méthodologies d'analyses LC-MS/MS pour l'identification et la caractérisation des protéines et des glycoprotéines</b> .....	69

<b>Chapitre 1 : Développement de nouvelles méthodes pour l'évaluation des performances d'un couplage nanoLC-MS/MS</b> .....	71
---	----

<b>1. Description du système nanoLC-Chip/HCT Ultra</b> .....	73
<b>2. Optimisation des conditions chromatographiques</b> .....	75
2.1. Optimisation des méthodes chromatographiques .....	75
2.2. Choix du modifieur organique : Optimisation d'une méthode nanoLC-MS utilisant le méthanol comme modifieur organique de la phase mobile .....	76
2.2.1. Conditions expérimentales .....	77
2.2.2. La contrepression .....	77
2.2.3. La sensibilité .....	78
2.2.4. Pouvoir éluatif .....	79
2.2.5. Résolution chromatographique .....	81
2.3. Discussion et conclusion .....	82
<b>3. Optimisation des paramètres de la trappe ionique</b> .....	84
3.1. Optimisation de l'interface .....	84
3.1.1. La source .....	84
3.1.2. La transmission .....	84
3.2. Optimisation des paramètres d'acquisition en mode MS/MS automatique .....	84
3.2.1. Les cycles d'acquisition en mode MS/MS automatique .....	84
3.2.2. Conditions expérimentales .....	85
3.2.3. Le nombre de spectres moyennés .....	86
3.2.4. Nombre de parents sélectionnés pour être fragmentés en MS/MS .....	87
3.2.5. Exclusion temporaire des ions parents après sélection .....	88
3.3. Discussion et conclusion .....	88
<b>4. Développement de tests permettant d'évaluer les performances du couplage nanoLC-Chip/HCT Ultra</b> .....	90
4.1. Intensité des pics chromatographiques .....	91
4.2. Largeur des pics chromatographiques à mi-hauteur .....	92
4.3. Différence entre les temps de rétention des deux ions de référence .....	92
4.4. Résolution spectrale .....	93
4.5. Identification de la protéine .....	93
4.6. Discussion et conclusion .....	94
<b>5. Nouveauté instrumentale : la trappe amazon</b> .....	95
5.1. Géométrie de l'instrument .....	95

5.2. Vitesse de balayage.....	95
5.3. Résolution spectrale.....	96
5.4. Sensibilité.....	98
5.5. Conclusion.....	98
6. Discussion et conclusion.....	99

## **Chapitre 2 : Développements méthodologiques pour la détection et la caractérisation des protéines N-glycosylées.....101**

<b>1. Introduction.....</b>	<b>102</b>
1.1. Une glycoprotéine modèle : l'alpha-1 glycoprotéine acide humaine (AGP).....	102
1.2. Les ions diagnostiques.....	103
<b>2. Comparaison de différentes géométries d'instruments pour la génération d'ions diagnostiques.....</b>	<b>106</b>
<b>2.1. Génération d'ions diagnostiques par CID en source.....</b>	<b>106</b>
2.1.1. Interface comprenant un capillaire de transmission et un skimmer.....	106
2.1.2. Interface comprenant un capillaire de transmission et des funnels.....	108
2.1.3. Interface de type Z-spray.....	110
2.1.4. Discussion et conclusion.....	112
<b>2.2. La génération d'ions diagnostiques par CID dans la cellule de collision.....</b>	<b>113</b>
2.2.1. Les trappes ioniques.....	113
2.2.2. Les Q-TOF.....	114
2.2.3. Discussion et conclusion.....	116
<b>2.3. CID en source vs CID dans la cellule de collision pour un Q-TOF.....</b>	<b>117</b>
2.3.1. Intensité des ions diagnostiques en fonction du mode de fragmentation.....	117
2.3.2. Type d'ion diagnostique généré en fonction du mode de fragmentation.....	118
2.3.3. Discussion et conclusion.....	118
<b>2.4. Discussion et conclusion.....</b>	<b>119</b>
<b>3. Stratégie développée pour la caractérisation des protéines N-glycosylées.....</b>	<b>121</b>
3.1. Digestion enzymatique : obtention de glycopeptides.....	121
3.2. Analyse LC-MS : génération d'ions diagnostiques.....	122
3.3. Analyse LC-MS/MS : caractérisation des glycopeptides.....	122
<b>4. Conclusion.....</b>	<b>125</b>

## **PARTIE 2 : Applications**

### **Application des développements méthodologiques pour l'identification et la caractérisation des protéines et des glycoprotéines à différentes thématiques biologiques.....133**

## **Chapitre 3 : Identification et caractérisation de protéines et de glycoprotéines du parasite *Toxoplasma gondii*.....135**

<b>1. Contexte général de l'étude.....</b>	<b>135</b>
1.1. Introduction.....	135
1.2. Le parasite <i>Toxoplasma gondii</i> .....	136
1.2.1. Cycle d'infection et de reproduction de <i>T. gondii</i> .....	136
1.2.2. Reconnaissance et invasion de la cellule hôte.....	137
<b>2. Etude des glycoprotéines impliquées dans le mécanisme de reconnaissance et d'infection de la cellule hôte par le parasite <i>T. gondii</i>.....</b>	<b>139</b>
2.1. Contexte de l'étude.....	139
2.2. Méthodologie.....	139
2.2.1. Préparation d'échantillon et analyse protéomique.....	139

2.2.2. Identification des glycoprotéines par analyse nanoLC-MS/MS et recherche dans le génome complet non annoté de <i>T. gondii</i> .....	140
2.2.3. Caractérisation de protéines N-glycosylées.....	142
<b>2.3. Résultats publiés</b> .....	142
<b>2.4. Conclusion et perspectives</b> .....	143
<b>3. Caractérisation des glycosylations de la protéine GAP50</b> .....	147
<b>3.1. Contexte de l'étude</b> .....	147
<b>3.2. Méthodologie</b> .....	147
3.2.1. Préparation d'échantillon.....	147
3.2.2. Caractérisation des glycosylations de GAP 50 .....	147
<b>3.3. Résultats</b> .....	148
3.3.1. Digestion enzymatique.....	148
3.3.2. Analyse nanoLC-MS : détection des ions diagnostiques.....	148
3.3.3. Analyse nanoLC-MS/MS : caractérisation des glycopeptides .....	149
3.3.4. Résultats .....	150
<b>3.4. Conclusion et perspectives</b> .....	152
<b>4. Etude des facteurs nucléaires impliqués dans la régulation de l'expression des gènes de <i>T. gondii</i></b> .....	153
<b>4.1. Contexte de l'étude</b> .....	153
<b>4.2. Méthodologie</b> .....	153
4.2.1. Préparation d'échantillon.....	154
4.2.2. Analyse protéomique .....	154
<b>4.3. Résultats</b> .....	154
<b>4.4. Discussion et conclusion</b> .....	155

## **Chapitre 4 : Application de la méthodologie protéomique développée à l'étude des protéines des vins de Champagne.....159**

<b>1. Analyse protéomique différentielle des vins de Champagne contaminés par <i>Botrytis cinerea</i></b> .....	160
<b>1.1. Contexte de l'étude</b> .....	160
<b>1.2. Stratégie protéomique</b> .....	160
1.2.1. Préparation d'échantillon.....	160
1.2.2. Analyse protéomique .....	160
<b>1.3. Résultats publiés</b> .....	161
<b>1.4. Conclusion et perspectives</b> .....	161
<b>2. Caractérisation des glycosylations de l'invertase vacuolaire de raisin</b> .....	165
<b>2.1. Contexte de l'étude</b> .....	165
<b>2.2. Méthodologie</b> .....	166
2.2.1. Purification de l'invertase vacuolaire de raisin .....	166
2.2.2. Caractérisation des glycosylations de l'invertase vacuolaire de raisin .....	166
2.2.2.1. Analyse des glycopeptides intacts .....	166
2.2.2.2. Analyse de la protéine déglycosylée .....	166
<b>2.3. Résultats</b> .....	166
2.3.1. Purification de l'invertase vacuolaire de raisin .....	166
2.3.2. Caractérisation des glycosylations de l'invertase vacuolaire de raisin .....	167
2.3.2.1. Analyse des glycopeptides intacts .....	167
2.3.2.2. Analyse de la protéine déglycosylée .....	169
<b>2.4. Conclusion et perspectives</b> .....	171

## **Chapitre 5 : Quantification de HbA2 et étude des mutations de l'hémoglobine humaine.....175**

<b>1. Introduction</b> .....	175
<b>2. Quantification de HbA2</b> .....	177
<b>2.1. Contexte de l'étude</b> .....	177

<b>2.2. Stratégie</b> .....	178
<b>2.3. Développement méthodologique</b> .....	179
2.3.1. Choix de l'endoprotéase.....	179
2.3.2. Protocole de digestion.....	180
2.3.3. Choix des peptides spécifiques pour la quantification .....	181
2.3.4. Séparation chromatographique.....	182
2.3.5. Limite de quantification .....	183
2.3.6. Etalonnage par LC-MS.....	184
<b>2.4. Résultats et discussion</b> .....	184
<b>2.5. Conclusion et perspectives</b> .....	187
<b>3. Caractérisation de mutations de l'hémoglobine</b> .....	191
<b>3.1. Contexte de l'étude</b> .....	191
<b>3.2. Résultats</b> .....	193
3.2.1. Caractérisation de la mutation "Noah Mehmet Oeztuerk" .....	193
3.2.2. Caractérisation d'une nouvelle mutation sur la chaîne beta 93Cys→Ser.....	194
<b>3.3. Conclusion et perspectives</b> .....	196
<b>4. Quantification de HbA<sub>2</sub> et détection des mutations les plus courantes tels que Hb C, E, S et A2' par une approche protéomique</b> .....	199
4.1. Contexte de l'étude.....	199
4.2. Résultats.....	200
4.3. Conclusion et perspectives.....	202
<b>5. Conclusion</b> .....	203
<b>CONCLUSION GENERALE</b> .....	211
<b>PARTIE EXPERIMENTALE</b> .....	217

# Liste des principales abréviations

ACN :	Acétonitrile
ADN :	Acide désoxyribonucléique
BPC :	Base Peak Chromatogram
BLAST :	Basic Local alignment Search Tool
BSA :	Bovin Serum Albumin
CID :	Collision-Induced Dissociation
Da :	Dalton
DDA :	Data Dependent Analysis
ECD :	Electron Capture Dissociation
EIC :	Extracted Ion Chromatogram
ESI :	Electrospray Ionisation
ETD :	Electron Transfer Dissociation
eV :	electron Volt
Gel 1D :	Gel d'électrophorèse monodimensionnel
Gel 2D :	Gel d'électrophorèse bidimensionnel
Hb :	Hémoglobine
HPLC :	High Performance Liquid Chromatography
ISCID :	Ion Source Collision Induced Dissociation
IT :	Ion Trap
LC :	Liquid Chromatography
LOD :	Limit Of Detection
LOQ :	Limit Of Quantification
MALDI :	Matrix-Assisted Laser Desorption/Ionisation
MS :	Spectrométrie de Masse
MS/MS :	Spectrométrie de masse en tandem
m/z :	mass to charge ratio
NanoLC-MS/MS :	Nano liquid chromatography tandem mass spectrometry
PFF :	Peptide Fragment Fingerprint
PMF :	Peptide Mass Fingerprint
ppm :	parties par million
PTM :	Post-translational Modification
Q :	Quadripôle
RP :	Reverse Phase
SCX :	Strong Cation Exchange
SDS-PAGE :	Sodium DodecylSulfate PolyAcrylamide Gel Electrophoresis
TOF :	Time of Flight
UHPLC :	Ultra High Pressure Liquid Chromatography

# Liste des acides aminés les plus courants

Burlingame A.L. and Carr S. A., Mass Spectrometry in the biological sciences, Humana Press, 1996, p. 542-545.

Acides aminés	Codes		Masse monoisotopique du résidu <sup>a</sup> (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

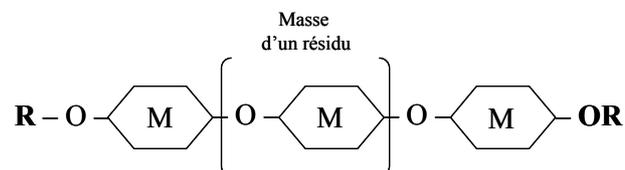
# Liste des monosaccharides les plus couramment observés

Burlingame A.L. and Carr S. A., Mass Spectrometry in the biological sciences, Humana Press, 1996, p. 542-545.

Famille	Monosaccharides	Abréviations	Masses monoisotopiques des résidus non dérivés (Da) <sup>a</sup>
Pentoses (Pent)	Xylose	Xyl	132,04
Deoxyhexoses (DeoxyHex)	Fucose	Fuc	146,06
Hexoses (Hex)	Galactose	Gal	162,05
	Glucose	Glc	162,05
	Mannose	Man	162,05
N-acétylhexosamines (HexNAc)	N-acétylglucosamine	GlcNAc	203,08
	N-acétylgalactosamine	GalNAc	203,08
Acides sialiques (SA)	Acide N-acétylneuraminique	NeuAc	291,10

<sup>a</sup> Masse d'un résidu + Masse des groupements terminaux non dérivés (R+OR ; 18,01 Da) = Masse d'un monosaccharide.

Addition des masses de plusieurs résidus + 18,01 = masse d'un polysaccharide.



M: représentation schématique d'un monosaccharide quelconque

R et OR : groupements terminaux d'un polysaccharide.

R = H pour un monosaccharide non dérivé.

# **INTRODUCTION GENERALE**





# INTRODUCTION GENERALE

Après plus de 10 ans d'utilisation en biologie, la protéomique est souvent considérée comme une méthodologie de routine. Ce point de vue est quotidiennement infirmé puisque, bien souvent, l'analyse protéomique ne permet pas d'apporter toutes les données expérimentales nécessaires pour répondre aux questions posées par les biologistes. Les limitations de l'analyse protéomique concernent bien sûr la sensibilité, la précision dans les mesures de masse mais aussi la possibilité de réaliser et d'exploiter de grandes séries d'analyses. A ces limitations, s'ajoute celle liée à la caractérisation complète des modifications post-traductionnelles, avec leur hétérogénéité, qui reste pour l'instant difficilement réalisable.

Il est donc clair qu'il est encore nécessaire de développer les méthodologies de l'analyse protéomique, qui ne répond pas, et de loin, à tous les problèmes posés en biologie. C'est dans ce cadre que se situe mon travail de thèse.

Mon travail thèse s'est principalement focalisé sur un aspect du développement méthodologique en protéomique : l'amélioration de la méthodologie instrumentale en nanoLC-MS/MS et plus particulièrement dans le cadre de la caractérisation des protéines N-glycosylées et de la micro-hétérogénéité des N-glycosylations. Ces dernières années, il a été montré que, contrairement à ce qui a longtemps été supposé, des différences apparemment minimes dans la structure des glycanes des N-glycosylations pouvaient entraîner des changements radicaux dans l'activité biologique des protéines concernées. J'ai donc tenté de développer des moyens nouveaux pour mieux caractériser la micro-hétérogénéité des sites de glycosylations, c'est-à-dire de décrire l'ensemble des structures glycaniques présentes sur un même site de N-glycosylation. J'ai en particulier essayé de comprendre pourquoi il était difficile de détecter les glycopeptides obtenus après digestion des protéines et tenté d'optimiser cette détection en utilisant l'approche des "ions diagnostiques" qui devait, selon la littérature, apporter plus de sensibilité. Enfin, nous avons orienté nos développements méthodologiques pour répondre à une demande très fréquente des biologistes : fournir des résultats fiables sur des échantillons qui ne peuvent être obtenus qu'une seule fois et disponibles en quantités très faibles. Les stratégies analytiques à développer dans ces cas sont évidemment très différentes de celles couramment utilisées lors de travaux où les échantillons biologiques sont disponibles en quantités suffisantes pour pouvoir répéter les expériences en optimisant les conditions d'analyse. Les développements méthodologiques obtenus pendant mon travail ont ensuite été appliqués dans le cadre de plusieurs collaborations avec des biologistes. Mon manuscrit de thèse est présenté comme suit :

**L' INTRODUCTION BIBLIOGRAPHIQUE** de ce manuscrit de thèse présente d'abord, dans un premier chapitre, les méthodes de caractérisation des N-glycosylations utilisées lorsque j'ai commencé mon travail et, dans un second chapitre, l'instrumentation et les stratégies protéomiques utilisées en général. Cette introduction tentera également de mettre en évidence les axes pour lesquels des progrès méthodologiques sont à faire en analyse protéomique des glycoprotéines.

**Les RESULTATS** obtenus pendant mon travail de thèse sont ensuite exposés en deux parties : une partie "Développements méthodologiques" suivie d'une partie "Application à des problèmes biologiques".

**La partie "Développements méthodologiques"** est focalisée sur la LC-MS/MS et comprend 2 chapitres :

- **Le chapitre 1** porte sur l'optimisation d'un couplage nanoLC-MS/MS et sur le développement de méthodes permettant de vérifier les performances de ce système. Les optimisations ont porté à la fois sur la partie chromatographique et sur le paramétrage du spectromètre de masse. Ce travail m'a permis de développer un ensemble de tests permettant de vérifier les performances de chacun des modules du couplage et d'établir des diagnostics en cas de panne.

- **Le chapitre 2** est consacré au développement d'une méthodologie pour la caractérisation des protéines N-glycosylées. Notre stratégie est basée sur l'utilisation de techniques protéomiques classiques (digestion enzymatique, LC-MS et LC-MS/MS) et sur la génération d'ions dits "diagnostiques" qui sont spécifiques de la fragmentation des glycopeptides. Plusieurs méthodes d'obtention des ions diagnostiques ont été expérimentées (CID en source ou CID dans la cellule de collision) sur différents spectromètres de masse disponibles au laboratoire, chacun ayant des géométries de source et d'analyseur différentes.

**La partie "Application à des problèmes biologiques"** décrit les résultats obtenus lors de 3 collaborations avec des biologistes où nous avons appliqué les méthodes d'analyse protéomique que nous avons mises au point et décrites dans la partie "Développements méthodologiques".

- **Le chapitre 3** présente l'utilisation séquentielle de plusieurs stratégies appliquées à l'étude du parasite *Toxoplasma gondii* et se divise en trois points :

- Nous avons étudié les glycoprotéines impliquées dans le mécanisme d'invasion (le glideosome) de la cellule hôte par le parasite et nous avons caractérisé un site de glycosylation de l'une d'entre elles, GAP50;

- Dans un deuxième temps, une stratégie plus adaptée, nous a permis de caractériser plus finement la protéine GAP50 en décrivant la micro-hétérogénéité pour chacun des 3 sites de glycosylation;

- Enfin, grâce à une augmentation de sensibilité obtenue par une série d'optimisations, nous avons pu identifier une série de facteurs de transcription impliqués dans le mécanisme

d'interconversion entre la forme virulente tachyzoïte et la forme dormante bradyzoïte du parasite. Ce travail a permis d'identifier un facteur nucléaire qui a été nommé TgNF3 et dont le rôle a pu être étudié de manière plus approfondie par nos collaborateurs.

▪ **Le chapitre 4** présente la première étude protéomique réalisée sur des vins de Champagne ainsi qu'une caractérisation des glycosylations d'une protéine importante pour la moussabilité du vin, l'invertase vacuolaire :

○ Une première partie est consacrée à une analyse différentielle par gel d'électrophorèse bidimensionnel entre un vin vinifié à partir de raisins contaminés par le champignon *Botrytis cinerea* et un vin sain;

○ La seconde partie de cette étude a porté sur la caractérisation de l'invertase vacuolaire de raisin qui constitue une des protéines majeures du vin et qui possède 12 sites potentiels de N-glycosylation.

▪ Enfin, **le chapitre 5** porte sur la mise au point de stratégies de quantification d'une forme d'hémoglobine (HbA<sub>2</sub>) et sur la caractérisation de formes mutées des hémoglobines :

○ Une méthode de quantification fine de l'hémoglobine humaine HbA<sub>2</sub> a ainsi été développée dans le but d'obtenir une méthode de référence pour certifier des échantillons standards pour étalonner les instruments couramment utilisés pour quantifier l'hémoglobine HbA<sub>2</sub> dans les laboratoires d'hématologie;

○ Nous avons également développé une stratégie permettant de caractériser de nouvelles mutations de l'hémoglobine lorsque les méthodes d'analyse classiques sont inefficaces;

○ Enfin, nous avons élargi la méthode de quantification à la détection des mutations les plus courantes.

**INTRODUCTION**  
**BIBLIOGRAPHIQUE**





**INTRODUCTION**  
**BIBLIOGRAPHIQUE**

---

# Chapitre 1

## L'analyse des protéines N-glycosylées par spectrométrie de masse

---

### 1. La glycosylation une des modifications post-traductionnelles les plus courantes

#### 1.1. Les modifications post-traductionnelles

Une modification post-traductionnelle d'une protéine se définit comme une modification chimique covalente catalysée par des enzymes, qui se produit après que l'ADN ait été transcrit en ARN, elle-même traduite en protéines.

Les modifications post-traductionnelles (Post-Translational Modifications PTMs) modulent souvent profondément les propriétés biologiques d'une protéine. Ces PTMs peuvent prendre plusieurs formes. Elles peuvent consister en :

- Une modification de la structure primaire de la protéine par addition ou clivage d'un ou plusieurs acides aminés;
- Une modification des extrémités N- ou C-terminales par fixation d'un groupement;
- Une modification de la chaîne latérale d'un acide aminé par fixation covalente d'un groupement;
- Un clivage protéolytique.

Le Tableau 1 présente les PTMs les plus courantes. Plusieurs PTMs, différentes ou non, peuvent affecter la même protéine. Environ 80% des cellules de mammifère contiennent des PTMs. Elles vont déterminer l'activité d'une protéine et donc son rôle au sein de la cellule. Elles sont impliquées dans tous les processus biologiques tels que la division cellulaire, la différenciation ou les interactions hôtes-pathogènes [M. Mann et Jensen 2003]. C'est pourquoi leur caractérisation est un enjeu fondamental pour l'étude du fonctionnement cellulaire.

Les PTMs 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 se produit sur des acides aminés donnés et entraîne des différences de masse par rapport à l'acide aminé non-modifié. Ces différences de masse sont détectables par spectrométrie de masse [Jensen 2006].

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

**Tableau 1.** PTMs les plus couramment observées, différences de masse générées et exemples de leurs fonctions biologiques [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 différente 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'enrichissements, de marquages chimiques spécifiques ou des combinaisons d'enzymes de digestions notamment [Hoffman et al. 2008].

## 1.2. Le cas des glycosylations

### 1.2.1. Les glycosylations

Une glycosylation correspond à la liaison covalente d'un glycane sur la chaîne latérale d'un acide aminé d'une protéine. Ces glycanes sont constitués d'une succession de monosaccharides qui forment une structure polysaccharidique. Les résidus les plus fréquemment observés sont décrits dans le Tableau 2.

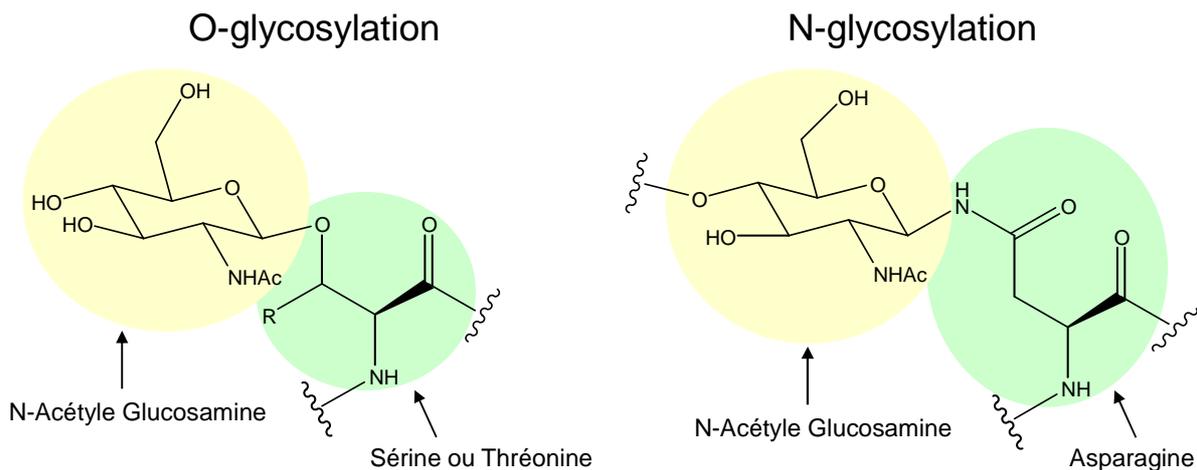
Monosaccharides	Abréviations	Symboles	Masses monoisotopiques des résidus (Da)
Xylose	Xyl	★	132,04
Fucose	Fuc	▲	146,06
Galactose	Gal	●	162,05
Glucose	Glc	●	162,05
Mannose	Man	●	162,05
N-acétylglucosamine	GlcNAc	■	203,08
N-acétylgalactosamine	GalNAc	■	203,08
Acide N-acétylneuraminique	NeuAc	◆	291,10

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

La glycosylation est présente aussi bien chez les eucaryotes que chez les procaryotes [Upreti et al. 2003], et représente une des PTMs les plus courantes. En effet, d'après une étude réalisée par Apweiler à partir de la banque protéique SWISS-PROT, plus de la moitié des protéines existantes seraient glycosylées [R. Apweiler et al. 1999]. Elles représentent la majorité des marqueurs de surface cellulaires ainsi que des protéines sécrétées. Elles ont une influence fondamentale sur les propriétés physicochimiques de la protéine telles que sa stabilité, sa conformation ou sa capacité à être dégradée par une protéase par exemple [Rudd et al. 2001]. De plus, elles jouent un rôle clef dans un grand nombre de processus biologiques comme le développement embryonnaire ou la communication cellulaire [Dwek 1996]. De nombreuses études ont permis d'observer des corrélations entre des glycosylations anormales et des cas de cancer, de maladies inflammatoires ou du foie

[Dube et Bertozzi 2005, Blomme et al. 2009]. La caractérisation des glycoprotéines peut ainsi être utilisée pour développer des vaccins, des médicaments ou des méthodes de diagnostics pour ces maladies [Dube et Bertozzi 2005].

Il existe deux principaux types de liaisons sucre-protéine qui dépendent de la nature de la liaison covalente entre le glycane et l'acide aminés (Figure 1) : les O-glycosylations et les N-glycosylations.



**Figure 1.** Structures chimiques des liaisons sucre-protéine pour les O et N-glycosylations

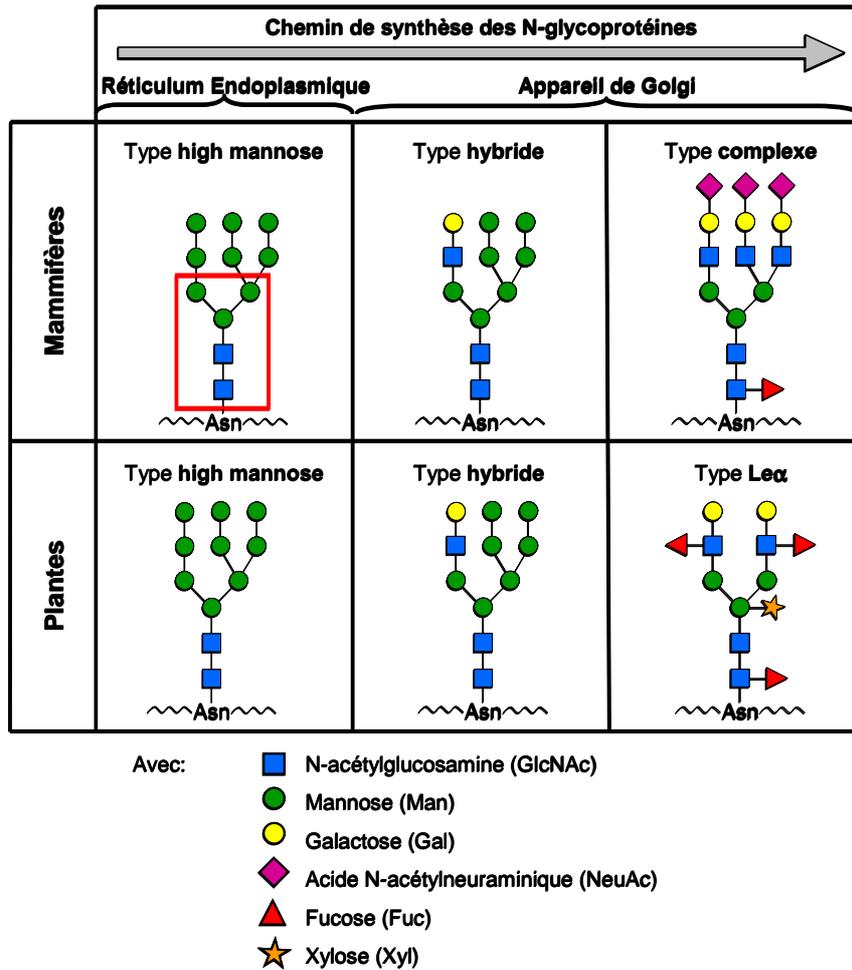
**Les O-glycosylations :** le résidu monosaccharide se lie à la protéine par le groupement hydroxyl des résidus sérine (Ser), thréonine (Thr), hydroxyproline (Hyp) ou hydroxylysine. La plus répandue est la O-glycosylation de type mucine où le résidu N-acétylglucosamine (GlcNAc) du glycane se lie à la sérine ou à 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 (X pouvant être n'importe quel acide aminé sauf une proline) qui est appelé séquence consensus. Dans de rares cas l'acide aminé Ser/Thr est remplacé par une cystéine [Satomi et al. 2004]. Leur chemin de biosynthèse étant commun, tous les N-glycanes ont un squelette pentasaccharide identique qui est constitué de 3 mannoses et de 2 N-acétylglucosamines ( $\text{Man}_3\text{GlcNAc}_2$ ) [Helenius et Aebi 2001].

Dans la suite de ce chapitre nous nous intéresserons plus particulièrement au cas des N-glycosylations.

### 1.2.2. Les N-glycosylations

Les N-glycosylations sont constituées d'un tronc commun dont la séquence est :  $\text{Man}_3\text{GlcNAc}_2$  (entourée en rouge sur la Figure 2). Le GlcNAc proximal est fixé sur l'asparagine de la séquence consensus.



**Figure 2.** Exemple de structures de N-glycanes couramment trouvés dans les plantes et les mammifères aux différents stades de leur biosynthèse [Bednarczyk 2008]

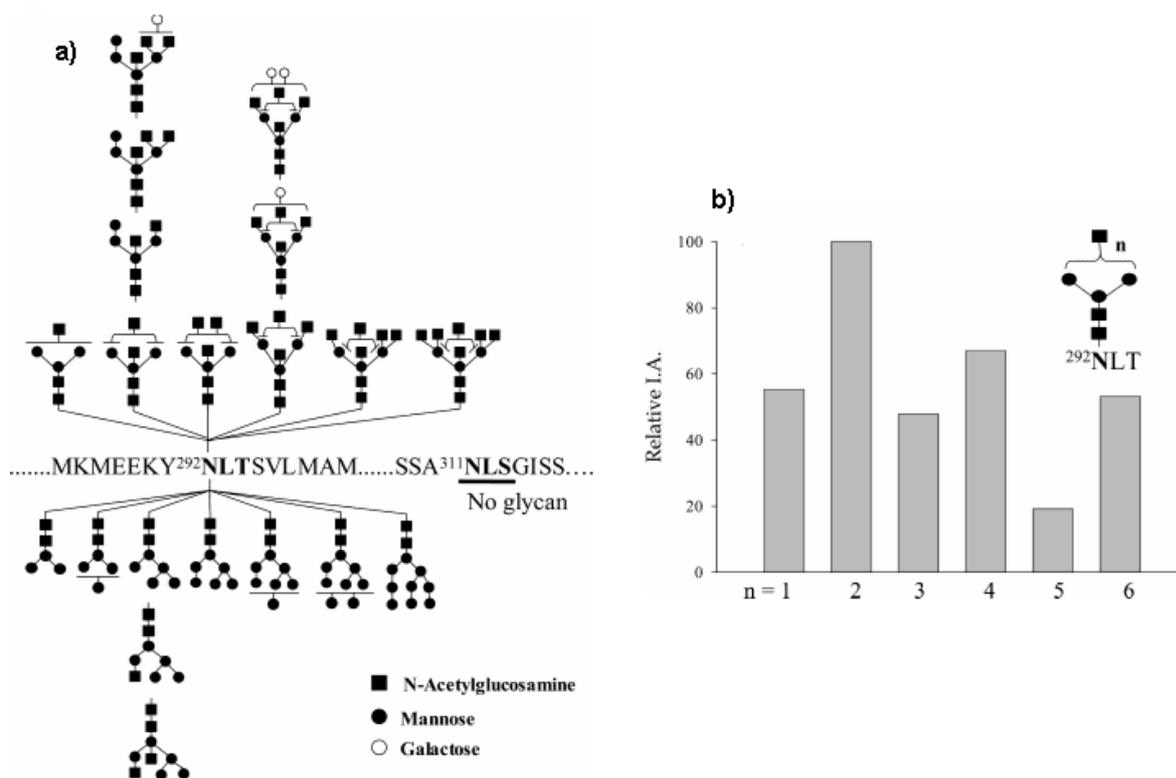
Les N-glycanes peuvent être classés en 3 types en fonction de la nature et de la localisation des ramifications, appelées antennes [Dalpathado et Desaire 2008] :

- Les **high-mannoses** quand les ramifications sont constituées uniquement de mannoses;
- Les N-glycanes **complexes** quand les antennes contiennent des N-acétylglucosamines, des galactoses, des fucoses et/ou des acides sialiques. Ils constituent la classe la plus abondante des oligosaccharides présents chez les mammifères et leur structure est caractérisée par un nombre variable d'antennes;
- Les N-glycanes **hybrides**, comportent une antenne de type high-mannose alors que la ou les autres sont de type complexe.

Il existe donc de nombreux types de N-glycosylations chez les eucaryotes dont la structure varie considérablement selon qu'elles soient biosynthétisées dans les cellules de mammifères ou de plantes [Gomord et Faye 2004] (Figure 2). De plus, les structures glycaniques ne sont pas statiques et varient de manière qualitative et quantitative en fonction de l'avancement d'un processus biologique

[Dwek 1996]. En effet, ce qu'on appelle une glycoprotéine est en fait un mélange de très nombreuses molécules qui peuvent avoir des fonctions biologiques très différentes, bien qu'elles aient toutes exactement la même séquence en acides aminés. Cette hétérogénéité est donc uniquement liée à des différences de glycosylations.

On distingue deux types d'hétérogénéité : la macro- et la micro-hétérogénéité [Morelle et al. 2006]. Une N-glycosylation se situe obligatoirement sur une asparagine comprise dans une séquence consensus (Asn-X-Ser/Thr) que l'on nomme site de glycosylation. Pourtant ce site peut ne pas être occupé par une glycosylation. C'est ce qu'on appelle la **macro-hétérogénéité** d'une glycoprotéine. Ensuite, chaque site consensus glycosylé peut porter plusieurs glycanes différents (longueur des antennes, ramification, etc). Ceci correspond à la **micro-hétérogénéité**. Par exemple, pour une glycoprotéine contenant 3 sites de glycosylations pouvant chacun porter 10 glycanes, il est possible d'observer jusqu'à 1000 glycoformes différentes de cette protéine.



**Figure 3.** (a) Structures de glycopeptides de l'ovalbumine (b) Abondance relative de chaque glycoforme sur 292 NLT, d'après [An et al. 2003]

La Figure 3 illustre la grande micro-hétérogénéité qui peut affecter un seul site de glycosylation. Toutes les structures glycaniques représentées (Figure 3a) sont fixées sur l'asparagine en position 292 de l'ovalbumine dans les proportions indiquées sur le diagramme (Figure 3b).

## 2. Stratégies pour la caractérisation des protéines N-glycosylées par spectrométrie de masse

Pour caractériser une glycoprotéine, il est nécessaire d'identifier la séquence primaire de la protéine, mais aussi de déterminer les sites de glycosylations et d'établir la structure des glycanes et leur hétérogénéité [Temporini et al. 2008]. Ceci pose un certain nombre de difficultés car :

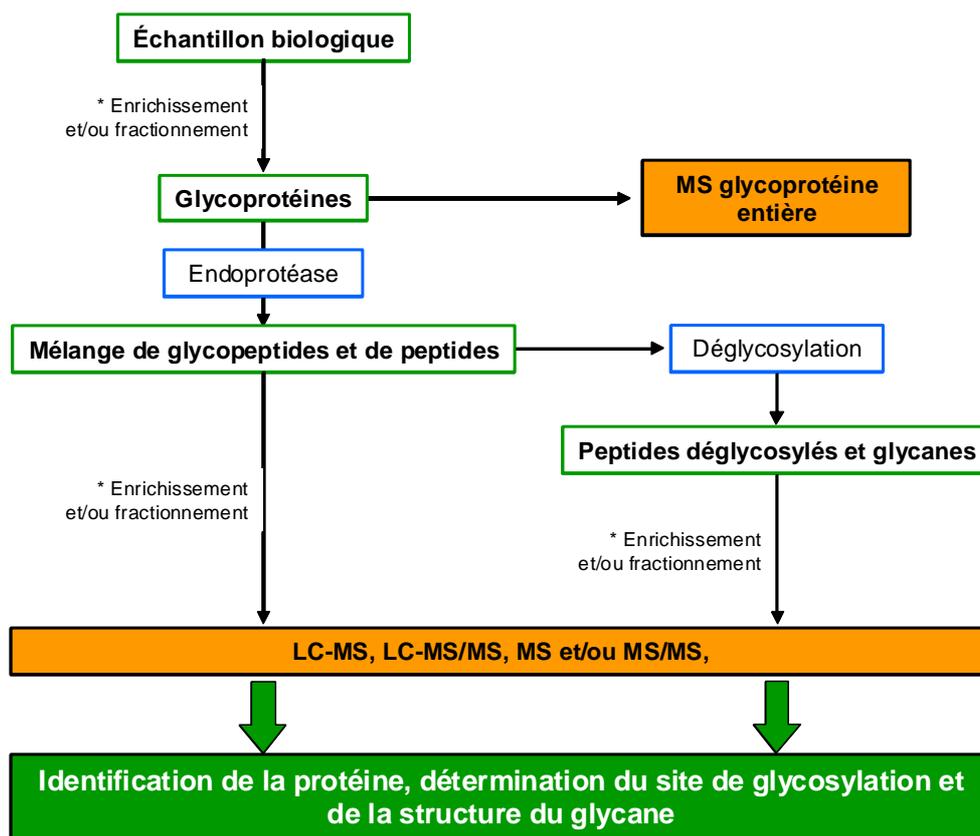
- Une glycoprotéine peut posséder plusieurs séquences consensus glycosylées ou non;
- Chaque site peut porter différents glycanes possédant des antennes de taille, de nature ou en nombre différent; c'est ce qu'on appelle l'anténarité;
- Les glycosylations ne sont pas statiques; elles varient de manière qualitative et quantitative en fonction de l'avancement d'un processus biologique.

La présence de plusieurs sites de glycosylation, la grande hétérogénéité de structure glycanique, et le nombre élevé de stéréoisomères et de régioisomères possibles rendent l'analyse des glycoprotéines plus complexe que celle des protéines non glycosylées. C'est pourquoi, pour les études protéomiques, les glycoprotéines sont souvent déglycosylées avant leur analyse [Brooks 2009]. Ceci est en train de changer grâce à de nombreux développements instrumentaux [Wuhrer et al. 2007, Dalpathado et Desaire 2008].

La Figure 4 présente les techniques les plus couramment utilisées pour l'analyse de glycoprotéines par spectrométrie de masse. Ces méthodologies comportent plusieurs étapes qui peuvent être résumées de la manière suivante :

- Purification des glycoprotéines à partir de l'échantillon biologique (facultatif);
  - Analyse des glycoprotéines entières par spectrométrie de masse afin d'obtenir des informations sur son hétérogénéité;
    - Digestion des glycoprotéines par une endoprotéase afin d'obtenir un ensemble de peptides et de glycopeptides;
- Et / ou {
- Analyse des glycopeptides entiers (purifiés ou non) par spectrométrie de masse;
  - Déglycosylation des glycopeptides afin d'obtenir un ensemble de peptides "nus" et de glycanes "libres" qui sont ensuite analysés par spectrométrie de masse;
- Interprétation des résultats permettant l'identification de la protéine, la détermination des sites de glycosylation et la structure de glycanes.

Dans ce chapitre, après avoir décrit les techniques de purification et d'analyse des glycoprotéines entières, nous ferons le parallèle entre les deux stratégies utilisant ou non la déglycosylation des glycopeptides.



\* : étapes facultatives en fonction de la nature de l'échantillon

**Figure 4.** Les stratégies les plus utilisées pour l'analyse de protéines N-glycosylées par spectrométrie de masse

## 2.1. Analyse des glycoprotéines entières

### 2.1.1. De l'échantillon biologique complexe aux glycoprotéines : techniques de séparations et d'enrichissements

Les glycoprotéines sont très hétérogènes et chaque glycoforme est généralement présente dans de faibles proportions dans un échantillon biologique. Il est donc très souvent nécessaire d'utiliser une ou plusieurs étapes de fractionnement afin de faciliter leur caractérisation.

La chromatographie d'affinité est une technique couramment utilisée qui permet d'enrichir l'échantillon en glycoprotéines et ainsi de les analyser plus facilement. Une des phases stationnaires les plus communes est constituée de lectine fixée sur de la sépharose ou de l'agarose. Les lectines sont des protéines qui interagissent de manière sélective avec un type de glycoforme sans le modifier.

Différentes lectines peuvent être utilisées pour isoler des glycoprotéines ayant des glycoformes différentes [Nawarak et al. 2004]. Il existe par exemple la ConA (concanavaleine A) qui est spécifique des glycanes contenant des mannoses ou des glucoses, ou la WGA (wheat germ agglutinin) qui

reconnait les N-acetylglucosamines et les acides sialiques. Il existe également des approches utilisant une combinaison de différentes lectines permettant de capturer plusieurs types de glycoformes [Yang et Hancock 2004].

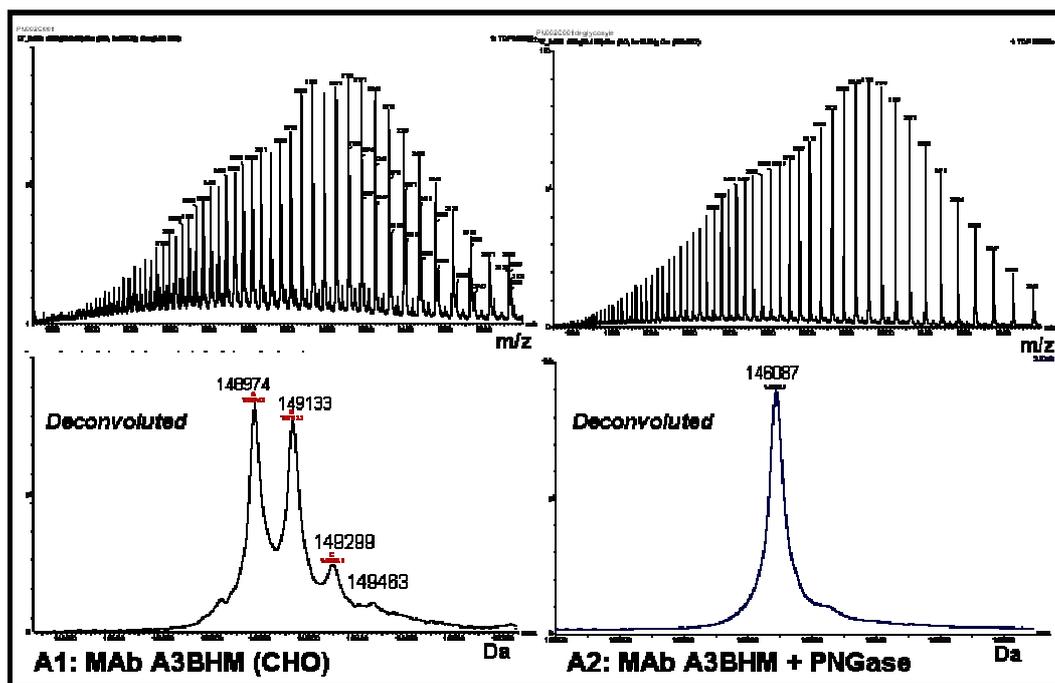
### **2.1.2. Analyse de glycoprotéines entières par spectrométrie de masse**

La caractérisation des glycoprotéines intactes par spectrométrie de masse pose un certain nombre de difficultés qui sont liées à leur hétérogénéité. Comme il a été mentionné précédemment, une glycoprotéine est un mélange d'espèces moléculaires avec la même séquence en acides aminés mais des glycosylations différentes. Par conséquent, lorsqu'on analyse une glycoprotéine par spectrométrie de masse, on obtient un spectre très hétérogène. De plus, l'intensité des ions est divisée par le nombre de glycoformes. Ces contraintes sont d'autant plus grandes que le nombre de glycoformes est important.

Les glycoprotéines intactes peuvent être analysées par MALDI-TOF-MS. Ce type d'instrument ne permet d'obtenir la résolution sur les glycoformes que pour des protéines de petite taille (20 à 40 kDa) et ne comportant que peu de glycoformes. Pour les glycoprotéines très hétérogènes, l'analyse MALDI-TOF ne permet d'obtenir qu'un spectre complexe difficilement interprétable. Cependant, en optimisant le type de matrice et la préparation du dépôt MALDI, il est possible d'obtenir des résultats intéressants [Gimenez et al. 2007].

D'autres études utilisent l'ESI-MS pour l'analyse de glycoprotéines intactes [Wagner-Rousset et al. 2008]. L'ionisation électrospray génère des ions multichargés (Chapitre 2, 1.1.1.). Avec ce type de source, il est donc nécessaire d'utiliser un analyseur possédant une résolution suffisante pour séparer des glycoformes de masse très proche. L'ESI-MS permet d'évaluer l'hétérogénéité glycanique pour des glycoprotéines ne comportant qu'un site de glycosylation, et est généralement utilisé pour l'analyse de protéines recombinantes (Figure 5).

De manière générale, l'analyse des glycoprotéines intactes ne permet pas d'obtenir d'informations significatives dès lors que la glycoprotéine comporte plus d'un site de glycosylation. C'est pourquoi il est nécessaire d'utiliser une approche qui passe par la digestion de la glycoprotéine à l'aide d'une endoprotéase de façon à générer des glycopeptides portant chacun un seul site de glycosylation pour permettre la caractérisation de la micro-hétérogénéité.



**Figure 5.** Spectres correspondant à l'analyse par électrospray positif d'un anticorps monoclonal. A1 : protéine intacte; A2 : protéine après traitement à la PNGase F. En haut les ions multichargés et en bas les spectres déconvolués qui représentent la masse moléculaire de l'anticorps et de ses différentes glycoformes, d'après [Wagner-Rousset et al. 2008]

## 2.2. Des glycoprotéines aux glycopeptides

La caractérisation des glycoprotéines passe par la détermination de la séquence primaire de la protéine, des sites de glycosylations et enfin des glycoformes avec les micro-hétérogénéités. Comme il a été montré dans le paragraphe précédent il est nécessaire d'effectuer une digestion enzymatique avec une endoprotéase pour avoir accès à ces informations. L'ensemble des peptides et glycopeptides obtenus forme un mélange complexe qui pose un certain nombre de difficultés pour leur analyse en LC-MS :

- Les glycopeptides sont plus faiblement ionisables que les peptides non glycosylés et génèrent donc des ions moins intenses;
- La micro-hétérogénéité provoque une division de l'intensité des ions entre les différentes glycoformes;
- Les glycopeptides ont souvent des poids moléculaires plus élevés que les peptides non glycosylés. Les paramètres de transmission du spectromètre de masse doivent donc être spécifiquement adaptés pour les peptides glycosylés;
- Les glycosylations augmentent souvent le caractère hydrophile des glycopeptides qui sont donc insuffisamment retenus et mal séparés sur les colonnes de phase C18 utilisés généralement en protéomique;
- Chaque glycoforme ayant un tant de rétention légèrement différent, la micro-hétérogénéité des glycopeptides provoque un élargissement des pics chromatographiques.

Deux méthodologies complémentaires présentées dans ce chapitre permettent de surmonter ces difficultés:

- La déglycosylation permet d'analyser les glycanes et les peptides séparément [Geyer et Geyer 2006]. Il est ainsi possible d'obtenir des informations de séquence à la fois pour le peptide et le glycan. En revanche, cette méthodologie ne permet pas d'obtenir d'information sur la micro-hétérogénéité s'il y a plusieurs sites de glycosylation.
- L'analyse des glycopeptides intacts [Dalpathado et Desaire 2008], permet de déterminer à la fois le site de glycosylation mais également la micro-hétérogénéité. Néanmoins la séquence en acides aminés du glycopeptide n'est que rarement déterminée dans sa totalité.

Ces deux méthodes nécessitent l'utilisation d'une endoprotéase afin de digérer la glycoprotéine en peptides et glycopeptides. De plus, une étape d'enrichissement du digestat en glycopeptides peut être nécessaire.

### **2.2.1. Utilisation de différentes endoprotéases pour générer les glycopeptides**

La protéolyse des glycoprotéines peut s'effectuer avec une grande variété d'enzymes. La plus couramment utilisée est la trypsine qui a l'avantage d'être très spécifique. La séquence peptidique des glycopeptides générés est donc prévisible. Cependant certains sites de coupure peuvent ne pas être atteints à cause de l'encombrement stérique généré par les glycosylations. De plus certains glycopeptides peuvent contenir plusieurs sites de glycosylations. C'est pourquoi il est également possible d'utiliser d'autres enzymes spécifiques (Glu C, Asp N, Chymotrypsine) ou aspécifiques (Pronase, Protéinase K) afin d'obtenir les glycopeptides les mieux adaptés à l'analyse MS. Enfin, la combinaison de plusieurs enzymes est parfois nécessaire pour parvenir à localiser tous les sites de glycosylation de la protéine [Larsen et al. 2005].

### **2.2.2. Méthodes de fractionnement et d'enrichissement des glycopeptides**

Une fois la glycoprotéine digérée avec une ou des protéases, on obtient un mélange complexe de glycopeptides et de peptides non glycosylés.

Les glycopeptides sont moins facilement ionisables que les peptides non glycosylés, cela peut donc générer une suppression de signal lors de l'analyse MS. De plus, la micro-hétérogénéité fait que l'intensité des ions des glycopeptides est divisée par le nombre de glycoformes. C'est pourquoi il peut être nécessaire d'enrichir le mélange en glycopeptides afin d'en faciliter l'analyse.

Il existe un grand nombre de méthodes permettant de fractionner ou d'enrichir un échantillon en glycopeptides. Comme pour l'enrichissement des glycoprotéines, l'une des plus courantes consiste à utiliser des colonnes d'affinité avec de la lectine qui peut être greffée sur différents supports comme de la sépharose [Alvarez-Manilla et al. 2009], de l'agarose [Bunkenborg et al. 2004] ou de la silice [Xiong et al. 2003]. Cette technique a l'avantage d'être spécifique d'un type de glycan donné [Alvarez-Manilla et al. 2009].

Une autre approche est basée sur l'utilisation des propriétés physico-chimiques des glycopeptides. Ainsi, plusieurs méthodes ont été développées en utilisant la chromatographie d'interaction hydrophile (HILIC) [Calvano et al. 2008]. Les glycopeptides étant plus hydrophiles que les peptides non glycosylés, ils vont être élués plus tard que sur une phase inverse C18. Il est alors possible de récupérer la fraction ou de coupler la chromatographie avec un spectromètre de masse équipé d'une source nano-ESI. Cela permet une analyse directe de l'échantillon enrichi en glycopeptides.

Une autre approche consiste en l'utilisation de micro-colonnes contenant de la poudre de carbone graphite [Larsen et al. 2005]. Cette technique permet de dessaler et de concentrer l'échantillon en glycopeptides. Larsen et al [Larsen et al. 2007] utilisent le même principe de micro-colonne mais avec une phase TiO<sub>2</sub> qui retient spécifiquement les acides sialiques.

Il existe également d'autres matrices comme des billes magnétiques fonctionnalisées avec de l'acide diboronique qui permettent un enrichissement en N- et O-glycopeptides [Sparbier et al. 2005], [Xu et al. 2010]. Les billes fonctionnalisées avec de l'hydrazide nécessitent une étape d'oxydation des glycanes qui augmente la durée expérimentale et la complexité de l'échantillon [H. Zhang et al. 2003, Tian et al. 2007].

Enfin les glycopeptides étant généralement de taille plus importante que les peptides, cette propriété peut être mise à profit en utilisant une colonne d'exclusion stérique pour les collecter [Alvarez-Manilla et al. 2006].

L'analyse des glycopeptides peut donc se faire après une ou plusieurs étapes de purification. Pourtant en pratique, les glycopeptides sont souvent analysés directement à partir du digestat total donc dans un mélange complexe de peptides et de glycopeptides.

### **2.3. Analyse des peptides déglycosylés**

Cette stratégie utilisant la déglycosylation nécessite plusieurs étapes qui comprennent la déglycosylation, le marquage des glycanes et l'analyse par spectrométrie de masse [Mechref et Novotny 2002, Mechref et al. 2005, Geyer et Geyer 2006, Morelle et Michalski 2007, Brooks 2009].

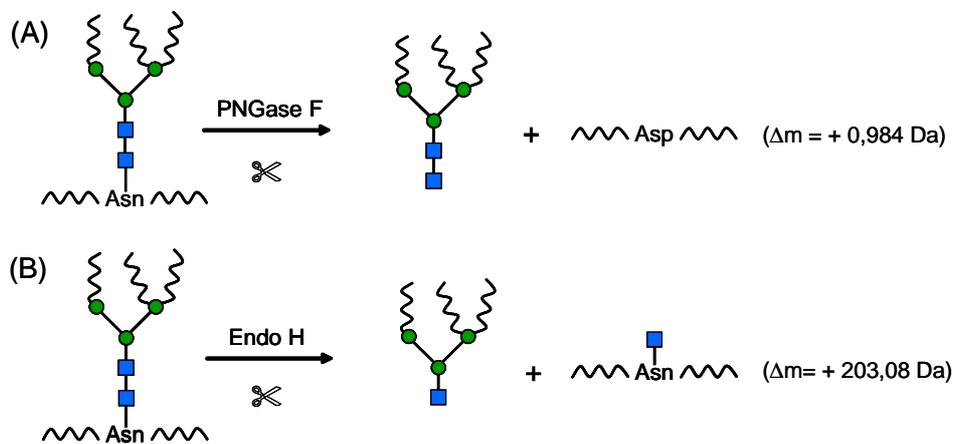
La libération du glycane peut se faire chimiquement ( $\beta$ -élimination en milieu alcalin, hydrazinolise, acide trifluoro-methanesulfobique (TFMS)) ou avec une enzyme [Fryksdale et al. 2002, Morelle et Michalski 2007]. La méthode enzymatique a pour avantage de conserver à la fois le glycane et le peptide intact. En revanche, chaque enzyme est spécifique d'un type de glycosylation, il n'en existe pas d'universelle [O'Neill 1996].

La PNGase F (peptide N-glycosydase F) est l'enzyme la plus couramment utilisée. Elle coupe une grande variété de glycanes à l'exception de ceux portant un fucose en  $\alpha$ 1-3 sur le GlcNAc de l'extrémité réduite [Tretter et al. 1991]. Dans ce cas il est possible d'utiliser la PNGase A. Ce type d'enzyme coupe la liaison entre le GlcNAc et l'asparagine. L'asparagine (Asn) glycosylée est alors

déamidée pour de venir un acide aspartique (Asp) (Figure 6A). La différence de masse entre ces deux acides aminés est de 1 Da ce qui permet d'identifier le site de glycosylation en effectuant une analyse MS/MS du peptide déglycosylé.

Afin de ne pas confondre une asparagine oxydée en acide aspartique avec une déamidation *in vitro* [Robinson et Robinson 2001], la déglycosylation peut être effectuée dans de l'eau lourde  $H_2^{18}O$  [Gonzalez et al. 1992, Küster et Mann 1999, Kaji et al. 2003]. La différence de masse obtenue est alors de 3 Da ce qui permet de lever l'ambiguïté.

Il existe également des endoglycosidases qui coupent la liaison glycosydique entre les deux résidus GlcNAc du tronc commun. La plus couramment utilisées est l'endoglycosydase H (endo H) qui libère les high-mannoses et une grande partie des glycanes de type hybrides. L'avantage de ce type d'enzyme est de laisser un résidu GlcNAc sur l'asparagine (Figure 6B). Le site de glycosylation est ainsi très clairement marqué.



**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

Les glycanes libres peuvent ensuite être marqués au niveau de leur extrémité réduite avec un groupement chromophore, fluorophore ou une fonction ionisable afin d'améliorer leur détection [Geyer et Geyer 2006]. L'un des marquages les plus courants utilise la 2-amino-benzamide ou 2-AB [Bigge et al. 1995].

Une analyse par spectrométrie de masse permet alors d'obtenir une carte glycanique en caractérisant les glycanes et de séquencer les peptides déglycosylés pour trouver les sites de glycosylation [Alvarez-Manilla et al. 2009]. Toutes ces étapes demandent du temps et surtout une quantité de matériel importante. De plus l'information sur la micro-hétérogénéité est perdue.

## 2.4. Analyse des glycopeptides intacts

Le séquençage des peptides par MS/MS est aujourd'hui devenu une technique de routine. Il n'en est pas de même pour les glycopeptides pour lesquels il s'agit d'obtenir des informations de séquence sur deux types de structures oligomériques différentes : le squelette peptidique et le squelette oligosaccharidique, pour lesquels les règles de fragmentation sont différentes. En théorie la MS/MS peut fournir la séquence en acides aminés du glycopeptide, les glycoformes et les sites de glycosylation. Mais en pratique, la MS/MS des glycopeptides présente des difficultés et n'est de loin pas encore une méthode complètement maîtrisée.

En effet, les glycopeptides sont plus difficiles à ioniser par la technique d'électrospray habituellement utilisée que les peptides non glycosylés. De plus, comme les structures glycaniques sont très hétérogènes le signal en spectrométrie de masse est réparti sur de nombreux ions moléculaires dont l'intensité est bien plus faible que celle du peptide correspondant non glycosylé. Par conséquent, ils sont en général, particulièrement difficiles à détecter.

En réponse à ce problème de détection difficile, une approche a été développée par l'équipe de Carr [Huddleston et al. 1993]. Il s'agit de détecter les glycopeptides dans le spectre de masse à l'aide d'ions spécifiques de la fragmentation des glycosylations. Ces ions sont généralement nommés "ions diagnostiques" ou "ions marqueurs".

### 2.4.1. Détection des glycopeptides grâce aux ions diagnostiques

Les ions diagnostiques sont générés par fragmentation préférentielle des liaisons osidiques qui sont plus faibles que les liaisons peptidiques. Il s'agit d'ions oxoniums tels que  $m/z$  204 (HexNAc<sup>+</sup>), 366 (HexHexNAc<sup>+</sup>), 163 (Hex<sup>+</sup>) ou 292 (Neu5Ac<sup>+</sup>) [Huddleston et al. 1993] (Figure 7).

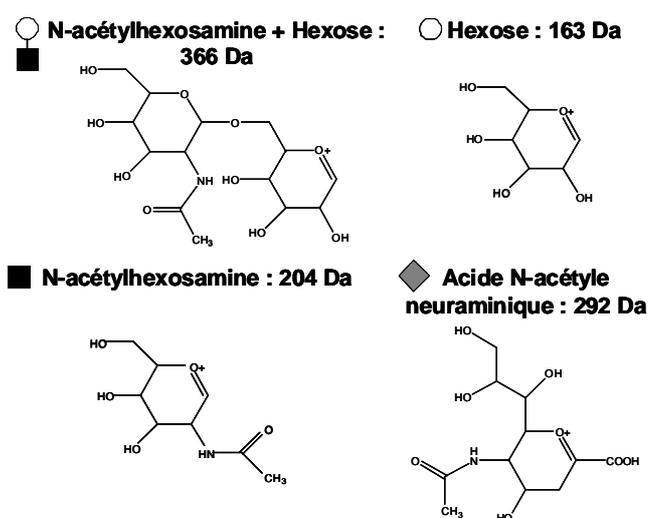
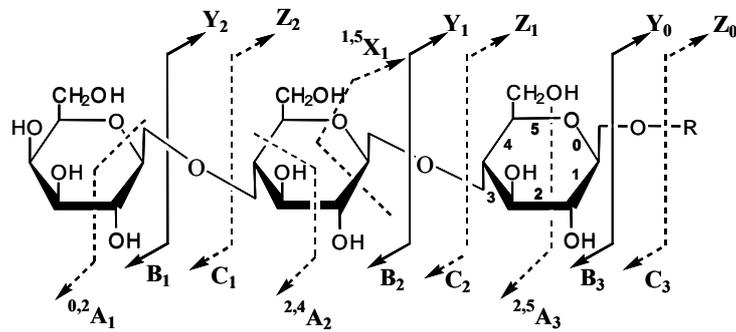


Figure 7. Structures chimiques des ions oxoniums servant d'ions diagnostiques

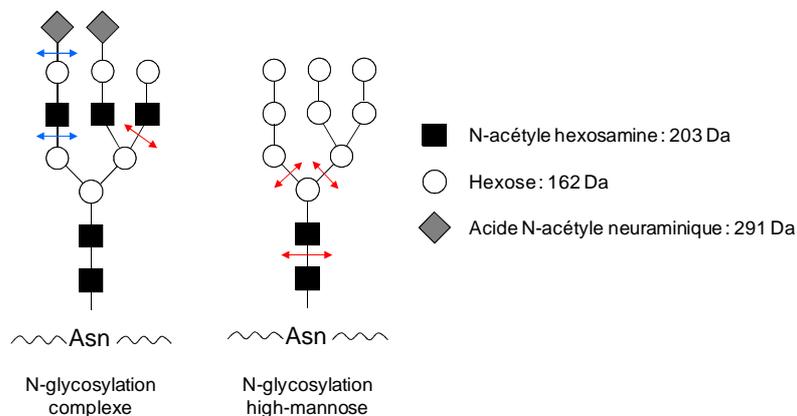
Domon et Costello [Domon et Costello 1988] ont établi une nomenclature pour la fragmentation des sucres décrite dans la Figure 8. Les ions diagnostiques sont de type B et Y.



**Figure 8.** Nomenclature de fragmentation des polysaccharides (Domon et Costello, 1988). Pour les fragments de type Ai et Xi, les numéros en exposants correspondent à des fragmentations intracycliques. [Domon et Costello 1988]

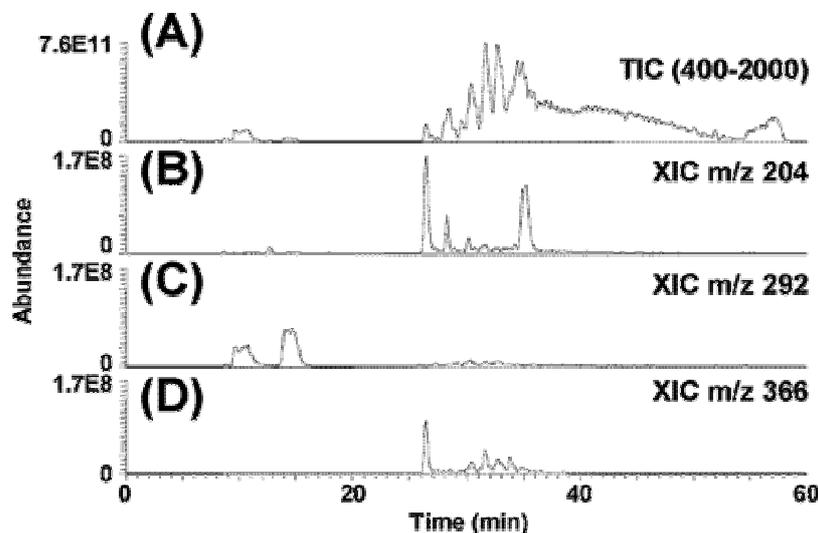
Ces ions peuvent être générés de deux manières : soit par fragmentation dans la cellule de collision [Carr et al. 1993, Huddleston et al. 1993], soit par fragmentation en source [Huddleston et al. 1993, Sullivan et al. 2004, Peterman et Mulholland 2006].

Ils sont issus de la fragmentation interne du glycane et ont donc des masses moléculaires différentes en fonction du type de glycosylation. Par exemple dans le cas des high-mannoses, on observe l'ion  $m/z$  366 (HexHexNAc<sup>+</sup>) dans des proportions moins importantes que dans les cas des glycosylations complexes [Huddleston et al. 1993]. En effet, comme le montre la Figure 9 la génération de cet ion nécessite la coupure de 3 liaisons dans le cas des high-mannoses et seulement d'une ou deux liaisons dans les cas des glycanes complexes. De même, la présence ou non d'ions tels que l'ion  $m/z$  292 (Neu5Ac<sup>+</sup>) ou  $m/z$  657 (Neu5AcHexHexNAc<sup>+</sup>) indique si le glycane porte ou non des acides sialiques (Figure 10) [Sullivan et al. 2004].



**Figure 9.** Liaisons devant être fragmentées pour la génération de l'ion  $m/z$  366

La recherche de ces ions dans un chromatogramme permet de détecter la zone d'éluion des glycopeptides (Figure 10) d'où leur nom de "marqueurs" ou "diagnostiques" [Conboy et Henion 1992, Carr et al. 1993, Medzihradzky et al. 1997]. Il est ainsi possible de localiser les spectres MS/MS correspondants à des glycopeptides dans le chromatogramme [Peterman et Mulholland 2006].



**Figure 10.** (A) LC-MS d'un digestat trypsique d'un anticorps monoclonal dans des conditions optimisées pour la détection des ions diagnostiques. (B) (C) et (D) correspondent aux chromatogrammes d'ions extraits m/z 204, 292 et 366 qui indiquent l'éluion de glycopeptides. L'absence de signal pour l'ion m/z 292 indique que ces glycopeptides ne portent pas d'acide sialique. D'après [Sullivan et al. 2004]

Ces ions diagnostiques vont donc permettre à la fois de détecter les glycopeptides dans le chromatogramme mais aussi d'obtenir une information sur le type de glycosylation.

La littérature montre, en fait, que les ions diagnostiques n'ont que peu été utilisés dans la stratégie de caractérisation des glycoprotéines. Bien que de nombreux travaux décrivent l'utilité des ions diagnostiques, très peu de méthodes permettant de les générer à partir de mélanges complexes et en très petites quantités ont été publiées. Nous pensons que les méthodes instrumentales pour former ces ions doivent encore être optimisées, surtout avec les progrès instrumentaux en termes de sensibilité, de précision de mesure de masse et de résolution. Le chapitre 2 des résultats sera consacré à ce point.

#### 2.4.2. L'analyse MS<sup>n</sup> des glycopeptides

La fragmentation des N-glycopeptides permet de déterminer la composition de chaque glycane 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 :

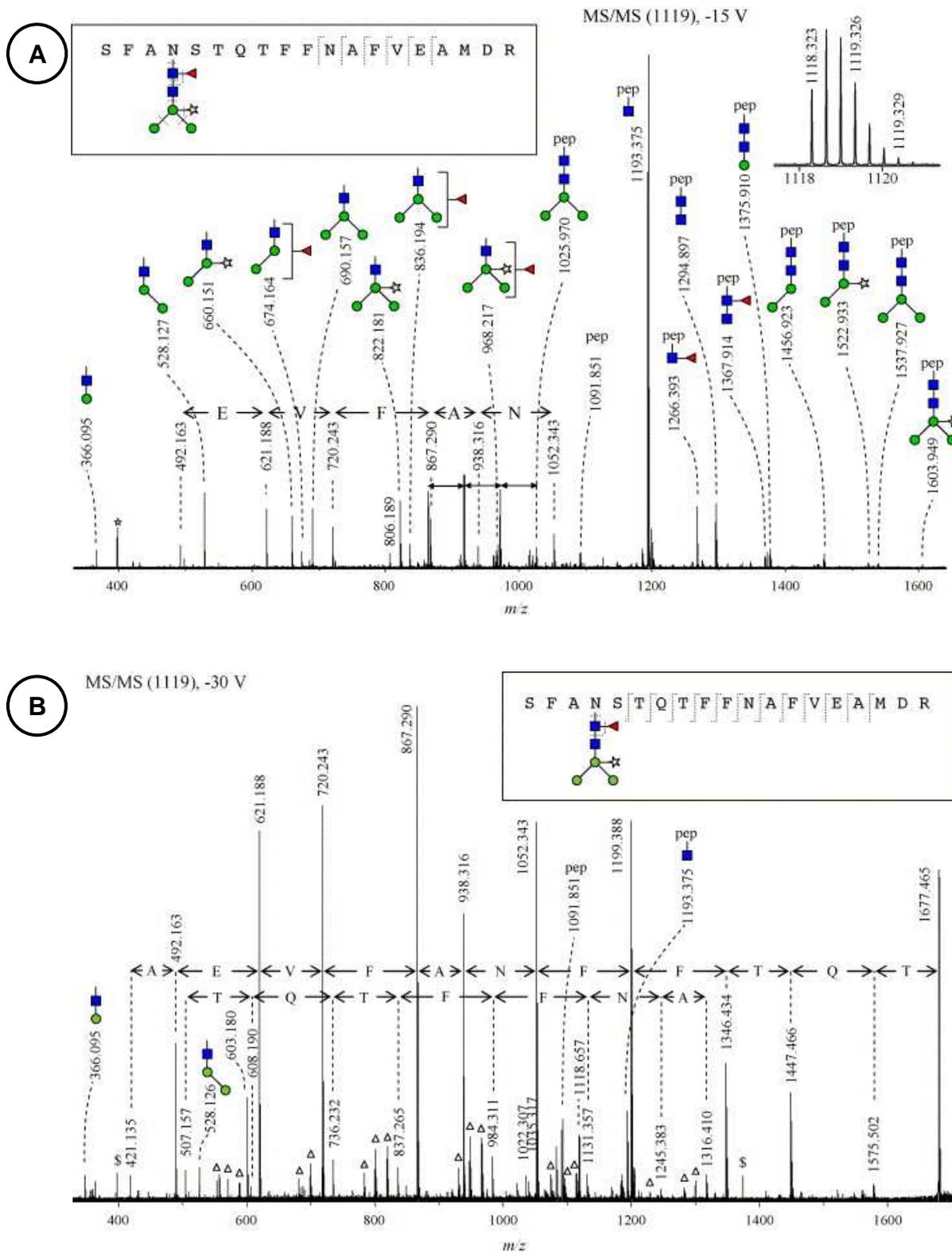
- Des ions fragments correspondant à des pertes de sucres successifs sur la chaîne peptidique : ces ions permettent de déterminer les enchaînements de monosaccharides et ainsi de remonter à la composition de chaque glycanes;
- Des ions oxoniums dits "diagnostiques" de type Bi et Yi qui sont issus de la fragmentation interne du glycanes. Ces ions donnent des informations supplémentaires sur la composition du glycanes mais permettent surtout d'identifier rapidement si un spectre MS/MS correspond à la fragmentation d'un glycopeptide;
- L'ion correspondant au peptide non-glycosylé.

Dans certains cas, il est possible d'observer des ions correspondant à la fragmentation du peptide non glycosylé permettant d'identifier la séquence en acides aminés du peptide (Figure 11A). 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.

Plusieurs articles sur l'analyse de glycopeptides par MS [H. Jiang et al. 2004, Harvey 2005, Wührer et al. 2007], montrent que la fragmentation CID d'un glycopeptide a tendance à varier selon le type de spectromètre de masse, les paramètres instrumentaux, la séquence du glycopeptide, l'état de charge, etc. Par exemple, sur les instruments de type Q-TOF, le type d'ions généré est très dépendant de l'énergie de collision appliquée [Wührer et al. 2007]. La Figure 11 montre les spectres de fragmentation d'un glycopeptide à basse (Figure 11A) et haute énergie (Figure 11B). A basse énergie, les ions observés correspondent majoritairement à la fragmentation du glycanes. A haute énergie, les ions générés correspondent à la fragmentation de la partie peptidique et permettent ainsi d'obtenir la séquence du peptide.

Une autre solution consiste à effectuer de la MS3 en utilisant une trappe ionique [Demelbauer et al. 2004, Wührer et al. 2007, Segu et Mechref 2010]. En MS/MS, l'ion correspondant au peptide avec un GlcNAc lié à l'asparagine est généralement le plus intense. Il est alors possible de sélectionner cet ion pour le fragmenter une seconde fois (MS3) et ainsi déterminer la séquence du peptide [Wührer et al. 2005].

Bien que plusieurs logiciels existent pour identifier et caractériser des glycopeptides à partir d'analyses MS/MS, l'automatisation de l'interprétation de leurs spectres de fragmentation reste compliquée. Tout d'abord, bien qu'il existe certaines banques de données pour les glycanes [Aoki-Kinoshita 2008] il n'y en a aucune pour les glycopeptides. De plus, étant donné que très peu de fragments correspondant au peptide sont observés dans les spectres, il est difficile d'établir la séquence en acides aminés du glycopeptide.

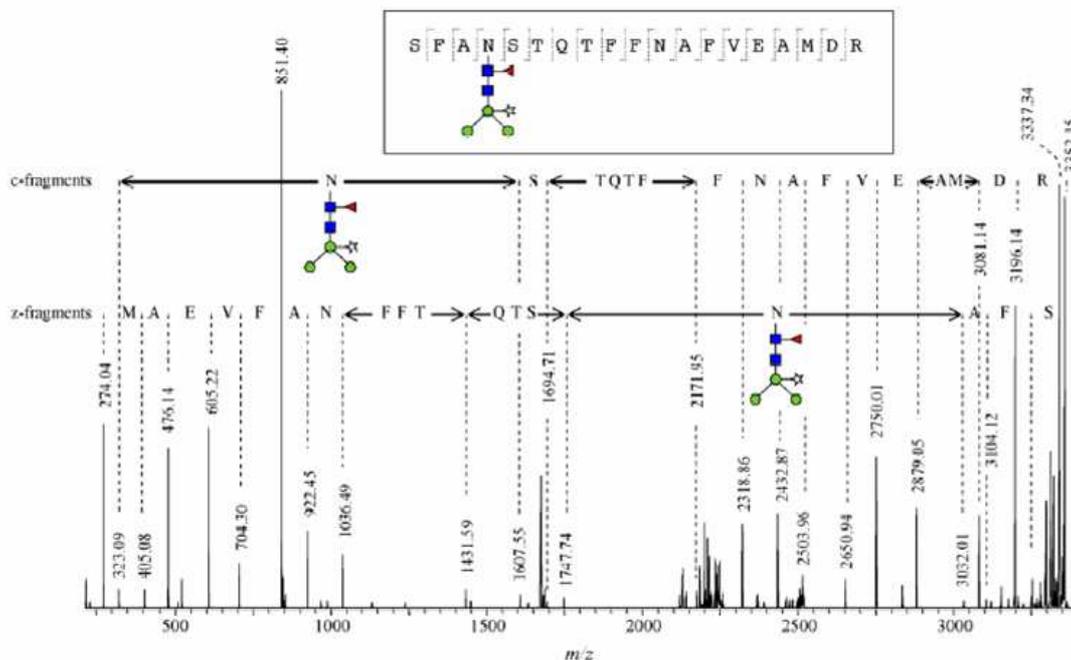


**Figure 11.** Exemple de spectres MS/MS obtenus par CID de l'ion trichargé de  $m/z$  1119 correspondant au glycopeptide tryptique Ser<sup>295</sup>-Arg<sup>313</sup> de la peroxidase de raifort (horseradish peroxidase HRP) obtenu sur un Q-FT-ICR-MS/MS ; (A) Fragmentation à basse énergie (15 V) ; (B) Fragmentation à haute énergie (30 V). D'après [Wuhrer et al. 2007]

C'est pourquoi, beaucoup de logiciels développés nécessitent de connaître la séquence de la glycoprotéine [Goldberg et al. 2007, Irungu et al. 2007, Deshpande et al. 2010]. De plus, afin de limiter le taux de faux positifs, il est souvent recommandé d'utiliser un spectromètre de masse à très haute résolution, type FTICR [Irungu et al. 2007, Wu et al. 2010]. Des problèmes se posent également en ce qui concerne la compatibilité avec les formats des listes de masses. Ainsi, certains logiciels ne sont utilisables qu'avec un certain type d'instruments [Ozohanics et al. 2008]. Enfin certains programmes semblent prometteurs mais ne sont pas utilisables sur des échantillons en très petite quantité [Ozohanics et al. 2008, Wu et al. 2010].

Ainsi, jusqu'à maintenant, il n'existe pas de logiciel permettant d'effectuer une interprétation de spectres MS/MS pour des glycopeptides totalement inconnus et présents dans un mélange en petite quantité.

Plus récemment avec le développement de la source ETD [Syka et al. 2004], une méthode alternative s'est ouverte pour la caractérisation de N-glycopeptides par MSn. En effet, la fragmentation par ETD étant non ergodique elle permet de préserver les liaisons osidiques tandis que le squelette peptidique est fragmenté en générant principalement de ions c et z (Figure 12).



**Figure 12.** Exemple d'un spectre MS/MS obtenu par ETD de l'ion trichargé de  $m/z$  1119 correspondant au glycopeptide tryptique Ser<sup>295</sup>-Arg<sup>313</sup> 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 plus, en combinant la fragmentation alternée ETD /CID il est possible de remonter à la structure du glycane et à la séquence du peptide [Catalina et al. 2007, Alley et al. 2009]. Cependant, l'interprétation des spectres de fragmentation ETD reste très complexe et entièrement manuelle.

### 3. Conclusion

La glycosylation fait partie des modifications post-traductionnelles les plus courantes. Elle possède un rôle majeur dans la fonction biologique des protéines. Cependant, la caractérisation des glycoprotéines reste complexe en grande partie à cause de leur grande hétérogénéité.

Afin de caractériser finement les protéines N-glycosylées, plusieurs critères sont à prendre en compte comme : la quantité de matériel, la taille de la glycoprotéine, son hétérogénéité et quelle est précisément l'information recherchée. Ainsi, la caractérisation de protéines N-glycosylées nécessite la mise en place d'une méthodologie combinant plusieurs techniques aussi bien au niveau de la préparation d'échantillon, de l'acquisition des données MS, que de leur interprétation. Chacune de ces étapes demande une grande expertise et doit être adaptée au contexte de l'étude.

De nombreuses méthodologies existent déjà pour la caractérisation des protéines N-glycosylées. Jusqu'à très récemment, la stratégie la plus courante consistait à déglycosyler les glycopeptides obtenus après digestion de la glycoprotéine par une endoprotéase et d'analyser les glycanes libres et les peptides séparément. Les progrès récents au niveau de la sensibilité et de la résolution de la nouvelle génération de spectromètres de masse tendent aujourd'hui à favoriser l'analyse des glycopeptides intacts par LC-MS/MS. Cette stratégie permet de travailler avec de petites quantités de matériel et de déterminer à la fois la séquence du glycane, le site de glycosylation, mais surtout la micro-hétérogénéité ce qui n'est pas possible lorsque l'on travaille sur les glycanes libres.

Nous avons utilisé dans la suite de ce travail les ions diagnostiques formés à partir de glycopeptides. Ces ions diagnostiques permettent de localiser dans le courant d'ions total les spectres correspondants à des glycopeptides mineurs. La littérature montre en fait que les ions diagnostiques ont été très peu utilisés dans la stratégie de caractérisation des glycoprotéines dans les cas d'échantillons en très faibles quantités (par exemple extrait de bandes de gel 1D ou 2D). C'est pourquoi nous avons consacré un chapitre à l'amélioration de la formation et de la détection de ces ions diagnostiques de ces glycopeptides.

Ainsi, la caractérisation des glycoprotéines qui porte de plus en plus, sur des échantillons en très petites quantités, dans des mélanges complexes et sur des espèces très hétérogènes, nécessite le développement de nouvelles méthodologies adaptées à ce type d'échantillon.

## Chapitre 2

### Outils et stratégies pour l'analyse protéomique : application à l'analyse des glycoprotéines

L'essor de la protéomique a eu lieu grâce à de nombreux développements technologiques dans différentes disciplines académiques qui ont ainsi permis de faire de grands progrès au niveau de :

- L'instrumentation pour l'analyse par spectrométrie de masse;
- La séparation des protéines et des peptides avec la chromatographie liquide ou les gels d'électrophorèse;
- L'interprétation de données avec le développement de nouveaux outils bioinformatiques.

L'ensemble de ces techniques (Figure 1) est utilisé pour l'identification des protéines mais peut également être optimisé pour la caractérisation de modifications post-traductionnelles comme les glycosylations.

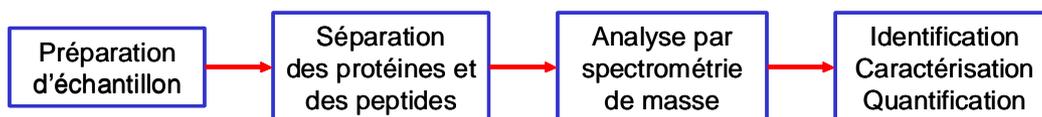


Figure 1. Principales étapes d'une analyse protéomique

Les étapes de séparation d'échantillon, d'analyse par spectrométrie de masse et d'interprétation des données sont décrites dans ce chapitre.

## 1. La spectrométrie de masse : un outil central pour la caractérisation des protéines.

La spectrométrie de masse date du début du siècle dernier avec les travaux de Thompson et Aston [Thompson 1913] mais ne s'est réellement développée pour l'analyse de biomolécules que depuis 20 ans.

En général, un spectromètre de masse est composé de trois parties (Figure 2) : une source d'ionisation qui volatilise et ionise les molécules, un analyseur qui sépare les ions et mesure leur rapport masse sur charge ( $m/z$ ) et un détecteur qui mesure le courant induit par les ions.

Une fois les ions produits dans la source, ils sont transférés jusqu'à l'analyseur à l'aide de gradients de pression et de champs électriques. Dans l'analyseur, les ions sont séparés en fonction de leur rapport  $m/z$ . Plusieurs analyseurs peuvent être combinés, on parle alors de spectrométrie de masse en tandem. Ce type de configuration permet d'obtenir des informations structurales par exemple par fragmentation induite par collision (CID). Dans ce cas, dans un deuxième temps, les fragments générés sont aussi séparés en fonction de leur rapport  $m/z$ .

Finalement, les ions atteignent le détecteur où un signal est généré puis traduit en spectre de masse dans lequel l'intensité du signal est mise en relation avec le rapport  $m/z$ .

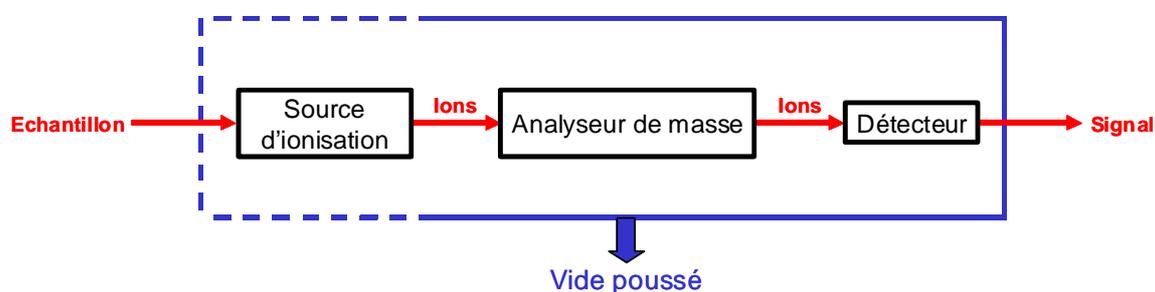


Figure 2. Schéma des différents constituants d'un spectromètre de masse

### 1.1. Les sources d'ionisation

Le succès de la spectrométrie de masse en biologie repose largement sur l'apparition de nouvelles techniques d'ionisation douce à la fin des années 1980 : l'ionisation Electrospray (ESI) [Yamashita et Fenn 1984, Fenn et al. 1989] et le MALDI (Matrix Assisted Laser Desorption Ionisation) [Karas et al. 1987, Tanaka et al. 1988b]. Ces techniques ont rendu possible le transfert de biomolécules intactes en phase gazeuse pour l'analyse par spectrométrie de masse. John Fenn et Koichi Tanaka ont été récompensés par un prix Nobel de chimie en 2002 pour leurs travaux sur ces nouvelles méthodes d'ionisation.

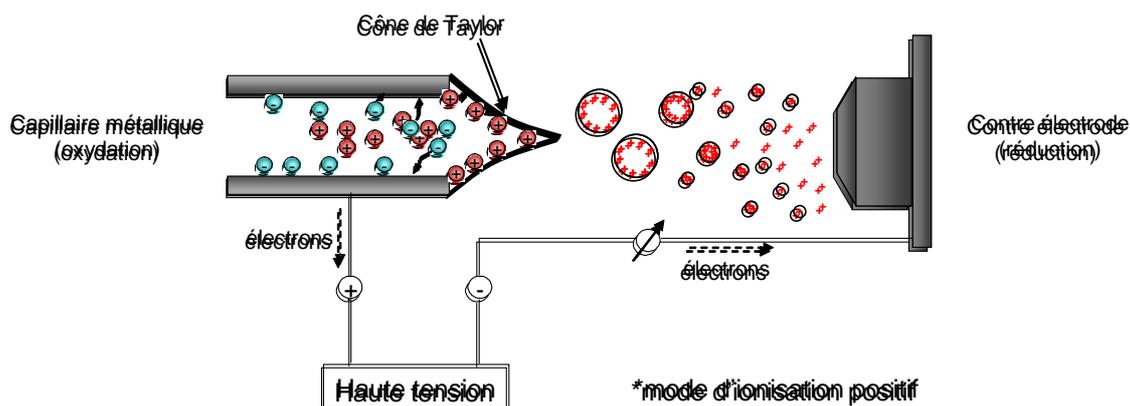
### 1.1.1. La source électrospray (ESI)

#### 1.1.1.1. Processus d'ionisation-désorption ESI

Les expériences de Dole à la fin des années 60 ont suggérées que le mécanisme électrospray pouvait être utilisé pour générer des ions à partir de molécules en solution et ainsi de les analyser par spectrométrie de masse [Dole et al. 1968]. Dans les années 80, l'équipe de Fenn démontre l'utilité de l'ESI qui permet de générer des ions à partir de molécules non volatiles et de haut poids moléculaire sans les fragmenter [Yamashita et Fenn 1984, Meng et al. 1988].

Le processus d'ionisation-désorption ESI s'effectue à pression atmosphérique et se divise en 3 étapes majeures décrites dans la Figure 3 [Kearle et Verkerk 2009] :

- La production des gouttelettes chargées à partir de l'électrolyte en solution;
- La fission des gouttelettes chargées par explosions coulombiennes;
- Le transfert des ions désolvatés en phase gazeuse.



**Figure 3.** L'ionisation électrospray, production et fission des gouttes chargées, transfert des ions en phase gazeuse

Le mode d'ionisation électrospray permet de faire l'interface entre une phase liquide et une phase gazeuse, ce qui le rend parfaitement adapté pour les couplages entre une HPLC et un spectromètre de masse [Whitehouse et al. 1985]. De plus, il génère des ions multichargés, ce qui permet de mesurer des molécules de haut poids moléculaires avec un analyseur possédant une gamme de  $m/z$  limitée.

#### 1.1.1.2. La source nano-électrospray

En mode d'ionisation électrospray, le spectromètre de masse mesure un courant d'ion qui est fonction de la concentration des ions présents dans l'échantillon et pas du débit avec lequel l'échantillon est infusé [Ikononou et al. 1990]. Pour les faibles quantités d'échantillon, ce qui est

souvent le cas pour les échantillons biologiques, afin de gagner en sensibilité, il est préférable de travailler à des débits plus faibles et avec des solutions plus concentrées.

Ces observations ont conduit au développement des sources micro-électrospray (débits 300-800 nL/min) [Emmett et Caprioli 1994] et nano-électrospray (débit ~20 nL/min) [Wilm et Mann 1996, Wilm et al. 1996] qui utilisent des capillaires de diamètre interne de quelques  $\mu\text{m}$  (~100  $\mu\text{m}$  pour le mode électrospray) et qui ne nécessitent plus d'assistance gazeuse. L'utilisation des débits plus faibles conduit à l'émission de gouttelettes plus petites (environ 10 fois) possédant une plus grande densité de charge et donc à une augmentation du rendement d'ionisation-désorption des analytes [Wilm et Mann 1996, Karas et al. 2000].

Ces sources miniaturisées permettent de réduire la consommation d'échantillon tout en améliorant la sensibilité, elles sont donc particulièrement bien adaptées pour les analyses protéomiques.

### **1.1.1.3. L'ionisation des glycoprotéines et des glycopeptides**

La première étude utilisant l'électrospray pour la caractérisation d'une glycoprotéine recombinante est publiée par [Ling et al. 1991]. Depuis comme pour la protéomique l'utilisation de ce type de source s'est très largement répandue. Ce mode d'ionisation doux est en effet particulièrement adapté pour l'analyse de groupements labiles comme les glycosylations. De plus grâce à la génération d'ions multichargés, il est possible de mesurer des molécules de hauts poids moléculaires avec un analyseur possédant une gamme de masse réduite. Pourtant l'analyse de protéines glycosylées reste un déficit [Temporini et al. 2008]. Les glycosylations posent un certain nombre d'obstacles à l'ionisation par électrospray [Mechref et Novotny 2002] :

- La présence de glycosylation peut rendre le phénomène de désolvation moins efficace;
- Certains sites potentiels de protonation peuvent être masqués par les glycosylations qui sont des groupements provoquant un encombrement stérique important. L'ionisation de la glycoprotéine se fera donc moins facilement;
- Les glycosylations ont une forte affinité pour les sels. Il n'est donc pas rare d'observer des adduits qui compliquent l'interprétation des spectres.

De plus, l'intensité de la glycoprotéine est divisée entre les différentes glycoformes mais aussi entre les états multichargés induits par le mode d'ionisation électrospray. Il est donc difficile d'analyser des glycoprotéines entières par électrospray. La source nanospray permet de palier à certains de ces problèmes [Wilm et Mann 1996]. En effet, les gouttelettes formées étant plus petites, la désolvation est plus efficace et permet donc d'obtenir un meilleur signal. Enfin, les développements instrumentaux récents, principalement au niveau des analyseurs, améliorent grandement l'analyse des glycoprotéines [Damen et al. 2009].

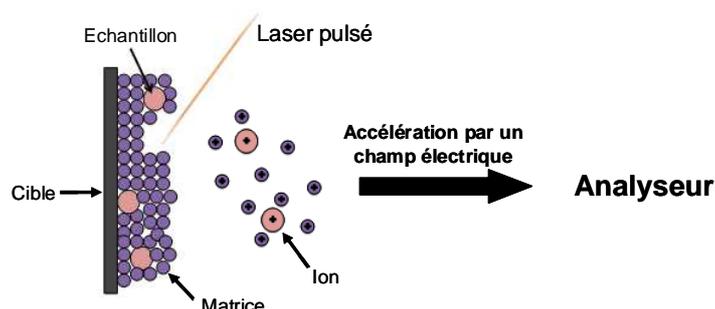
L'ensemble de ces contraintes reste valable pour l'analyse des glycopeptides, mais dans une moindre mesure. En effet, les glycopeptides ne comportent généralement qu'un seul site de glycosylation et donc l'effet de l'hétérogénéité est moins important, de même pour le nombre de formes multichargées qui est plus restreint [Mechref et Novotny 2002, Temporini et al. 2008].

### 1.1.2. MALDI

Les travaux sur la source MALDI date des années 80 [Karas et Hillenkamp 1988, Tanaka et al. 1988a], où deux groupes allemands et japonais montrent son intérêt pour l'analyse de composés de hauts poids moléculaires.

La technique MALDI utilise un faisceau laser pulsé travaillant généralement dans l'UV, pour désorber et ioniser un mélange matrice/échantillon co-cristallisé sur une surface métallique. Les matrices les plus couramment utilisées en protéomique sont l'acide  $\alpha$ -cyano-4-hydroxycinnamique et l'acide 2,5-dihydrobenzoïque (DHB). La matrice DHB est également utilisée pour l'analyse des glycosylations avec l'acide sinapinique et le 2,4,6-trihydroxyacetophenone (THAP) [Gimenez et al. 2007]. La qualité des spectres dépend de plusieurs facteurs liés à la préparation de l'échantillon comme le choix de la matrice, la méthode de dépôt pour la co-cristallisation matrice-analyte ou le pH des solutions.

Le processus d'ionisation par MALDI s'effectue sous vide ( $\approx 10^{-7}$  mbar). L'énergie transmise par le laser est absorbée par la matrice et cet apport d'énergie provoque l'éjection des molécules d'échantillon et de matrice en phase gazeuse (Figure 4). L'échantillon est alors ionisé par transfert de protons [Zenobi et Knochenmuss 1998]. Les ions générés sont majoritairement monochargés. Les spectres obtenus sont donc moins complexes qu'en ESI.



**Figure 4.** La source MALDI

La source MALDI est généralement couplée à un analyseur en temps de vol (TOF pour Time-Of-Flight) qui est bien adapté à l'ionisation pulsée par désorption laser ainsi qu'à l'analyse de composés de haut poids moléculaire. En effet, le processus MALDI est peu dépendant des propriétés de désorption et de la taille des composés à analyser, il permet donc l'ionisation d'analytes de très haut poids moléculaire (>100000 Da).

L'ensemble de ces propriétés rend la source MALDI bien adaptée pour l'analyse de glycoprotéines peu complexes. Elle est plus particulièrement utilisée pour l'analyse d'oligosaccharides libres [Wada et al. 2007].

## 1.2. Les analyseurs

Les résultats obtenus lors d'une analyse protéomique sont principalement conditionnés par les performances de l'analyseur de masse utilisé. Les paramètres qui permettent de juger ces performances sont :

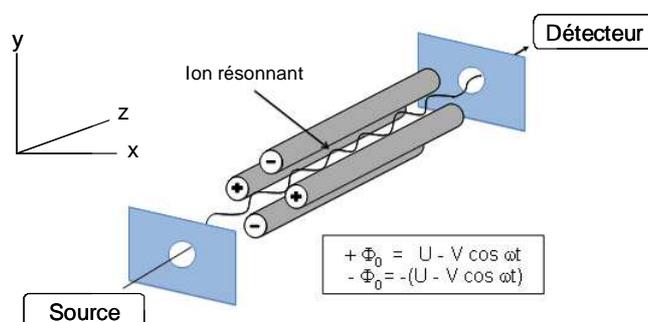
- La résolution
- La précision de mesure de masse
- La sensibilité
- La gamme de masse
- La vitesse de balayage
- La capacité à effectuer des analyses en tandem

Au cours de ce travail de thèse, deux types d'instruments ont principalement été utilisés : des trappes ioniques (HCT Ultra et AmaZon, Bruker Daltonique) et des Q-TOF (Synapt, Waters; MaXis et MicrOTOF-Q, Bruker Daltonique). C'est pourquoi seuls ces types d'analyseurs seront décrits.

### 1.2.1. Le quadripôle

Le quadripôle a été développé par Paul dans les années 60 [Paul et Steinwedel 1960]. Aujourd'hui les quadripôles servent d'analyseurs mais aussi de guide d'ions ou de filtres de masse par exemple [Douglas 2009]. Il est généralement associé à une source ESI et peut être couplé avec d'autres quadripôles afin de former un triple-quadripôle, ou un TOF, pour former un Q-TOF. Ces instruments hybrides permettent d'effectuer des analyses en tandem.

Un analyseur quadripolaire est constitué de quatre électrodes parallèles connectées entre elles afin d'établir un champ électrique en opposition de phase entre deux barreaux adjacents (Figure 5). La trajectoire des ions reste stable pour un courant donné et un rapport masse/charge ( $m/z$ ) donné. Ainsi les ions peuvent être isolés en fonction de leur masse. Un balayage en amplitude du champ électrique permet à des ions possédant des  $m/z$  différents de tous ressortir au bout du quadripôle. Le principe de fonctionnement de l'analyseur quadripolaire a été décrit par Campana [Campana 1980].

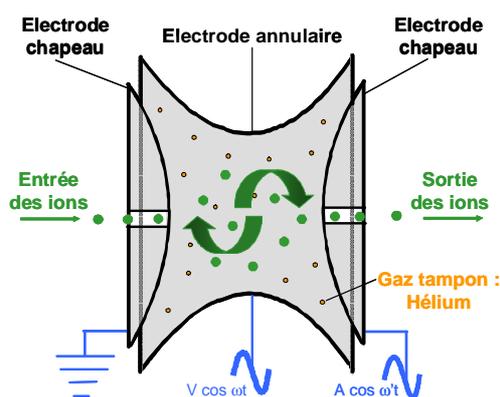


**Figure 5.** Schéma d'un analyseur quadripolaire. La trajectoire des ions se fait selon l'axe z [Campana 1980]

### 1.2.2. La trappe ionique

Le principe de la trappe ionique est issu du même brevet que celui du quadripole déposé par Paul dans les années 60 [Paul et Steinwedel 1960]. Par la suite, Stafford et Finnigan y ont apporté de nombreux développements technologiques [Stafford Jr et al. 1984] pour aboutir à la trappe ionique qui a été commercialisée en 1983 comme détecteur de chromatographie gazeuse. Depuis de nombreux groupes ont apporté leur contribution à l'amélioration de cet analyseur en matière de sensibilité, de résolution et de gamme de masse [March 2009].

La trappe ionique (Figure 6) est constituée de deux électrodes quasi hyperboliques dites électrodes chapeaux, et d'une électrode annulaire. Cet assemblage définit un espace d'environ  $1 \text{ cm}^3$  sous un vide d'environ  $10^{-6} \text{ mbar}$ . Les ions circulent dans cet espace en empruntant les orifices situés au centre de chaque électrode chapeau. Une pression résiduelle d'hélium (environ  $5.10^{-3} \text{ mbar}$ ) règne au sein de l'analyseur permettant d'une part d'augmenter l'efficacité de piégeage des ions et d'autre part d'assurer le rôle de gaz de collision.



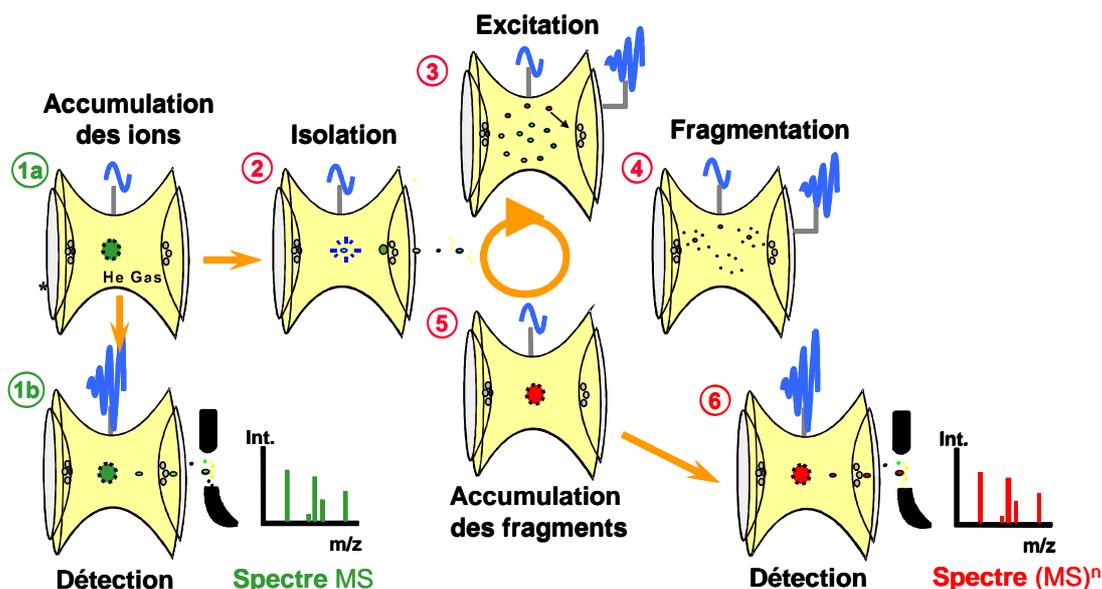
**Figure 6.** Représentation schématique d'une trappe ionique.  $V$  = amplitude de la radiofréquence appliquée sur l'électrode annulaire;  $A$  = amplitude de la tension sinusoïdale appliquée sur l'électrode chapeau de sortie;  $\omega$  et  $\omega'$  = fréquences d'oscillation des radiofréquences

Les différentes étapes de piégeage, éjection, isolation, fragmentation sont décrites dans la Figure 7.

#### ➤ Piégeage des ions (Figure 7-1a et 5)

Lors de l'injection des ions dans la trappe, les électrodes chapeaux sont à la terre ( $A = 0$ ) tandis que l'électrode annulaire est soumise à une tension oscillante qui permet le piégeage des ions en créant un champ quadripolaire dans la trappe ionique. Afin d'éviter que les ions ne ressortent directement, ils sont thermalisés par les molécules d'hélium. Les ions se trouvent alors piégés dans un puits de potentiel dont la profondeur est proportionnelle à l'amplitude de la tension appliquée ( $V$ ). Ils

adoptent une trajectoire stable régie par les équations de Mathieu. Cette trajectoire prend la forme d'un huit dans l'espace en trois dimensions (courbe de Lissajou).



**Figure 7.** Principe d'isolation, de fragmentation et de détection d'ions dans une trappe d'ions pour les expériences de MS et MSn

L'amplitude de la radiofréquence ( $V$ ) détermine la profondeur du puits de potentiel. Plus l'amplitude  $V$  augmente, plus le puits de potentiel est profond et meilleur est le piégeage des ions. D'autre part pour une amplitude  $V$  constante, les ions de différents  $m/z$  ne subissent pas le même potentiel de piégeage. Ainsi, une grande amplitude ( $V$ ) de la radiofréquence sera favorable au piégeage des ions de hauts  $m/z$ , mais inadaptée au piégeage des ions de faible  $m/z$ . Le choix de l'amplitude de la radiofréquence détermine donc la gamme de  $m/z$  d'analyse. Pour une amplitude  $V$  donnée, on définit le "cut-off" comme étant la valeur du plus petit  $m/z$  pouvant être piégé dans l'analyseur. Tous les ions de  $m/z$  inférieur au "cut-off" adoptent une trajectoire instable et sont défectés sur les électrodes de la trappe. Le changement de polarité des électrodes n'est pas suffisamment rapide pour éviter la collision des ions avec les électrodes.

➤ L'éjection des ions (Figure 7-1b et 6)

L'éjection des ions peut s'effectuer de deux manières :

- Une augmentation de l'amplitude  $V$  de la radiofréquence sur l'électrode annulaire va permettre d'éjecter les ions par ordre de  $m/z$  croissants. Il s'agit dans ce cas d'une éjection simple. Cette technique est limitée par la valeur des tensions applicables (20 kV) et ne permet pas d'éjecter des ions de plus de  $m/z$  650.
- En appliquant une tension alternative ( $A \cos \omega t$ ) sur l'électrode chapeau de sortie, les ions soumis à ce potentiel d'amplitude croissante vont successivement être déstabilisés et être éjectés. On parle alors d'éjection résonante. Cette approche permet d'augmenter la gamme de  $m/z$  d'analyse.

➤ Isolément des ions précurseurs pour les analyses MSn (Figure 7-2)

L'isolement résulte de deux étapes successives.

- Dans un premier temps les ions de  $m/z$  plus faibles que les ions à isoler sont éjectés en procédant à un balayage croissant de l'amplitude  $V$  de la radiofréquence de l'électrode annulaire jusqu'à atteindre  $V_{\max}$  (amplitude limite pour laquelle le précurseur n'est pas éjecté).
- Ensuite, les ions de  $m/z$  plus élevés que le précurseur sont mis en résonance et éjectés par un balayage de fréquence en partant d'une fréquence proche de la fréquence des ions précurseurs à isoler.

➤ Excitation et fragmentation des ions précurseurs (Figure 7-3 et 4)

La fragmentation est provoquée par l'ajustement de la fréquence de la tension alternative sur l'électrode chapeau de sortie de manière à ce qu'elle coïncide avec la fréquence propre de résonance des ions précurseurs. Ainsi dopé en énergie cinétique, ils rentrent en collision avec les molécules d'hélium. Au fur et à mesure des collisions, leur énergie cinétique est convertie en énergie interne et l'ion parent finit par se fragmenter. L'étape de fragmentation est suffisamment longue pour permettre aux ions d'emprunter les voies de fragmentation les plus stables énergétiquement. De plus, les ions fils issus de la fragmentation ne peuvent pas se fragmenter simultanément en ions petits-fils car l'énergie qui est transmise aux ions précurseurs n'est délivrée qu'à leur seule fréquence de résonance.

➤ Séquence d'analyse de peptides en MS/MS

Pour les analyses protéomiques, les différentes opérations décrites précédemment, sont réalisées successivement. Ainsi tout au long de l'élution chromatographique des différents peptides analysés, des cycles (MS + MS/MS) sont alternés selon la séquence suivante :

- |  |   |              |
|--|---|--------------|
| <ul style="list-style-type: none"> <li>- Piégeage des ions dans la trappe</li> <li>- Ejection des ions</li> </ul>  | } | <b>MS</b>    |
| Sélection des ions d'intérêt   |   |              |
| <ul style="list-style-type: none"> <li>- Piégeage des ions</li> <li>- Isolations des ions d'intérêt</li> <li>- Excitation et fragmentations des ions isolés</li> <li>- Ejection des ions fragments.</li> </ul> | } | <b>MS/MS</b> |

Les étapes d'isolement et de fragmentation peuvent en théorie être répétées à l'infini. La seule limitation étant la quantité d'ions que peut contenir la trappe. Il est ainsi possible de réaliser des analyses MS<sup>n</sup>.

La trappe ionique présente une grande capacité de piégeage des ions et donc une grande sensibilité. Elle possède de plus une grande vitesse de balayage et permet d'effectuer des cycles MS-MS/MS dans des temps très réduits (~ 1 s/cycle). Cette caractéristique en fait un instrument

particulièrement bien adapté pour les analyses en couplage nanoLC-MS/MS réalisées en protéomique.

En revanche, les trappes ioniques pêchent par leur résolution limitée (environ 3000) et leur faible justesse de mesure de masse. De plus elles possèdent une faible dynamique au niveau de leur gamme de masse ce qui pose un problème pour l'analyse de molécules de hauts poids moléculaires comme les glycopeptides. Le "cut-off" est également une limitation importante pour l'analyse des glycopeptides bien qu'il puisse être réglé à 17% sur des trappes récentes. Les ions diagnostiques (cf. Chapitre 1, 2.6.1.) étant des ions de petits  $m/z$ , ils sont difficilement détectables sur ce type d'instrument. En effet, l'efficacité du piégeage ne peut pas être optimal à la fois pour les ions de rapport  $m/z$  faible et pour les ions de rapport  $m/z$  élevé. Il est donc nécessaire d'effectuer un compromis.

### **1.2.3. Le Q-TOF : un analyseur en tandem**

Cet appareil est constitué de deux analyseurs, un premier de type quadripôle suivi d'un tube de vol ou analyseur temps de vol ("Time Of Flight", TOF), les deux étant séparés par une cellule de collision. Dans cette configuration le quadripôle peut fonctionner soit comme filtre, soit comme guide d'ions.

Un analyseur temps de vol établit le rapport  $m/z$  des ions en mesurant le temps nécessaire à chaque ion pour parcourir toute la longueur du tube de vol et atteindre le détecteur. Les ions entrent tous dans le tube de vol libre de champ avec la même énergie cinétique ( $zeE$ ). Étant donné que cette dernière est égale à  $\frac{1}{2}mv^2$ , la vitesse des ions est inversement proportionnelle à la racine carrée de leur masse et proportionnelle à la racine carrée de leur charge. Par conséquent, des ions de rapport  $m/z$  différent auront des temps de vol différents. Un spectre de masse est généré après avoir mesuré le temps de vol de chaque ion. Généralement, un réflectron permettant de refocaliser les ions de même rapport  $m/z$  est utilisé pour améliorer la résolution de l'instrument.

Lorsque le quadripôle fonctionne comme un guide d'ions ("RF-only"), les ions sont analysés par l'analyseur TOF. Cependant, la transmission des ions dans le quadripôle est dépendante de l'amplitude de la radiofréquence qui doit être réglée pour transmettre les ions dans la gamme de masse souhaitée.

Pour faire de la MS/MS, le quadripôle est utilisé en tant que filtre à ions. Il sélectionne les ions précurseurs qui sont ensuite fragmentés dans la cellule de collision. Les fragments générés sont analysés par l'analyseur TOF. On parle alors de spectromètre de masse en tandem.

Dans le cadre de l'analyse protéomique, les Q-TOF ont plusieurs atouts. Ils possèdent tout d'abord une excellente résolution (en moyenne entre 8000 et 20000 en fonction de l'instrument). Ils peuvent également être très précis en mesure de masse particulièrement quand le système est équipé d'un système d'étalonnage interne. Enfin, les analyseurs TOF sont en théorie illimités en gamme de masse. C'est donc le quadripôle qui est limitant mais les Q-TOF possèdent tout de même une bien meilleure dynamique au niveau de leur gamme de masse que les trappes ioniques.

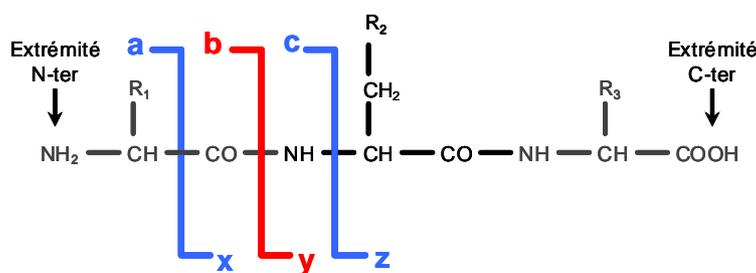
Les Q-TOF sont particulièrement bien adaptés à l'analyse des glycopeptides. En effet la bonne dynamique au niveau de la gamme de masse permet de voir correctement des ions lourds comme des glycopeptides et des ions légers comme des ions oxoniums issus de la fragmentation des glycosylations. De plus, la résolution et la précision de masse permettent d'obtenir des masses monoisotopiques pour des ions lourds multichargés comme le sont généralement les glycopeptides.

### 1.3. La fragmentation peptidique

Dans le cadre des études protéomiques, les analyses MS/MS effectuées avec des trappes ioniques ou des Q-TOF, permettent de générer des spectres de fragmentation de peptides. La fragmentation peptidique permet d'obtenir des informations de séquence et d'avoir ainsi une grande spécificité d'identification en analyse protéomique. Les spectres de fragmentation sont également très informatifs en ce qui concerne les modifications post-traductionnelles comme les glycosylations. Cependant, l'identification des séquences peptidiques à partir de données MS/MS demande une connaissance des mécanismes de fragmentation des peptides en phase gazeuse.

#### 1.3.1. Nomenclature des fragments peptidiques

La nomenclature acceptée pour les ions fragments fut proposée pour la première fois par Roepstorff [Roepstorff et Fohlman 1984] puis modifiée par Johnson et Biemann [R. S. Johnson et al. 1987, Biemann 1988]. La nomenclature proposée par Biemann reste toujours celle utilisée officiellement (Figure 8).



Ions de série **a**, **b** et **c** : charge positive portée par la partie N-terminal

Ions de série **x**, **y** et **z** : charge positive portée par la partie C-terminal

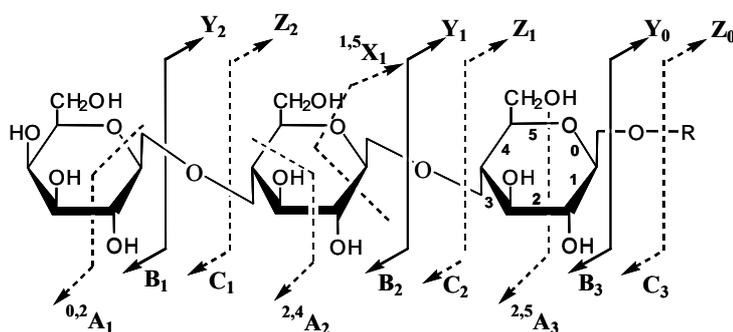
**Figure 8.** Nomenclature de fragmentation de Biemann [Biemann 1988] pour un peptide de 3 acides aminés (R1, R2, R3)

Deux types d'ions sont observés pour les fragmentations à basse énergie :

- Les ions pour lesquels la charge positive est portée par la partie contenant l'acide aminé N-terminal : les ions des séries a, b, et c;
- Les ions pour lesquels la charge positive est portée par la partie contenant l'acide aminé C-terminal : les ions des séries x, y et z.

Les différences de masse entre des ions consécutifs d'une même série permettent de déterminer l'identité des acides aminés consécutifs et donc de déduire la séquence du peptide à l'exception du couple d'acides aminés isobares Leucine/Isoleucine.

De la même manière Domon et Costello [Domon et Costello 1988] ont établi une nomenclature pour la fragmentation des oligosaccharides (Figure 9), qui permet d'interpréter des spectres de fragmentation de glycanes mais aussi de glycopeptides. Suivant le même principe que pour les acides aminés, les différences de masse entre des ions consécutifs d'une même série permettent d'identifier les sucres d'un glycane (hexose, N-acetylhexosamine, deoxyhexose, etc).



**Figure 9.** Nomenclature de fragmentation des polysaccharides. Pour les fragments de type Ai et Xi, les numéros en exposants correspondent aux fragmentations intracycliques [Domon et Costello 1988]

### 1.3.2. La fragmentation induite par collision (CID) : le modèle du proton mobile

Le transfert des ions en phase gazeuse par ESI n'est pas un processus hautement énergétique, les ions entrant dans le spectromètre de masse ont donc des énergies internes faibles [Gaskell 1997]. L'énergie acquise par les ions pour être dissociés est la plupart du temps apportée par activation par collisions avec des molécules de gaz non chargées (Collision Induced Dissociation, CID).

Le transfert d'énergie par activation collisionnelle implique une délocalisation du proton à différents endroits du squelette peptidique. La présence de ce proton initie une fragmentation au niveau des liaisons amides ou liaisons peptidiques pour former des ions b et y. On dit que ces fragmentations basses énergies sont dirigées par la charge [Cox et al. 1996] et suivent le modèle du proton mobile [Dongre et al. 1996].

Pour des peptides monochargés, la présence d'acides aminés basiques (Arginine, Lysine) augmente considérablement l'énergie nécessaire à leur fragmentation : le proton est "séquestré" sur la chaîne latérale de l'acide aminé basique et l'énergie requise pour le délocaliser sur le squelette peptidique est plus élevée.

Pour des peptides tryptiques doublement chargés (espèces majoritaires à 65-80% dans le cas de peptides tryptiques [Savitski et al. 2005]), les deux charges sont généralement portées par l'extrémité N-terminale et par la chaîne latérale de la lysine ou de l'arginine C-terminale. La barrière énergétique pour transférer le proton N-terminal sur le squelette peptidique est plus faible (par rapport

au proton séquestré par l'acide aminé basique en C-terminal) ce qui favorise la génération d'ions b et y. Les ions y sont souvent plus intenses car le proton séquestré par le résidu C-terminal augmente leur stabilité par rapport aux ions b.

Dans les approches mettant en œuvre des digestions avec d'autres enzymes que la trypsine (AspN, GluC par exemple), les peptides - non tryptiques - subissent des fragmentations moins régulières car ils ne portent pas systématiquement des acides aminés N- et C-terminaux basiques. Les spectres de fragmentation sont dans ce cas bien plus difficiles à interpréter [Paizs et Suhai 2005, Barton et Whittaker 2009].

En CID, la fragmentation s'effectue d'abord au niveau des liaisons les plus faibles. Dans le cas des glycopeptides, les liaisons osidiques sont plus faibles que les liaisons peptidiques. Le glycane fragmente donc préférentiellement au détriment du peptide (Chapitre 1).

### **1.3.3. La dissociation par transfert d'électron (ETD)**

Il existe un autre type de fragmentation, la dissociation par transfert d'électron (Electron Transfert Dissociation, ETD) qui a été développée par l'équipe de Hunt [Syka et al. 2004]. Cette méthode de fragmentation met en jeu des mécanismes comparables à ceux de l'Electron Capture Dissociation (ECD) [Zubarev et al. 1998].

Le principe de l'ETD consiste à transférer un électron sur le peptide protoné par l'intermédiaire d'un réactif (radical anion) et d'engendrer ainsi sa fragmentation.

Ce processus de fragmentation ne suit pas le chemin énergétique préférentiel pour la dissociation des molécules. Il n'y a pas de redistribution d'énergie sur le peptide, la fragmentation est donc "statistique" et ne dépend pas de la séquence en acides aminés contrairement à la fragmentation CID. Les ions résultants de cette fragmentation sont de type c et z.

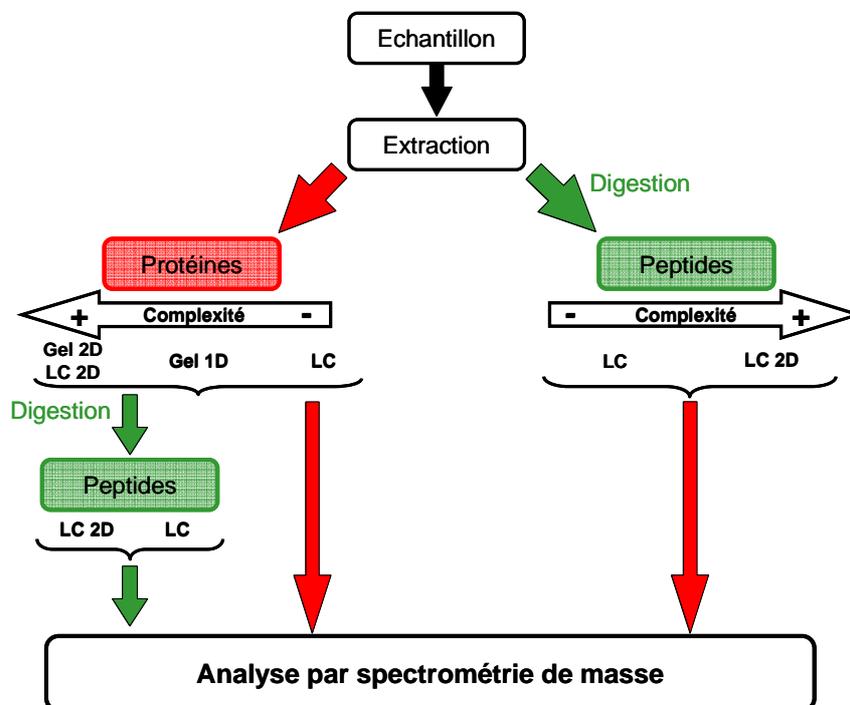
Cette fragmentation statistique permet de conserver les modifications post-traductionnelles comme les glycosylations. La fragmentation ETD permet alors d'obtenir des informations à la fois sur la séquence en acides aminés et en sucre [Alley et al. 2009]. Pourtant, même si ces spectres sont très informatifs, ils sont également assez complexes à interpréter.

## 2. Les stratégies protéomiques basées sur les techniques séparatives

La spectrométrie de masse ne peut, à elle seule, permettre la caractérisation directe de mélanges complexes de protéines. En effet, ces protéines peuvent être présentes sous de multiples isoformes et porter des modifications post-traductionnelles. De plus leurs concentrations peuvent s'étendre sur de larges gammes dynamiques, souvent au-delà de celles qui peuvent être appréhendées en spectrométrie de masse. Un fractionnement des échantillons avant leur analyse par spectrométrie de masse est alors requis pour que leur complexité soit compatible avec les performances des instruments utilisés.

La Figure 10 décrit les techniques séparatives les plus utilisées en analyse protéomique. Il existe deux approches :

- Les protéines sont séparées à l'aide de différentes techniques puis digérées pour obtenir des peptides qui peuvent eux-mêmes être séparés puis analysés par spectrométrie de masse;
- La digestion directe des protéines (sans séparation préalable) pour obtenir des peptides qui sont séparés puis analysés par spectrométrie de masse.



**Figure 10.** Les techniques séparatives les plus utilisées en analyse protéomique par spectrométrie de masse

## 2.1. Séparation des protéines

### 2.1.1. L'électrophorèse sur gel bidimensionnel (Gel 2D)

Introduit dans les années 70 [Kenrick et Margolis 1970, O'Farrell 1975] et largement améliorée depuis [Rabilloud et al. 2009], le gel 2D fait partie des outils les plus couramment utilisés en analyse protéomique. Il peut permettre de séparer plusieurs milliers de protéines et de détecter et de quantifier des protéines à moins de 1ng par spot [Görg et al. 2004].

Cette technique consiste à séparer les protéines, contenues dans un échantillon complexe, en fonction de leur point isoélectrique (pI) selon une première dimension (iso-électrofocalisation, IEF), puis en fonction de leur masse moléculaire selon une deuxième dimension (électrophorèse SDS-PAGE). Après coloration du gel, les spots d'intérêt sont découpés, et les protéines sont digérées avec une endoprotéase. Les peptides issus de la digestion sont ensuite analysés par MALDI-TOF ou par nanoLC-MS/MS afin d'obtenir des informations de séquence et identifier la ou les protéines contenues dans les spots.

Bien que très utilisée cette technique possède quelques limitations. Le grand nombre de spots générés rend difficile la découpe systématique des gels 2D. Généralement seuls les spots les plus intenses sont donc analysés. De plus, la gamme dynamique visible sur le gel est restreinte, et certaines catégories de protéines (membranaires, pI extrêmes) sont exclues du gel [Braun et al. 2007].

Pourtant le gel 2D reste une méthode de choix pour l'analyse protéomique car il est très résolutif et donne des indications quantitatives. Il permet également de séparer et de visualiser un grand nombre de variants post-traductionnels comme les glycosylations par exemple [Görg et al. 2004].

### 2.1.2. L'électrophorèse sur gel monodimensionnel (Gel 1D ou SDS-PAGE)

Le gel SDS (Sodium DodecylSulfate)-PAGE (PolyAcrylamide Gel Electrophoresis) permet de séparer les protéines en fonction de leurs masses apparentes [Rabilloud et al. 2009]. Contrairement au gel 2D, cette technique permet de séparer tous types de protéines quels que soient leurs propriétés physico-chimiques. Cependant, son pouvoir résolutif reste relativement faible. Sur un extrait biologique complexe plusieurs centaines de protéines peuvent co-migrer dans la même zone du gel.

Mais cette technique est facile et rapide à mettre en œuvre et peut servir de préfractionnement des échantillons complexes de protéines avant leur analyse par LC-MS/MS [Fang et al. 2010].

### **2.1.3. Séparation des protéines par chromatographie liquide (LC)**

Dans le cas de mélanges peu complexes, les protéines peuvent être séparées par simple chromatographie liquide. Plusieurs types de chromatographies peuvent être utilisés comme la chromatographie d'exclusion stérique [Zhongli Zhang et al. 2001], la chromatographie d'échange d'ions [Schluesener et al. 2005] ou la chromatographie en phase inverse [Le Coutre et al. 2000].

Cette technique est peu résolutive et n'est applicable qu'à des protéines ne nécessitant pas de détergent pour leur solubilisation. Pourtant la chromatographie peut être utile dans les cas où un échantillon doit être enrichi pour la caractérisation d'une protéine par exemple. Elle peut de plus être couplée avec un spectromètre de masse ce qui permet d'obtenir directement des spectres de protéines entières [Roussel et al. 2008].

## **2.2. Séparation des peptides**

Les protéines entières sont difficiles à identifier par spectrométrie de masse MS ou MS/MS, c'est pourquoi les analyses sont la majeure partie du temps effectuées sur les digestats peptidiques de ces protéines. La digestion enzymatique contrôlée (généralement avec de la trypsine) augmente la complexité de l'échantillon mais facilite l'analyse par spectrométrie de masse [Sandra et al. 2008]. L'HPLC est la technique la plus couramment utilisée pour séparer les mélanges peptidiques. Elle est en effet très résolutive, flexible, et permet un couplage direct avec un spectromètre de masse via une source électrospray. De nombreux modes de séparations existent avec des sélectivités différentes comme la chromatographie d'échange d'ions (IEC), la chromatographie d'exclusion stérique (SEC) ou la chromatographie d'interaction hydrophile (HILIC). Mais la chromatographie de phase inverse (RPLC) avec un greffage C18 est la principale méthode séparative utilisée pour les peptides.

### **2.2.1. La séparation des peptides par nano-chromatographie liquide (nanoLC)**

Actuellement la méthode la plus classiquement utilisée pour la séparation de peptides est la RPLC. Ce type de chromatographie est très résolutif, dessale les échantillons et utilise des solvants compatibles avec l'ESI et la MS.

Les techniques de séparation par chromatographie liquide ont fait l'objet de nombreuses adaptations pour rendre robuste, sensible et résolutif le couplage LC-MS via la source ESI [Sandra et al. 2008]. Avec les sources ESI, l'intensité du signal MS est dépendant de la concentration des échantillons. Les problématiques biologiques posées nécessitant toujours plus de sensibilité, des systèmes de chromatographie liquide ont du être développés pour descendre à des débits de l'ordre de 100 à 300 nL/min utilisant des colonnes de très petit diamètre (75 µm). Ces développements instrumentaux permettent au couplage nanoLC-MS/MS de détecter quelques femtomoles de peptides [Han et al. 2008].

Les systèmes nanoLC-MS travaillent à des débits très faibles (100 à 600 nL/min). Les échantillons doivent donc être concentrés avant d'être séparés sur la colonne analytique. C'est

pourquoi 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 la colonne analytique par un gradient de solvants eau/acétonitrile ou eau/méthanol. La résolution chromatographique ainsi que la séparation des peptides sont déterminés par le gradient chromatographique, mais surtout par le type de colonne utilisée (taille des particules, type de phase stationnaire, longueur, diamètre, taille des pores).

Les développements technologiques de ces dernières années ont visé à améliorer la qualité de la séparation des composés [Sandra et al. 2008] mais également à gagner en temps d'analyse. Plusieurs voies ont été explorées :

- La **diminution de la taille des particules**, permet d'augmenter la résolution chromatographique et donc la sensibilité. En revanche ce type de colonne nécessite un système chromatographique adapté car réduire la taille des particules provoque une importante augmentation de pression (nanoAcquity UPLC Waters par exemple). Ce système est appelé UHPLC (Ultra High Pressure Liquid Chromatography) [Shen et al. 2005, Kay et al. 2007, Liu et al. 2007].
- Des **systèmes de puces microfluidiques** intégrant colonne de chargement, colonne analytique et aiguille nanospray (nanoLC-Chip/MS Agilent Technologies par exemple) permettent la réduction des volumes morts et donc à la fois une meilleure résolution chromatographique et un gain en temps d'analyse [Yin et al. 2004, Fortier et al. 2005, Yin et Killeen 2007]. Ce système sera détaillé dans la partie Résultats – Partie 1 – chapitre 1.
- Les **colonnes monolithiques** permettent de travailler avec des débits élevés car elles ont une très faible perte de charge et donc génèrent peu de pression, d'où une amélioration de la résolution chromatographique et un gain en temps d'analyse [Premstaller et al. 2001, Rieux et al. 2005, Xiaogang Jiang et al. 2008].

### 2.2.2. La chromatographie liquide multidimensionnelle : l'approche "shotgun"

L'approche dite "shotgun" désigne l'analyse directe de mélanges complexes de protéines pour identifier l'ensemble des protéines présentes dans un échantillon biologique. Le mélange complexe de protéines est directement digéré en peptides qui sont séparés par des étapes de chromatographie multiples avant d'être analysés en MS et fragmentés en MS/MS.

Cette stratégie est issue de l'apparition de la chromatographie multidimensionnelle MudPIT (Multi-Dimensional Protein Identification Technology) introduite par Yates [Link et al. 1999, Washburn et al. 2001, Wolters et al. 2001] en réponse aux limitations des seules techniques utilisées jusqu'alors : les gels d'électrophorèse mono et bidimensionnels.

Le principe de la chromatographie multidimensionnelle consiste à coupler des systèmes de chromatographie basés sur des mécanismes de rétention différents pour obtenir des séparations en théorie orthogonales entre les différentes dimensions du système. La plupart du temps cette séparation est effectuée en 2 dimensions. Généralement la dernière étape de séparation avant

l'analyse par spectrométrie de masse est la RPLC. En revanche lors de la première étape il est possible d'utiliser une grande variété de techniques comme la RPLC à pH basique [Delmotte et al. 2007] mais la plus courante reste la chromatographie d'échanges de cations [Motoyama et Yates 2008, Sandra et al. 2009].

Le nombre de protéines pouvant être identifiées par ces approches MudPIT peut largement dépasser le nombre de protéines identifiées par des approches de gel 2D [Motoyama et Yates 2008]. Cependant, avec ces approches, le nombre de peptides servant à l'identification des protéines est souvent faible et les identifications peuvent contenir des taux de faux positifs non négligeables [Nesvizhskii et Aebersold 2005]. De plus dans les approches "shotgun", la digestion des protéines constitue la première étape et les séparations ne se font donc qu'au niveau peptidique. La relation entre les peptides et la protéine de départ est donc perdue [Nesvizhskii et Aebersold 2005]. Or un peptide d'une séquence donnée peut être présent dans plusieurs protéines. Il est donc impossible de distinguer des isoformes ou des hétérogénéités au niveau des modifications post-traductionnelles.

Dans le cas de l'analyse de protéines N-glycosylées l'approche "shotgun" pose un certain nombre de difficultés :

- Une étape d'enrichissement de l'échantillon en glycopeptides est nécessaire afin de pouvoir les analyser (voir Chapitre 1) [Cantin et Yates 2004, Sun et al. 2007];
- Comme il est décrit dans le chapitre 1 § 2.7.3. la fragmentation des glycopeptides par CID ne donne que peu voire pas d'information sur la séquence en acides aminés du glycopeptide. Il est donc souvent difficile voire impossible de déterminer à quelle protéine appartient le glycopeptide lorsque l'on utilise une approche "shotgun".

### 3. Les stratégies d'identification en analyse protéomique par spectrométrie de masse

L'analyse protéomique par spectrométrie de masse couplée aux méthodes de séparation des protéines et des peptides fournit un ensemble de spectres MS et MS/MS reflétant l'ensemble des protéines détectables présentes dans un échantillon. L'interprétation de ces spectres s'effectue à l'aide d'outils bioinformatiques et permet d'identifier les protéines correspondantes. Pour cela les données mesurées (par MS et/ou MS/MS) sont comparées à l'aide de moteurs de recherche aux données théoriques issues de banques de données protéiques.

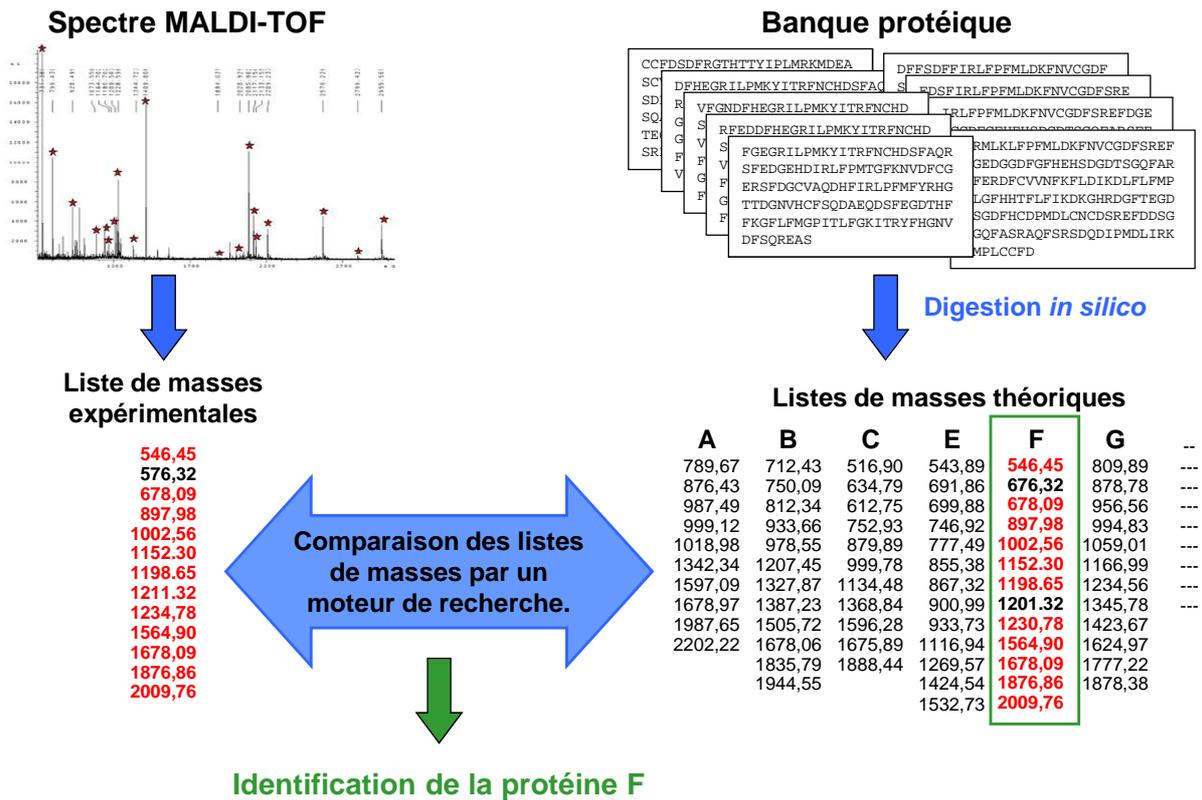
#### 3.1. Les stratégies d'identification de protéines

##### 3.1.1. L'empreinte peptidique massique (Peptide Mass Fingerprint (PMF))

L'empreinte peptidique massique a été décrite par 5 groupes différents simultanément [Henzel et al. 1993, James et al. 1993, Matthias Mann et al. 1993, Pappin et al. 1993, Yates et al. 1993] au début des années 90. Cette stratégie consiste en l'identification des protéines par la mesure des masses de peptides générés suite à leur digestion enzymatique. Ces masses sont ensuite comparées à celles de peptides théoriques obtenus à partir de la digestion *in silico* des protéines présentes dans une banque de séquences (Figure 11).

Généralement, la stratégie PMF est utilisée pour identifier des protéines contenues dans un extrait peu complexe comme un spot de gel 2D par exemple. Les protéines sont digérées et analysées par MS avec un instrument généralement de type MALDI-TOF qui a l'avantage d'être sensible, résolutif, facilement automatisable, rapide, avec une bonne précision de mesure de masse et relativement tolérant aux sels. Des instruments de dernière génération combinant des sources ESI ou MALDI avec des analyseurs à haute voire très haute résolution (TOF, Q-TOF, FTICR, Orbitrap) peuvent aussi réaliser du PMF.

Malgré ces avantages, l'empreinte peptidique massique présente certaines limites. Il est par exemple impossible d'analyser directement un échantillon complexe. De plus, étant donné la croissance exponentielle des banques de données protéiques, la spécificité du PMF est de plus en plus limitée [Gattiker et al. 2002]. Ainsi, bien que cette stratégie soit efficace dans bien des cas, le séquençage MS/MS des peptides est une méthode d'identification bien plus sensible et spécifique [Steen et Mann 2004].



**Figure 11.** Stratégie d'identification des protéines par empreinte peptidique massique à partir d'une analyse MALDI-TOF

### 3.1.2. Identification par LC-MS/MS

Cette stratégie consiste à utiliser les données des spectres MS/MS pour réaliser l'identification des séquences des peptides générés par digestion des protéines de l'échantillon. Elle représente aujourd'hui le cœur de l'analyse protéomique par spectrométrie de masse.

A partir d'un échantillon biologique plus ou moins complexe, on réalise une digestion enzymatique qui génère des peptides qui sont ensuite analysés par LC-MS/MS. Lors des analyses LC-MS/MS, la sélection des ions parents qui seront fragmentés au cours de l'analyse est réalisée de manière automatique sur la base de leur intensité, leur rapport m/z ou encore leur état de charge. C'est ce qu'on appelle le mode DDA (Data Dependant Acquisition).

Cette stratégie présente de nombreux avantages par rapport à la stratégie PMF. En effet, en ajoutant une étape de chromatographie liquide, cette stratégie permet d'analyser des mélanges de peptides beaucoup plus complexes puisque les peptides sont séparés avant d'être analysés par spectrométrie de masse réduisant considérablement les phénomènes de suppression de signal. De plus l'étape de fragmentation MS/MS permet d'obtenir des informations de séquence qui ne sont pas accessibles avec le PMF. L'ensemble des spectres de fragmentation permettra alors d'identifier des protéines de manière beaucoup plus fiable qu'avec le PMF. Avec cette méthode il est également

possible de mettre en évidence des modifications post-traductionnelles comme des glycosylations ou des mutations dans la séquence peptidique.

### 3.1.2.1. L'empreinte de fragmentation peptidique (Peptide Fragment Fingerprinting (PFF))

La dénomination "Peptide Fragment Fingerprinting" (PFF) est introduite en 2004 [Blueggel et al. 2004] par analogie avec la dénomination PMF. Cette approche est basée sur la comparaison des listes de masses des ions parents et fragments mesurés avec des ions parents et fragments théoriques contenus dans une banque protéique (Figure 12).

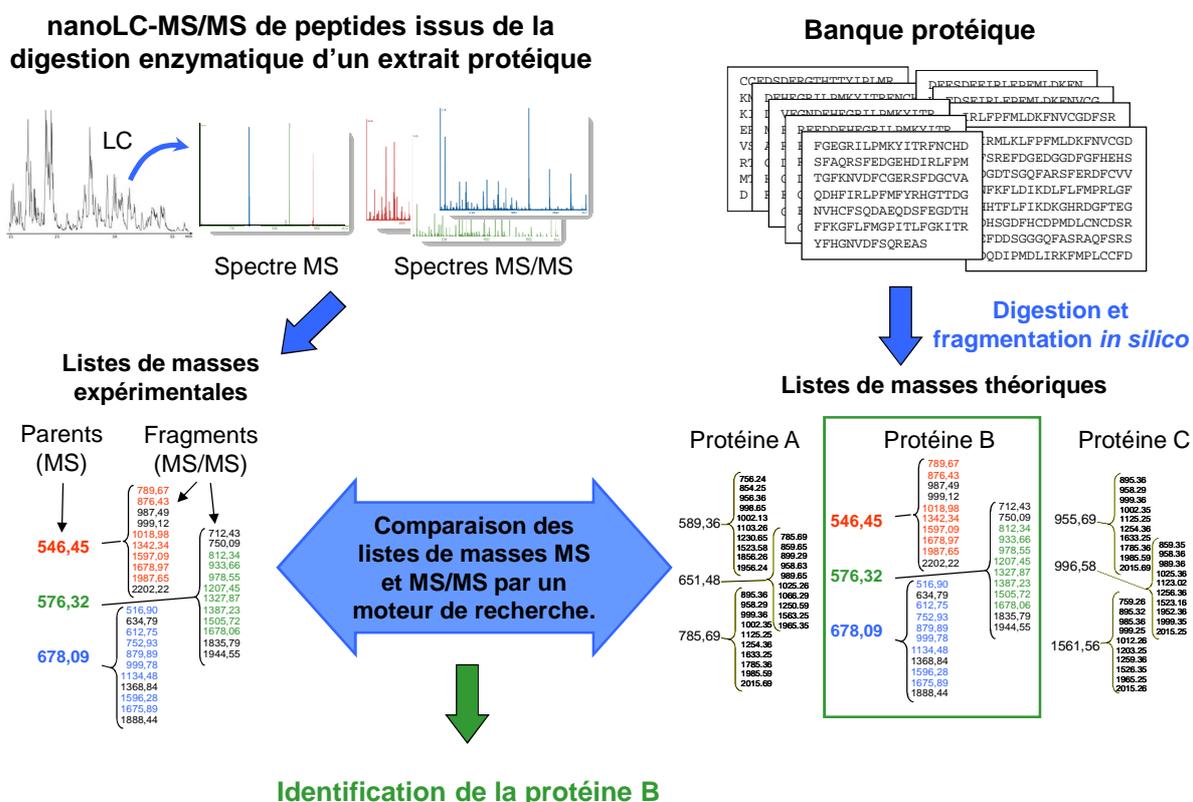


Figure 12. Stratégie d'identification de protéines par LC-MS/MS (PFF)

Un grand nombre de moteurs de recherche permettent de réaliser cette approche dont les plus utilisés sont :

- Mascot (<http://matrixscience.com>) [Perkins et al. 1999]
- OMSSA (<http://pubchem.ncbi.nlm.nih.gov/omssa>) [Geer et al. 2004]
- Sequest (<http://fields.scripps.edu/sequest/>) [Eng et al. 1994]
- ProteinProspector (<http://prospector.ucsf.edu>) [Clauser et al. 1999]
- X!Tandem (<http://www.thegpm.org>) [Craig et Beavis 2004]
- SpectrumMill (<http://www.chem.agilent.com>)
- Phenyx (<http://www.phenyx-ms.com>) [Colinge et al. 2003]

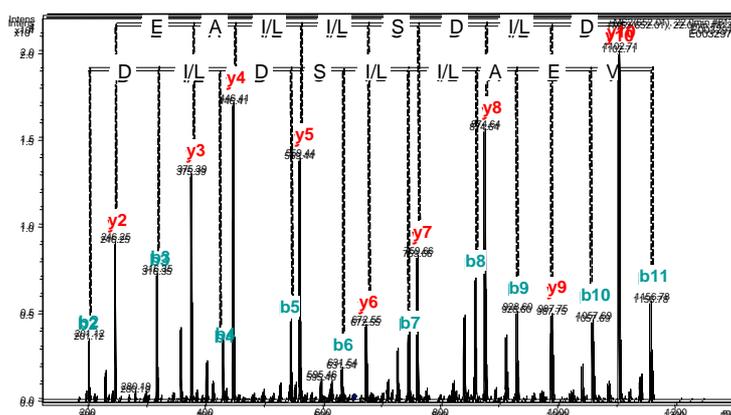
Ces programmes fournissent une liste de séquences peptidiques auxquelles sont attribuées un score mesurant la "qualité" de l'identification. Même si l'ensemble de ces moteurs de recherche ont un système de fonctionnement assez proche, ils sont basés sur des algorithmes de recherche différents. Quelques études ont été réalisées sur la comparaison de différents moteurs de recherche et ont permis d'observer qu'ils présentaient des performances différentes en fonction du type d'échantillon et du spectromètre de masse [Bakalarski et al. 2007, Balgley et al. 2007]. C'est pourquoi il peut être utile voire recommandé d'en utiliser plusieurs afin d'augmenter la fiabilité des identifications [Kapp et al. 2005, Searle et al. 2008, Jones et al. 2009].

Bien qu'extrêmement utilisée en analyse protéomique, le PFF possède certaines limites. En effet, ce type de méthode n'est utilisable que pour identifier des séquences présentes dans les banques protéiques. Les séquences provenant d'organismes non séquencés, séquencés mais non annotés, contenant des mutations inconnues ou des modifications post-traductionnelles complexes (type glycosylations) ne peuvent pas être identifiées. Une autre stratégie doit alors être adoptée : le séquençage *de novo*.

### **3.1.2.2. Le séquençage *de novo***

Dans cette approche, les séquences en acides aminés sont directement déduites des différences de masses entre les ions fragments successifs à partir des spectres MS/MS (§ 1.3.1.). Il est ainsi possible d'obtenir des fragments de séquence en acides aminés appelés "tags" (Figure 13). Ce type d'interprétation peut s'effectuer manuellement mais plusieurs logiciels ont été développés permettant d'automatiser l'opération comme Lutefisk [Richard Johnson et Taylor 2002], Pepnovo [Frank et Pevzner 2005] ou Peaks [Ma et al. 2003]. Ces "tags" peuvent être ensuite soumis à des programmes de recherche d'homologies de séquence comme MS-Blast [Shevchenko et al. 2001].

Malgré ces logiciels, cette approche reste très couteuse en temps et nécessite des spectres de fragmentation de très bonne qualité. En revanche, l'approche *de novo* permet l'interprétation des spectres pour lesquelles la séquence peptidique exacte correspondante n'est pas présente dans les banques de données. Ainsi avec cette approche, il est possible d'identifier des peptides modifiés (mutation, modifications post-traductionnelles...) ou provenant de protéines d'organismes dont le génome n'a pas été séquencé par exemple.



**Figure 13.** Exemple d'un spectre de fragmentation MS/MS illustrant la détermination de la séquence en acides aminés d'un peptide. (Spectre de fragmentation de l'ion  $m/z$  652.0, 2+ ; Séquence S-L/I-D-L-D-S-L/I-L/I-A-E-V-K).

### 3.2. Les banques de données

Les identifications et les interprétations biologiques qui découlent des analyses protéomiques dépendent directement de la disponibilité de banques de séquences et de la qualité de leur annotation. Le développement des banques de données protéiques a accompagné l'essor de l'analyse protéomique par spectrométrie de masse. En effet, la disponibilité croissante des banques de séquences a permis l'émergence des stratégies d'identification automatique des protéines.

De nombreuses banques protéiques sont aujourd'hui disponibles [Rolf Apweiler et al. 2004]. Chacune est caractérisée par son exhaustivité, son degré de redondance, la qualité de ses annotations.

On peut ainsi distinguer deux catégories de banques protéiques :

**1.** Les banques généralistes de dépôt contenant essentiellement des séquences protéiques issues de la traduction de gènes prédits *in silico* par des algorithmes de prédiction sans vérification et avec des annotations fonctionnelles limitées voire inexistantes. Parmi ces banques, on peut citer la banque GenPept (GenBank Gene Products Data Bank) produite par le NCBI (National Center for Biotechnology Information) regroupant l'ensemble des protéines traduites à partir des banques nucléotidiques, sans annotation et avec une grande redondance.

**2.** Les banques protéiques couramment utilisées en protéomique sont des banques pour lesquelles un travail de nettoyage, de tri, de documentation et d'annotation a été réalisé par des experts. La liste de ces banques est longue et seules les banques les plus couramment utilisées en analyse protéomique sont brièvement détaillées ici :

- Le consortium UniProt :

Swiss-Prot (UniProtKB/SwissProt) qui est l'une des banques corrigées de référence. Il s'agit de la banque de séquence protéique la moins redondante et la mieux annotée, mais elle est incomplète;

TrEMBL (UniProtKB/TrEMBL) qui contient des séquences annotées *in silico* d'après la traduction de l'ensemble des séquences codantes des banques nucléotidiques hormis celles déjà présentes dans Swiss-Prot. Elle contient également des séquences protéiques issues de publications ou déposées directement par les utilisateurs. La banque TrEMBL est complémentaire de SwissProt et permet un accès rapide aux utilisateurs à de nouvelles séquences protéiques.

- La banque NCBI's Entrez Protein est également très utilisée. Cette banque généraliste se trouve à la jonction des deux types de banques. Elle contient à la fois des séquences issues de la traduction automatique des banques nucléotidiques comme une banque de dépôt mais également des séquences annotées de banques protéiques. Cette banque est donc volumineuse et redondante mais elle présente l'avantage d'être relativement complète.

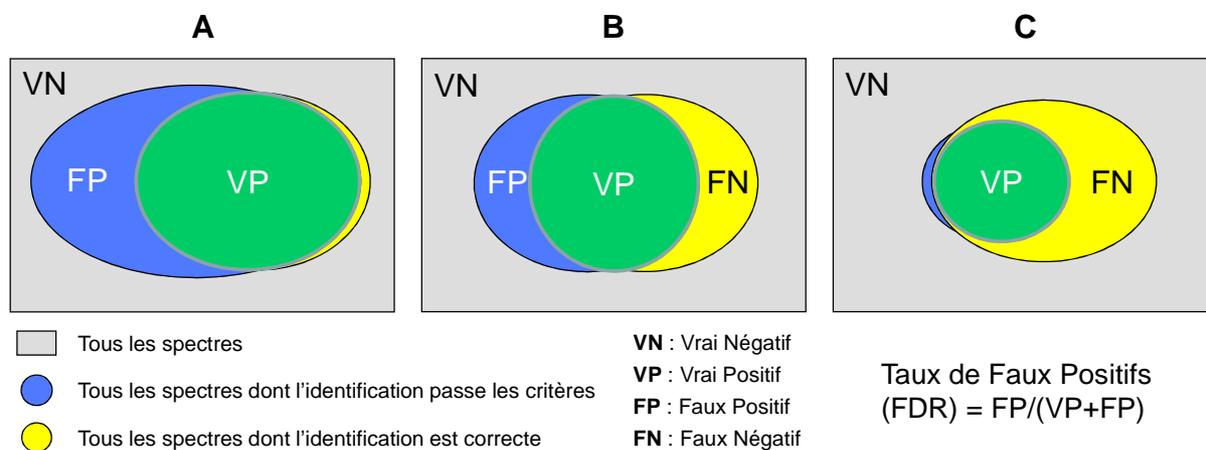
### 3.3. Evaluation des résultats

Les scores calculés par les moteurs de recherche ne constituent pas des indicateurs suffisamment fiables pour considérer la pertinence des identifications. Une étape supplémentaire de vérification est indispensable particulièrement pour les études nécessitant l'interprétation d'une grande quantité de données. Il est en effet important pour le protéomique d'associer une évaluation de la confiance accordée aux résultats d'identifications afin de ne pas biaiser les conclusions biologiques de l'étude. C'est pourquoi, les journaux spécialisés dans le domaine ("Molecular and Cellular Proteomics" ou "Proteomics") ont établi des directives sur les informations et la documentation à associer à l'analyse et l'identification des peptides et des protéines qui doivent être suivies pour pouvoir publier dans ces journaux. Ces directives permettent de formuler des critères standards pour que les éditeurs, les "reviewers" et les lecteurs puissent évaluer plus aisément les données issues de l'analyse protéomique.

La vérification des résultats obtenus à partir des moteurs de recherche peut être effectuée manuellement mais devant la nécessité de plus en plus pressante d'estimer empiriquement la confiance dans les identifications, des stratégies d'évaluation de taux de faux positif sont apparues.

L'une d'entre elles est l'approche "Target-Decoy" [Joshua E. Elias et Gygi 2007]. Cette stratégie consiste à effectuer les recherches dans une banque protéique classique "target / cible" concaténée avec une banque "decoy / leurre" qui contient toutes les séquences inversées (ou mélangées ou randomisées) de la même banque. L'approche suppose que les identifications de peptides dans la banque "leurre" et les identifications de faux positifs dans la banque "cible" originale suivent la même distribution. Dans un deuxième temps, les identifications de peptides sont filtrés en ajustant les seuils de scores du moteur de recherche afin d'obtenir un taux de faux positif le plus bas possible (Figure

14). La stratégie "Target-Decoy" ne permet pas de déterminer exactement quelles identifications sont correctes ou non mais elle évalue le taux de faux positifs (FDR). Cette stratégie permet d'estimer la probabilité qu'une identification soit correcte en considérant qu'elle fait partie d'un ensemble d'identifications.



## 4. Conclusion

Les analyses protéomiques comportent un grand nombre d'étapes qui nécessitent l'utilisation de plusieurs techniques que sont la préparation d'échantillon, la chromatographie liquide, la spectrométrie de masse et l'interprétation des données. Chacune des ces étapes doit être choisie et optimisée en fonction du type d'échantillon et de la question biologique posée. De plus, chaque technique utilisée peut induire des biais et nécessite donc d'être parfaitement maîtrisée. Plusieurs études montrent qu'il est difficile d'obtenir une bonne répétabilité et une bonne reproductibilité lors de l'analyse de protéines par LC-MS/MS [J. E. Elias et al. 2005, Van Midwoud et al. 2007, Aebersold 2009, Bell et al. 2009, Tabb et al. 2010]. Le nombre de peptides et de protéines identifiés ainsi que la confiance dans les identifications dépend étroitement de l'état du couplage nanoLC-MS/MS utilisé. Ainsi, la vérification des performances du système chromatographique et du spectromètre de masse fait partie des points clefs lors des expériences d'analyse protéomique, particulièrement lorsque l'échantillon est disponible en très faible quantité.

# **BIBLIOGRAPHIE**



- Aebersold, R. (2009). "A stress test for mass spectrometry-based proteomics." *Nat Methods* 6(6): 411-412.
- Alley, W. R., Jr., Mechref, Y. et Novotny, M. V. (2009). "Characterization of glycopeptides by combining collision-induced dissociation and electron-transfer dissociation mass spectrometry data." *Rapid Commun Mass Spectrom* 23(1): 161-170.
- Alvarez-Manilla, G., Atwood, J., 3rd, Guo, Y., Warren, N. L., Orlando, R. et Pierce, M. (2006). "Tools for glycoproteomic analysis: size exclusion chromatography facilitates identification of tryptic glycopeptides with N-linked glycosylation sites." *J Proteome Res* 5(3): 701-708.
- Alvarez-Manilla, G., Warren, N. L., Atwood, J. A., Dalton, S., Orlando, R. et Pierce, M. (2009). "Glycoproteomic analysis of embryonic stem cells: identification of potential glycobiomarkers using lectin affinity chromatography of glycopeptides." *J Proteome Res*.
- An, H. J., Peavy, T. R., Hedrick, J. L. et Lebrilla, C. B. (2003). "Determination of N-glycosylation sites and site heterogeneity in glycoproteins." *Anal Chem* 75(20): 5628-5637.
- Aoki-Kinoshita, K. F. (2008). "An introduction to bioinformatics for glycomics research." *PLoS Comput Biol* 4(5): e1000075.
- Apweiler, R., Bairoch, A. et Wu, C. H. (2004). "Protein sequence databases." *Current Opinion in Chemical Biology* 8(1): 76-80.
- Apweiler, R., Hermjakob, H. et Sharon, N. (1999). "On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database." *Biochim Biophys Acta* 1473(1): 4-8.
- Bakalarski, C., Haas, W., Dephoure, N. et Gygi, S. (2007). "The effects of mass accuracy, data acquisition speed, and search algorithm choice on peptide identification rates in phosphoproteomics." *Analytical and Bioanalytical Chemistry* 389(5): 1409-1419.
- Balgley, B. M., Laudeman, T., Yang, L., Song, T. et Lee, C. S. (2007). "Comparative Evaluation of Tandem MS Search Algorithms Using a Target-Decoy Search Strategy." *Molecular & Cellular Proteomics* 6(9): 1599-1608.
- Barton, S., J. et Whittaker, J., C. (2009). "Review of factors that influence the abundance of ions produced in a tandem mass spectrometer and statistical methods for discovering these factors." *Mass Spectrometry Reviews* 28(1): 177-187.
- Bednarczyk, A. (2008). "Nouvelles méthodologies en protéomique pour une caractérisation fine des protéines." Université de Strasbourg.
- Bell, A. W., Deutsch, E. W., Au, C. E., Kearney, R. E., Beavis, R., Sechi, S., Nilsson, T. et Bergeron, J. J. (2009). "A HUPO test sample study reveals common problems in mass spectrometry-based proteomics." *Nat Methods* 6(6): 423-430.
- Biemann, K. (1988). "Contributions of mass spectrometry to peptide and protein structure." *Biomed Environ Mass Spectrom* 16(1-12): 99-111.
- Bigge, J. C., Patel, T. P., Bruce, J. A., Goulding, P. N., Charles, S. M. et Parekh, R. B. (1995). "Nonselective and efficient fluorescent labeling of glycans using 2-amino benzamide and anthranilic acid." *Anal Biochem* 230(2): 229-238.
- Blomme, B., Van Steenkiste, C., Callewaert, N. et Van Vlierberghe, H. (2009). "Alteration of protein glycosylation in liver diseases." *J Hepatol* 50(3): 592-603.
- Blueggel, M., Chamrad, D. et Meyer, H. E. (2004). "Bioinformatics in proteomics." *Curr Pharm Biotechnol* 5(1): 79-88.

Braun, R., Kinkl, N., Beer, M. et Ueffing, M. (2007). "Two-dimensional electrophoresis of membrane proteins." *Analytical and Bioanalytical Chemistry* 389(4): 1033-1045.

Brooks, S. A. (2009). "Strategies for analysis of the glycosylation of proteins: current status and future perspectives." *Mol Biotechnol* 43(1): 76-88.

Bunkenborg, J., Pilch, B. J., Podtelejnikov, A. V. et Wisniewski, J. R. (2004). "Screening for N-glycosylated proteins by liquid chromatography mass spectrometry." *Proteomics* 4(2): 454-465.

Calvano, C. D., Zamboni, C. G. et Jensen, O. N. (2008). "Assessment of lectin and HILIC based enrichment protocols for characterization of serum glycoproteins by mass spectrometry." *J Proteomics* 71(3): 304-317.

Campana, J. E. (1980). "Elementary theory of the quadrupole mass filter." *International Journal of Mass Spectrometry and Ion Physics* 33(2): 101-117.

Cantin, G. T. et Yates, J. R., 3rd (2004). "Strategies for shotgun identification of post-translational modifications by mass spectrometry." *J Chromatogr A* 1053(1-2): 7-14.

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

Catalina, M. I., Koeleman, C. A., Deelder, A. M. et Wührer, M. (2007). "Electron transfer dissociation of N-glycopeptides: loss of the entire N-glycosylated asparagine side chain." *Rapid Commun Mass Spectrom* 21(6): 1053-1061.

Clauser, K. R., Baker, P. et Burlingame, A. L. (1999). "Role of accurate mass measurement ( $\pm 10$  ppm) in protein identification strategies employing MS or MS/MS and database searching." *Anal Chem* 71(14): 2871-2882.

Colinge, J., Masselot, A., Giron, M., Dessingy, T. et Magnin, J. (2003). "OLAV: towards high-throughput tandem mass spectrometry data identification." *Proteomics* 3(8): 1454-1463.

Conboy, J. J. et Henion, J. D. (1992). "The determination of glycopeptides by liquid chromatography/mass spectrometry with collision-induced dissociation." *Journal of the American Society for Mass Spectrometry* 3(8): 804-814.

Cox, K. A., Gaskell, S. J., Morris, M. et Whiting, A. (1996). "Role of the site of protonation in the low-energy decompositions of gas-phase peptide ions." *Journal of the American Society for Mass Spectrometry* 7(6): 522-531.

Craig, R. et Beavis, R. C. (2004). "TANDEM: matching proteins with tandem mass spectra." *Bioinformatics* 20(9): 1466-1467.

Dalpathado, D. S. et Desaire, H. (2008). "Glycopeptide analysis by mass spectrometry." *Analyst* 133(6): 731-738.

Damen, C. W., Chen, W., Chakraborty, A. B., van Oosterhout, M., Mazzeo, J. R., Gebler, J. C., Schellens, J. H., Rosing, H. et Beijnen, J. H. (2009). "Electrospray Ionization Quadrupole Ion-Mobility Time-of-Flight Mass Spectrometry as a Tool to Distinguish the Lot-to-Lot Heterogeneity in N-Glycosylation Profile of the Therapeutic Monoclonal Antibody Trastuzumab." *J Am Soc Mass Spectrom*.

Delmotte, N., Lasaosa, M., Tholey, A., Heinze, E. et Huber, C. G. (2007). "Two-dimensional reversed-phase x ion-pair reversed-phase HPLC: an alternative approach to high-resolution peptide separation for shotgun proteome analysis." *J Proteome Res* 6(11): 4363-4373.

Demelbauer, U. M., Zehl, M., Plematl, A., Allmaier, G. et 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(14): 1575-1582.

Deshpande, N., Jensen, P. H., Packer, N. H. et Kolarich, D. (2010). "GlycoSpectrumScan: fishing glycopeptides from MS spectra of protease digests of human colostrum sIgA." J Proteome Res 9(2): 1063-1075.

Dole, M., Mack, L. L., Hines, R. L., Mobley, R. C., Ferguson, L. D. et Alice, M. B. (1968). "Molecular Beams of Macroions." The Journal of Chemical Physics 49(5): 2240-2249.

Domon, B. et Costello, C. E. (1988). "A Systematic Nomenclature for Carbohydrate Fragmentations in FAB-MS/MS Spectra of Glycoconjugates." Glycoconjugate J 5: 397-409.

Dongre, A. R., Jones, J. L., Somogyi, A. et Wysocki, V. H. (1996). "Influence of Peptide Composition, Gas-Phase Basicity, and Chemical Modification on Fragmentation Efficiency: Evidence for the Mobile Proton Model." Journal of the American Chemical Society 118(35): 8365-8374.

Douglas, D. J. (2009). "Linear quadrupoles in mass spectrometry." Mass Spectrometry Reviews 28(6): 937-960.

Dube, D. H. et Bertozzi, C. R. (2005). "Glycans in cancer and inflammation--potential for therapeutics and diagnostics." Nat Rev Drug Discov 4(6): 477-488.

Dwek, R. A. (1996). "Glycobiology: Toward Understanding the Function of Sugars." Chem Rev 96(2): 683-720.

Elias, J. E. et Gygi, S. P. (2007). "Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry." Nat Meth 4(3): 207-214.

Elias, J. E., Haas, W., Faherty, B. K. et Gygi, S. P. (2005). "Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations." Nat Methods 2(9): 667-675.

Emmett, M. R. et Caprioli, R. M. (1994). "Micro-electrospray mass spectrometry: ultra-high-sensitivity analysis of peptides and proteins." Journal of the American Society for Mass Spectrometry 5(7): 605-613.

Eng, J. K., McCormack, A. L. et Yates, J. R., 3rd (1994). "An approach to correlate tandem mass-spectral data of peptides with amino-acid-sequences in a protein database." Journal of the American Society for Mass Spectrometry 5: 976-989.

Fang, Y., Robinson, D. P. et Foster, L. J. (2010). "Quantitative Analysis of Proteome Coverage and Recovery Rates for Upstream Fractionation Methods in Proteomics." Journal of Proteome Research 9(4): 1902-1912.

Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. et Whitehouse, C. M. (1989). "Electrospray ionization for mass spectrometry of large biomolecules." Science 246(4926): 64-71.

Fortier, M.-H., Bonneil, E., Goodley, P. et Thibault, P. (2005). "Integrated Microfluidic Device for Mass Spectrometry-Based Proteomics and Its Application to Biomarker Discovery Programs." Analytical Chemistry 77(6): 1631-1640.

Frank, A. et Pevzner, P. (2005). "PepNovo: de novo peptide sequencing via probabilistic network modeling." Anal Chem 77(4): 964-973.

Fryksdale, B., G., Jedrzejewski, P., T., Wong, D., L., Gaertner, A., L. et Miller, B., S. (2002). "Impact of deglycosylation methods on two-dimensional gel electrophoresis and matrix assisted laser desorption/ionization-time of flight-mass spectrometry for proteomic analysis." ELECTROPHORESIS 23(14): 2184-2193.

- Gaskell, S. J. (1997). "Electrospray: Principles and Practice." *Journal of Mass Spectrometry* 32(7): 677-688.
- Gattiker, A., Bienvenut, W., V., Bairoch, A. et Gasteiger, E. (2002). "FindPept, a tool to identify unmatched masses in peptide mass fingerprinting protein identification." *PROTEOMICS* 2(10): 1435-1444.
- Geer, L. Y., Markey, S. P., Kowalak, J. A., Wagner, L., Xu, M., Maynard, D. M., Yang, X., Shi, W. et Bryant, S. H. (2004). "Open mass spectrometry search algorithm." *J Proteome Res* 3(5): 958-964.
- Geyer, H. et Geyer, R. (2006). "Strategies for analysis of glycoprotein glycosylation." *Biochim Biophys Acta* 1764(12): 1853-1869.
- Jimenez, E., Benavente, F., Barbosa, J. et Sanz-Nebot, V. (2007). "Towards a reliable molecular mass determination of intact glycoproteins by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry." *Rapid Commun Mass Spectrom* 21(16): 2555-2563.
- Goldberg, D., Bern, M., Parry, S., Sutton-Smith, M., Panico, M., Morris, H. R. et Dell, A. (2007). "Automated N-Glycopeptide Identification Using a Combination of Single- and Tandem-MS." *Journal of Proteome Research* 6(10): 3995-4005.
- Gomord, V. et Faye, L. (2004). "Posttranslational modification of therapeutic proteins in plants." *Curr Opin Plant Biol* 7(2): 171-181.
- Gonzalez, J., Takao, T., Hori, H., Besada, V., Rodriguez, R., Padron, G. et Shimonishi, Y. (1992). "A method for determination of N-glycosylation sites in glycoproteins by collision-induced dissociation analysis in fast atom bombardment mass spectrometry: Identification of the positions of carbohydrate-linked asparagine in recombinant [alpha]-amylase by treatment with peptide-N-glycosidase F in 18O-labeled water." *Analytical Biochemistry* 205(1): 151-158.
- Görg, A., Weiss, W. et Dunn, M. J. (2004). "Current two-dimensional electrophoresis technology for proteomics." *PROTEOMICS* 4(12): 3665-3685.
- Han, X., Aslanian, A. et Yates Iii, J. R. (2008). "Mass spectrometry for proteomics." *Current Opinion in Chemical Biology* 12(5): 483-490.
- Harvey, D. J. (2005). "Structural determination of N-linked glycans by matrix-assisted laser desorption/ionization and electrospray ionization mass spectrometry." *Proteomics* 5(7): 1774-1786.
- Helenius, A. et Aebi, M. (2001). "Intracellular functions of N-linked glycans." *Science* 291(5512): 2364-2369.
- Henzel, W. J., Billeci, T. M., Stults, J. T., Wong, S. C., Grimley, C. et Watanabe, C. (1993). "Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases." *Proceedings of the National Academy of Sciences of the United States of America* 90(11): 5011-5015.
- Hoffman, M. D., Sniatynski, M. J. et Kast, J. (2008). "Current approaches for global post-translational modification discovery and mass spectrometric analysis." *Anal Chim Acta* 627(1): 50-61.
- Huddleston, M. J., Bean, M. F. et Carr, S. A. (1993). "Collisional fragmentation of glycopeptides by electrospray ionization LC/MS and LC/MS/MS: methods for selective detection of glycopeptides in protein digests." *Anal Chem* 65(7): 877-884.
- Ikonomou, M. G., Blades, A. T. et Kebarle, P. (1990). "Investigations of the electrospray interface for liquid chromatography/mass spectrometry." *Analytical Chemistry* 62(9): 957-967.
- Irungu, J., Go, E. P., Dalpathado, D. S. et Desaire, H. (2007). "Simplification of mass spectral analysis of acidic glycopeptides using GlycoPep ID." *Anal Chem* 79(8): 3065-3074.

- James, P., Quadroni, M., Carafoli, E. et Gonnet, G. (1993). "Protein Identification by Mass Profile Fingerprinting." *Biochemical and Biophysical Research Communications* 195(1): 58-64.
- Jensen, O. N. (2006). "Interpreting the protein language using proteomics." *Nat Rev Mol Cell Biol* 7(6): 391-403.
- Jiang, H., Desaire, H., Butnev, V. Y. et Bousfield, G. R. (2004). "Glycoprotein profiling by electrospray mass spectrometry." *J Am Soc Mass Spectrom* 15(5): 750-758.
- Jiang, X., Dong, J., Wang, F., Feng, S., Ye, M. et Zou, H. (2008). "Automation of nanoflow liquid chromatography-tandem mass spectrometry for proteome and peptide profiling analysis by using a monolithic analytical capillary column." *ELECTROPHORESIS* 29(8): 1612-1618.
- Johnson, R. et Taylor, J. (2002). "Searching sequence databases via De novo peptide sequencing by tandem mass spectrometry." *Molecular Biotechnology* 22(3): 301-315.
- Johnson, R. S., Martin, S. A., Biemann, K., Stults, J. T. et Watson, J. T. (1987). "Novel fragmentation process of peptides by collision-induced decomposition in a tandem mass spectrometer: differentiation of leucine and isoleucine." *Anal Chem* 59(21): 2621-2625.
- Jones, A., R. , Siepen, J., A. , Hubbard, S., J. et Paton, N., W. (2009). "Improving sensitivity in proteome studies by analysis of false discovery rates for multiple search engines." *PROTEOMICS* 9(5): 1220-1229.
- Kaji, H., Saito, H., Yamauchi, Y., Shinkawa, T., Taoka, M., Hirabayashi, J., Kasai, K., Takahashi, N. et Isobe, T. (2003). "Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify N-linked glycoproteins." *Nat Biotechnol* 21(6): 667-672.
- Kapp, E., A. , Schütz, F., Connolly, L., M. , Chakel, J., A. , Meza, J., E. , Miller, C., A. , Fenyo, D., Eng, J., K. , Adkins, J., N. , Omenn, G., S. et Simpson, R., J. (2005). "An evaluation, comparison, and accurate benchmarking of several publicly available MS/MS search algorithms: Sensitivity and specificity analysis." *PROTEOMICS* 5(13): 3475-3490.
- Karas, M., Bachmann, D., Bahr, U. et Hillenkamp, F. (1987). "Matrix-assisted ultraviolet laser desorption of non-volatile compounds." *Int J Mass Spectrom Ion Proc* 78: 53-68.
- Karas, M., Bahr, U. et Dülcks, T. (2000). "Nano-electrospray ionization mass spectrometry: addressing analytical problems beyond routine." *Fresenius' Journal of Analytical Chemistry* 366(6): 669-676.
- Karas, M. et Hillenkamp, F. (1988). "Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons." *Anal Chem* 60(20): 2299-2301.
- Kay, R., G. , Gregory, B., Grace, P., B. et Pleasance, S. (2007). "The application of ultra-performance liquid chromatography/tandem mass spectrometry to the detection and quantitation of apolipoproteins in human serum." *Rapid Communications in Mass Spectrometry* 21(16): 2585-2593.
- Kebarle, P. et Verkerk, U. H. (2009). "Electrospray: from ions in solution to ions in the gas phase, what we know now." *Mass Spectrom Rev* 28(6): 898-917.
- Kenrick, K. G. et Margolis, J. (1970). "Isoelectric focusing and gradient gel electrophoresis: A two-dimensional technique." *Analytical Biochemistry* 33(1): 204-207.
- Küster, B. et Mann, M. (1999). "18O-Labeling of N-Glycosylation Sites To Improve the Identification of Gel-Separated Glycoproteins Using Peptide Mass Mapping and Database Searching." *Analytical Chemistry* 71(7): 1431-1440.
- Larsen, M. R., Hojrup, P. et Roepstorff, P. (2005). "Characterization of gel-separated glycoproteins using two-step proteolytic digestion combined with sequential microcolumns and mass spectrometry." *Mol Cell Proteomics* 4(2): 107-119.

Larsen, M. R., Jensen, S. S., Jakobsen, L. A. et Heegaard, N. H. (2007). "Exploring the sialome using titanium dioxide chromatography and mass spectrometry." *Mol Cell Proteomics* 6(10): 1778-1787.

le Coutre, J., Whitelegge, J. P., Gross, A., Turk, E., Wright, E. M., Kaback, H. R. et Faull, K. F. (2000). "Proteomics on Full-Length Membrane Proteins Using Mass Spectrometry" *Biochemistry* 39(15): 4237-4242.

Ling, V., Guzzetta, A. W., Canova-Davis, E., Stults, J. T., Hancock, W. S., Covey, T. R. et Shushan, B. I. (1991). "Characterization of the tryptic map of recombinant DNA derived tissue plasminogen activator by high-performance liquid chromatography-electrospray ionization mass spectrometry." *Anal Chem* 63(24): 2909-2915.

Link, A. J., Eng, J., Schieltz, D. M., Carmack, E., Mize, G. J., Morris, D. R., Garvik, B. M. et Yates, J. R. (1999). "Direct analysis of protein complexes using mass spectrometry." *Nat Biotech* 17(7): 676-682.

Liu, H., Finch, J. W., Lavalley, M. J., Collamati, R. A., Benevides, C. C. et 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." *Journal of Chromatography A* 1147(1): 30-36.

Ma, B., Zhang, K., Hendrie, C., Liang, C., Li, M., Doherty-Kirby, A. et Lajoie, G. (2003). "PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry." *Rapid Communications in Mass Spectrometry* 17(20): 2337-2342.

Mann, M., Højrup, P. et Roepstorff, P. (1993). "Use of mass spectrometric molecular weight information to identify proteins in sequence databases." *Biological Mass Spectrometry* 22(6): 338-345.

Mann, M. et Jensen, O. N. (2003). "Proteomic analysis of post-translational modifications." *Nat Biotechnol* 21(3): 255-261.

March, R. E. (2009). "Quadrupole ion traps." *Mass Spectrom Rev* 28(6): 961-989.

Mechref, Y., Muzikar, J. et Novotny, M. V. (2005). "Comprehensive assessment of N-glycans derived from a murine monoclonal antibody: a case for multimethodological approach." *Electrophoresis* 26(10): 2034-2046.

Mechref, Y. et Novotny, M. V. (2002). "Structural investigations of glycoconjugates at high sensitivity." *Chem Rev* 102(2): 321-369.

Medzihradzky, K. F., Besman, M. J. et Burlingame, A. L. (1997). "Structural characterization of site-specific N-glycosylation of recombinant human factor VIII by reversed-phase high-performance liquid chromatography-electrospray ionization mass spectrometry." *Anal Chem* 69(19): 3986-3994.

Meng, C. K., Mann, M. et Fenn, J. B. (1988). "Of protons or proteins." *Zeitschrift für Physik D Atoms, Molecules and Clusters* 10(2): 361-368.

Morelle, W., Canis, K., Chirat, F., Faid, V. et Michalski, J. C. (2006). "The use of mass spectrometry for the proteomic analysis of glycosylation." *Proteomics* 6(14): 3993-4015.

Morelle, W. et Michalski, J. C. (2007). "Analysis of protein glycosylation by mass spectrometry." *Nat Protoc* 2(7): 1585-1602.

Motoyama, A. et Yates, J. R., 3rd (2008). "Multidimensional LC separations in shotgun proteomics." *Anal Chem* 80(19): 7187-7193.

Nawarak, J., Phutrakul, S. et Chen, S. T. (2004). "Analysis of lectin-bound glycoproteins in snake venom from the *Elapidae* and *Viperidae* families." *J Proteome Res* 3(3): 383-392.

Nesvizhskii, A. I. et Aebersold, R. (2005). "Interpretation of shotgun proteomic data: the protein inference problem." *Mol Cell Proteomics* 4(10): 1419-1440.

O'Farrell, P. H. (1975). "High resolution two-dimensional electrophoresis of proteins." *Journal of Biological Chemistry* 250(10): 4007-4021.

O'Neill, R. A. (1996). "Enzymatic release of oligosaccharides from glycoproteins for chromatographic and electrophoretic analysis." *J Chromatogr A* 720(1-2): 201-215.

Ozohanics, O., Krenyacz, J., Ludanyi, K., Pollreisz, F., Vekey, K. et Drahos, L. (2008). "GlycoMiner: a new software tool to elucidate glycopeptide composition." *Rapid Commun Mass Spectrom* 22(20): 3245-3254.

Paizs, B. et Suhai, S. (2005). "Fragmentation pathways of protonated peptides." *Mass Spectrometry Reviews* 24(4): 508-548.

Pappin, D. J. C., Hojrup, P. et Bleasby, A. J. (1993). "Rapid identification of proteins by peptide-mass fingerprinting." *Current Biology* 3(6): 327-332.

Paul, W. et Steinwedel, H. (1960). US Patent 2 939 952.

Perkins, D. N., Pappin, D. J., Creasy, D. M. et Cottrell, J. S. (1999). "Probability-based protein identification by searching sequence databases using mass spectrometry data." *Electrophoresis* 20(18): 3551-3567.

Peterman, S. M. et Mulholland, J. J. (2006). "A novel approach for identification and characterization of glycoproteins using a hybrid linear ion trap/FT-ICR mass spectrometer." *J Am Soc Mass Spectrom* 17(2): 168-179.

Premstaller, A., Oberacher, H., Walcher, W., Timperio, A. M., Zolla, L., Chervet, J.-P., Cavusoglu, N., van Dorsselaer, A. et Huber, C. G. (2001). "High-Performance Liquid Chromatography Electrospray Ionization Mass Spectrometry Using Monolithic Capillary Columns for Proteomic Studies." *Analytical Chemistry* 73(11): 2390-2396.

Rabilloud, T., Vaezzadeh, A., R. , Potier, N., Lelong, C., Leize-Wagner, E. et Chevallet, M. (2009). "Power and limitations of electrophoretic separations in proteomics strategies." *Mass Spectrometry Reviews* 28(5): 816-843.

Rieux, L., Niederländer, H., Verpoorte, E. et Bischoff, R. (2005). "Silica monolithic columns: Synthesis, characterisation and applications to the analysis of biological molecules." *Journal of Separation Science* 28(14): 1628-1641.

Robinson, N. E. et Robinson, A. B. (2001). "Deamidation of human proteins." *Proc Natl Acad Sci U S A* 98(22): 12409-12413.

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

Roussel, X., Béchade, G., Kriznik, A., Van Dorsselaer, A., Sanglier-Cianferani, S., Branlant, G. et Rahuel-Clermont, S. (2008). "Evidence for the Formation of a Covalent Thiosulfinate Intermediate with Peroxiredoxin in the Catalytic Mechanism of Sulfiredoxin." *Journal of Biological Chemistry* 283(33): 22371-22382.

Rudd, P. M., Elliott, T., Cresswell, P., Wilson, I. A. et Dwek, R. A. (2001). "Glycosylation and the immune system." *Science* 291(5512): 2370-2376.

Sandra, K., Moshir, M., D'Hondt, F., Tuytten, R., Verleysen, K., Kas, K., François, I. et Sandra, P. (2009). "Highly efficient peptide separations in proteomics: Part 2: Bi- and multidimensional liquid-based separation techniques." *Journal of Chromatography B* 877(11-12): 1019-1039.

Sandra, K., Moshir, M., D'Hondt, F., Verleysen, K., Kas, K. et Sandra, P. (2008). "Highly efficient peptide separations in proteomics: Part 1. Unidimensional high performance liquid chromatography." *Journal of Chromatography B* 866(1-2): 48-63.

Satomi, Y., Shimonishi, Y. et Takao, T. (2004). "N-glycosylation at Asn(491) in the Asn-Xaa-Cys motif of human transferrin." *FEBS Lett* 576(1-2): 51-56.

Savitski, M. M., Nielsen, M. L., Kjeldsen, F. et Zubarev, R. A. (2005). "Proteomics-Grade de Novo Sequencing Approach." *Journal of Proteome Research* 4(6): 2348-2354.

Schluesener, D., Fischer, F., Kruij, J., Rögner, M. et Poetsch, A. (2005). "Mapping the membrane proteome of *Corynebacterium glutamicum*." *PROTEOMICS* 5(5): 1317-1330.

Searle, B. C., Turner, M. et Nesvizhskii, A. I. (2008). "Improving Sensitivity by Probabilistically Combining Results from Multiple MS/MS Search Methodologies." *Journal of Proteome Research* 7(1): 245-253.

Segu, Z. M. et Mechref, Y. (2010). "Characterizing protein glycosylation sites through higher-energy C-trap dissociation." *Rapid Communications in Mass Spectrometry* 24(9): 1217-1225.

Shen, Y., Smith, R. D., Unger, K. K., Kumar, D. et Lubda, D. (2005). "Ultrahigh-Throughput Proteomics Using Fast RPLC Separations with ESI-MS/MS." *Analytical Chemistry* 77(20): 6692-6701.

Shevchenko, A., Sunyaev, S., Loboda, A., Shevchenko, A., Bork, P., Ens, W. et Standing, K. G. (2001). "Charting the Proteomes of Organisms with Unsequenced Genomes by MALDI-Quadrupole Time-of-Flight Mass Spectrometry and BLAST Homology Searching." *Analytical Chemistry* 73(9): 1917-1926.

Sparbier, K., Koch, S., Kessler, I., Wenzel, T. et Kostrzewa, M. (2005). "Selective isolation of glycoproteins and glycopeptides for MALDI-TOF MS detection supported by magnetic particles." *J Biomol Tech* 16(4): 407-413.

Stafford Jr, G. C., Kelley, P. E., Syka, J. E. P., Reynolds, W. E. et Todd, J. F. J. (1984). "Recent improvements in and analytical applications of advanced ion trap technology." *International Journal of Mass Spectrometry and Ion Processes* 60(1): 85-98.

Steen, H. et Mann, M. (2004). "The ABC's (and XYZ's) of peptide sequencing." *Nat Rev Mol Cell Biol* 5(9): 699-711.

Sullivan, B., Addona, T. A. et Carr, S. A. (2004). "Selective detection of glycopeptides on ion trap mass spectrometers." *Anal Chem* 76(11): 3112-3118.

Sun, B., Ranish, J. A., Utleger, A. G., White, J. T., Yan, X., Lin, B. et Hood, L. (2007). "Shotgun glycopeptide capture approach coupled with mass spectrometry for comprehensive glycoproteomics." *Mol Cell Proteomics* 6(1): 141-149.

Syka, J. E. P., Coon, J. J., Schroeder, M. J., Shabanowitz, J. et Hunt, D. F. (2004). "Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry." *Proceedings of the National Academy of Sciences of the United States of America* 101(26): 9528-9533.

Tabb, D. L., Vega-Montoto, L., Rudnick, P. A., Variyath, A. M., Ham, A. J., Bunk, D. M., Kilpatrick, L. E., Billheimer, D. D., Blackman, R. K., Cardasis, H. L., Carr, S. A., Clauser, K. R., Jaffe, J. D., Kowalski, K. A., Neubert, T. A., Regnier, F. E., Schilling, B., Tegeler, T. J., Wang, M., Wang, P., Whiteaker, J. R., Zimmerman, L. J., Fisher, S. J., Gibson, B. W., Kinsinger, C. R., Mesri, M., Rodriguez, H., Stein, S. E., Tempst, P., Paulovich, A. G., Liebler, D. C. et Spiegelman, C. (2010). "Repeatability and reproducibility in proteomic identifications by liquid chromatography-tandem mass spectrometry." *J Proteome Res* 9(2): 761-776.

# RESULTATS





# **Partie 1 : Développements**

**Recherche de l'amélioration des méthodologies  
d'analyses LC-MS/MS  
pour l'identification et la caractérisation  
des protéines et des glycoprotéines**





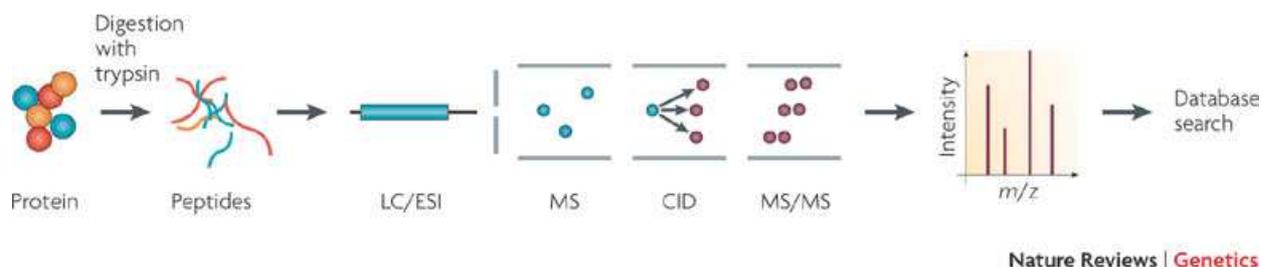
# Chapitre 1

## Développements de nouvelles méthodes pour l'évaluation des performances d'un couplage nanoLC-MS/MS

---

Le couplage nanoLC-MS/MS est aujourd'hui le système le plus utilisé pour les analyses protéomiques par spectrométrie de masse [Gstaiger et Aebersold 2009]. En effet, cette méthode est très sensible et permet d'obtenir des informations de séquence sur les peptides analysés et ainsi d'identifier les protéines de l'échantillon avec un degré de confiance important.

Une analyse protéomique par spectrométrie de masse comprend plusieurs étapes [Steen et Mann 2004, Gstaiger et Aebersold 2009] (Figure 1). L'échantillon biologique est digéré avec une endoprotéase (généralement de la trypsine) avant d'être analysé par nanoLC-MS/MS. Les spectres de fragmentation des peptides sont obtenus à l'aide du mode DDA (Data Dependant Acquisition) qui consiste à sélectionner les ions parents automatiquement en fonction de critères choisis par l'utilisateur. L'identification des protéines est déduite de la liste de peptides obtenus suite à l'interprétation des données par recherche dans une banque de données à l'aide d'un moteur de recherche. A chacune de ces étapes correspond un paramétrage complexe et qui aura une incidence sur le nombre de peptides et de protéines identifiés et sur le degré de confiance de ces identifications [Bell et al. 2009, Tabb et al. 2010].



**Figure 1.** Méthodologie la plus couramment utilisée en analyse protéomique par spectrométrie de masse, d'après [Gstaiger et Aebersold 2009]

Lorsque l'on effectue des analyses protéomiques, les échantillons donnés par les biologistes sont très souvent en petite quantité et/ou ne peuvent être fournis qu'une seule fois. Dans ce cas, seulement une ou deux analyses de LC-MS/MS pourront être réalisées. Il faut donc que le système LC-MS/MS soit paramétré de façon à avoir la sensibilité la plus élevée mais aussi pour garder la capacité de détecter des petits peptides (moins de 1000 Da), des peptides lourds (plus de 1500 Da) et des glycopeptides (plus de 2500 Da). Ce compromis ne peut être obtenu qu'à condition de comprendre l'incidence de tous les paramètres aussi bien de la partie chromatographique que de ceux du spectromètre de masse.

Dans ce chapitre, nous nous sommes donc intéressés au paramétrage du couplage nanoLC-MS/MS tant au niveau chromatographique, que du spectromètre de masse. En effet, un des problèmes rencontrés de manière récurrente concerne l'évaluation des performances nanoLC-MS/MS. Nous sommes souvent en "limite" de performances pour les analyses effectuées sur ce système, et avant de réaliser une expérience, il est absolument primordial de savoir si le couplage utilisé offre ses capacités maximales avec les paramètres sélectionnés pour l'échantillon à analyser.

A mon arrivée au laboratoire un couplage nanoLC-MS/MS entre une chaîne nanoHPLC 1100 (Agilent Technologies) et une trappe ionique HCT Ultra (Bruker Daltonics) était déjà en place avec une source nanospray Dual Nanospray ion source (Agilent Technologies). Peu de temps après le laboratoire a fait l'acquisition d'un système nanoLC-Chip/MS qui a remplacé l'interface Dual Nanospray.

Une partie de mon travail de thèse a consisté à développer et optimiser une méthode permettant d'évaluer les performances de ce nouveau système nano-chromatographique intégré sur puce microfluidique nanoLC-Chip en couplage avec la trappe ionique HCT Ultra.

Nous allons d'abord présenter l'impact des paramètres chromatographiques (§ 1 et 2), puis ceux des cycles d'acquisition des spectres MS/MS en mode automatique (§ 3), suivi des tests mis en place pour évaluer les performances du système complet (§ 4) et enfin les nouvelles possibilités des instruments récents (§ 5).

## 1. Description du système nanoLC-Chip/HCT Ultra

Un couplage nanoLC-MS est composé de :

- Une pompe nano-chromatographique qui délivre le gradient;
- Une pompe auxiliaire qui permet de charger l'échantillon sur la pré-colonne;
- Une pré-colonne d'enrichissement qui permet de concentrer l'échantillon et de le dessaler;
- Une colonne analytique sur laquelle est effectuée la séparation chromatographique;
- Une source nanospray qui permet l'ionisation des peptides pour leur analyse par spectrométrie de masse;
- Un spectromètre de masse.

Ces différentes composantes sont reliées entre elles à l'aide de plusieurs capillaires, vannes, etc. Ces connections constituent le point faible de ce type de système. En effet, chacune d'entre elles génèrent des volumes morts qui provoquent une dégradation de la résolution chromatographique et donc une perte de sensibilité. De plus chaque connexion est une source potentielle de fuite. Enfin, les aiguilles de nébulisation utilisées avec les sources nanospray on-line sont fragiles, elles ont des durées de vie très variables, et sont délicates à mettre en place. Ainsi, il n'est pas rare d'observer des pertes de spray aléatoires, ce qui est fortement handicapant pour l'analyse de longues séries d'échantillons.

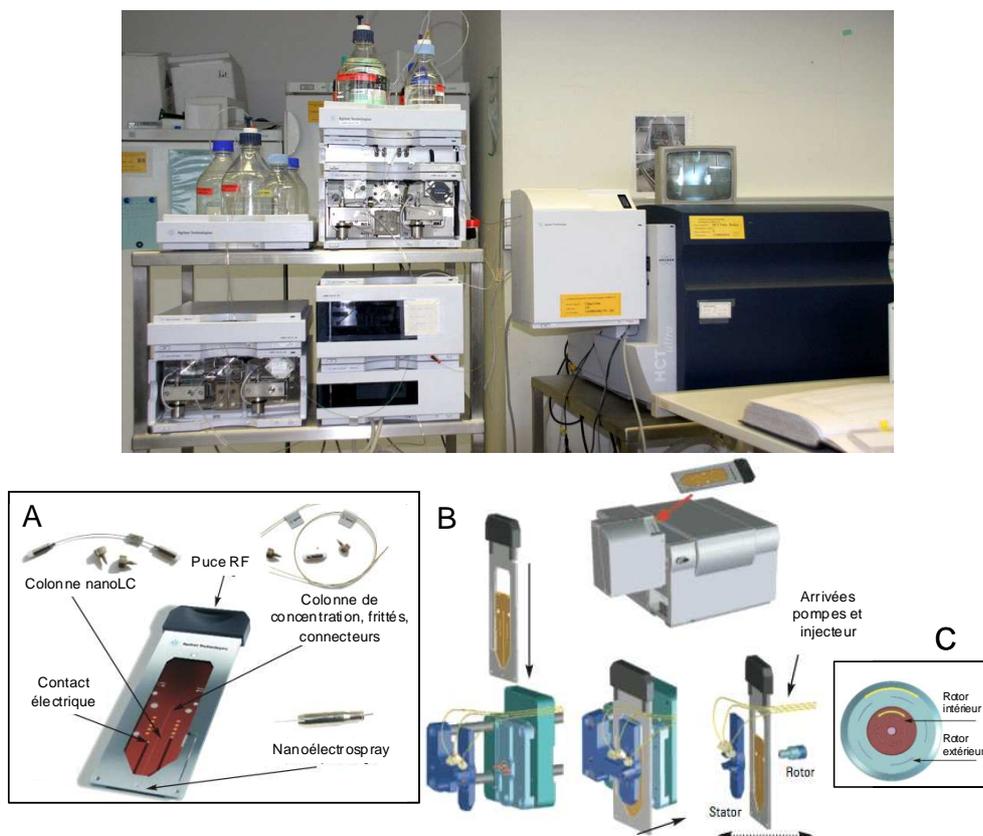
Ainsi, afin de palier à ces problèmes, un système de nano chromatographie intégrée sur puce microfluidique a été développé [Yin et al. 2004, Yin et Killeen 2007], et commercialisé par Agilent Technologies sous le nom de nanoHPLC-Chip/MS. Ce système est constitué :

- D'une puce microfluidique que nous appellerons chip dans la suite de ce manuscrit, qui comprend la pré-colonne, la colonne analytique ainsi que l'aiguille nanospray (Figure 2A);
- D'un cube qui sert d'interface entre la partie pompe chromatographique et le spectromètre de masse.

La chip est introduite dans l'interface et s'ajuste automatiquement entre un rotor et un stator présents à l'intérieur du cube (Figure 2B). La pression appliquée par le rotor et le stator sur la chip permettent d'assurer l'étanchéité du système. Néanmoins, il est possible d'observer des fuites à ce niveau si le rotor est sale ou si la pression des solvants est trop importante (> 200 bars).

Le rotor est constitué de deux parties (Figure 2C). La partie intérieure est utilisée pour des analyses monodimensionnelles c'est à dire avec des chip contenant une seule colonne analytique. La partie extérieure du rotor a été développée pour l'utilisation de chip contenant plusieurs colonnes. Par exemple Bynum et al utilisent une chip possédant une cellule contenant une enzyme (PNGase F) permettant de déglycosyler un anticorps, suivi d'une colonne remplie de phase inverse C8 qui retient

la protéine déglycosylée, et d'une colonne contenant une phase stationnaire carbone graphite pour séparer chromatographiquement les glycanes purifiés [Bynum et al. 2009].



**Figure 2.** Configuration de l'interface nanoHPLC-Chip/MS. L'ensemble des éléments nécessaires pour le système nanoHPLC classique est regroupé dans la chip : la pré-colonne d'enrichissement, la colonne analytique et l'aiguille nanospray (A). Disposée de manière automatisée en position orthogonale face à l'électrode d'entrée du spectromètre de masse (B), la chip est enserrée entre un stator et un rotor composé d'une double microvanne permettant la distribution des fluides sans connexion (C)

Dans ce chapitre nous traiterons de l'utilisation de chip contenant des colonnes analytiques de 43 ou 150 mm de long et 75  $\mu\text{m}$  de diamètre, avec des pré-colonnes de 40 nl et une phase stationnaire ZORBAX 300SB-C18 avec des particules de 5  $\mu\text{m}$ .

Le système nanoHPLC-Chip est couplé à une trappe ionique tridimensionnelle (HCT Ultra, Bruker Daltonics). Cet instrument a déjà été décrit dans une thèse effectuée au laboratoire [Carapito 2006].

## 2. Optimisation des conditions chromatographiques

### 2.1. Optimisation des méthodes chromatographiques

Une première optimisation du système nanoLC-Chip/MS a déjà été décrite dans une thèse réalisée au laboratoire [Carapito 2006]. Ce système ayant évolué (amélioration de la résolution des colonnes, capacité des colonnes de chargement, meilleure stabilité du nanospray, évolution des performances du spectromètre de masse), il a été nécessaire de reprendre l'étude de l'influence des principaux paramètres (débit, gradient, nature du modifieur organique de la phase mobile, etc.).

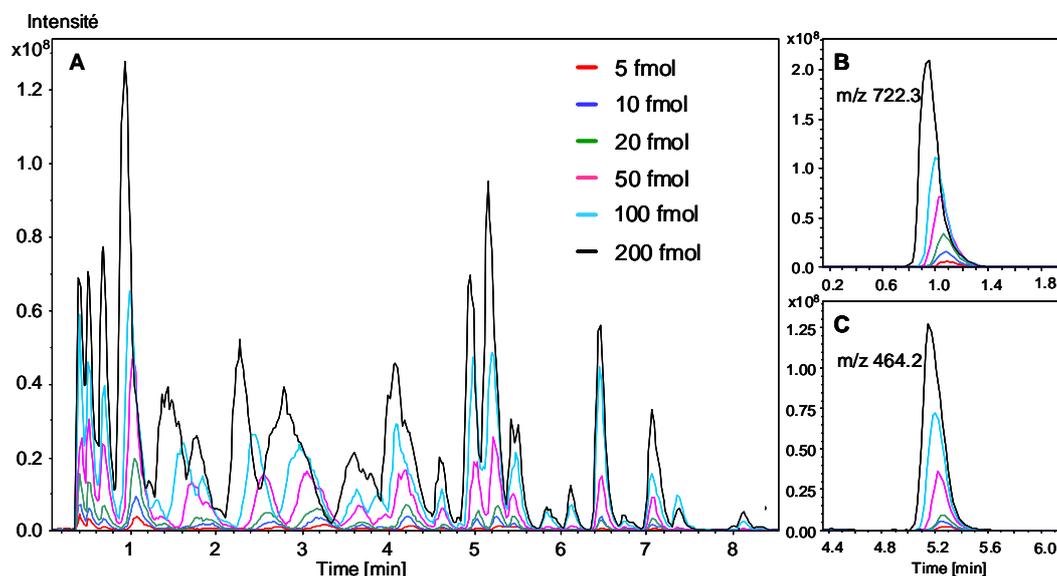
Notre objectif était d'obtenir des données MS/MS aussi informatives que possible grâce à une optimisation du système pour chaque type de colonne (meilleurs choix pour la largeur des pics et la capacité de pics). En effet, les paramètres chromatographiques et ceux du spectromètre de masse (cycle d'acquisition automatisées) sont étroitement liés, ce qui a une incidence très importante sur la quantité et la qualité des données MS/MS, et donc sur le nombre de peptides et protéines identifiés et sur le degré de confiance de ces identifications.

Le travail présenté dans ce chapitre est donc une étape essentielle pour comprendre comment obtenir les meilleures données expérimentales en LC-MS/MS.

Plusieurs débits ont été testés. C'est à 300nl/min que la meilleure résolution chromatographique et la meilleure séparation ont été obtenues. De plus, ce débit permet de générer des contrepressions plus basses (~40 bars pour les colonnes de 43 mm et ~60 bars pour les colonnes de 150 mm).

Les peptides tryptiques sont élués pour la grande majorité entre 10% et 40% d'acétonitrile. Plusieurs gradients linéaires ont donc été testés en injectant un digestat tryptique de BSA. Pour les colonnes de 43 mm un gradient linéaire de 7 min entre 10% et 40% d'acétonitrile permet une bonne séparation et une bonne résolution chromatographique (10 – 15 s à mi-hauteur). Pour les colonnes de 150 mm la même largeur de pic est obtenue avec un gradient linéaire à 1% (10% à 40% d'acétonitrile en 30 min).

La système nanoLC-Chip-MS présente une excellente répétabilité et des tests de sensibilité ont permis de montrer qu'il était possible d'observer des ions correspondants à des peptides de BSA pour seulement 5 fmol injectés (Figure 3).



**Figure 3.** LC/MS correspondants à des quantités croissantes injectées d'un digestat tryptique de BSA (A); extractions des chromatogrammes des ions m/z 722,3 (B) et m/z 464,2 (C)

## 2.2. Choix du modifieur organique : Optimisation d'une méthode nanoLC-MS utilisant le méthanol comme modifieur organique de la phase mobile

L'acétonitrile est le solvant organique le plus utilisé comme phase mobile pour l'analyse de protéines et de peptides par RPLC-MS. En 2009, le prix de ce solvant a très fortement augmenté à cause d'une pénurie mondiale provoquée par la crise économique [Tullo 2008]. L'acétonitrile étant un sous-produit de la synthèse de l'acrylonitrile qui est un polymère utilisé dans l'industrie automobile, sa production a directement subi la récession économique. Cette situation nous a obligés à remettre en question notre consommation en solvant et plus particulièrement en ce qui concerne le système nanoLC-Chip/MS. En effet, bien que le débit appliqué sur la colonne soit de seulement 300 nL/min, le débit primaire des pompes est de 200 à 500 µL/min. Ce système est donc très consommateur en solvant.

Afin de pallier à ce problème nous avons choisi de tester une méthode alternative et de remplacer l'acétonitrile (AcN) par le méthanol (MeOH) sur un de nos couplages nanoLC-Chip/MS. Le méthanol possédant des propriétés physicochimiques différentes de l'AcN, une méthode analytique a du être adaptée en prenant en compte les spécificités de ce solvant [Giorgianni et al. 2004, Fritz et al. 2009] :

- Le méthanol possède une viscosité plus importante que l'AcN. La contrepression générée par son utilisation comme phase mobile est donc plus importante. Ce point n'est pas négligeable car la pression est l'un des facteurs limitant du système nanoLC-Chip/MS (maximum 150-200 bars);

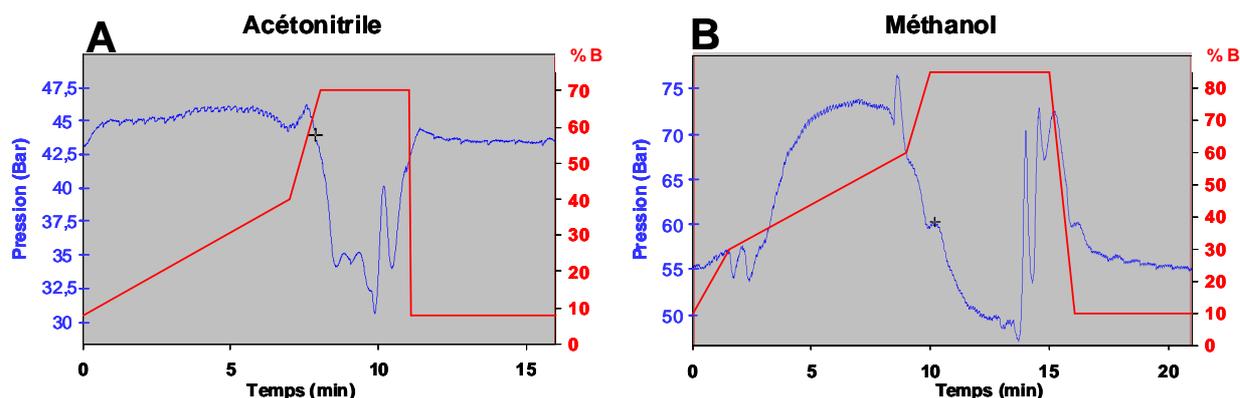
- Le méthanol possède un pouvoir éluant moindre que l'acétonitrile. Le gradient doit donc être optimisé en tenant compte de cette propriété qui risque de rallonger les temps de rétention et donc les temps d'analyse;
- La sélectivité du méthanol est différente de celle de l'acétonitrile, le gradient doit donc être optimisé afin de conserver voire d'améliorer la qualité de la séparation chromatographique.

### 2.2.1. Conditions expérimentales

- Echantillon modèle : 50 fmol d'un digestat tryptique de BSA (Bovin Serum Albumin)
- Chip utilisées :
  - Phase stationnaire : ZORBAX 300SB-C18 particules de 5  $\mu\text{m}$
  - Colonne analytique : 75  $\mu\text{m}$  de diamètre, 43 mm ou 150 mm de long
  - Pré-colonne de 40nl
- Pompe nano :
  - Solvants :
    - Voie A : 98%  $\text{H}_2\text{O}$ , 2% AcN ou MeOH, 0,1% HCOOH
    - Voie B : 98% AcN ou MeOH, 2%  $\text{H}_2\text{O}$ , 0,1% HCOOH
  - Débit après split: 300 nl/min
  - Débit primaire (avant split) : 200 à 500 nl/min
- Pompe de chargement (pompe capillaire):
  - Solvant : 98%  $\text{H}_2\text{O}$ , 2% AcN ou MeOH, 0,1% HCOOH
  - Débit après split: 3,75  $\mu\text{l}/\text{min}$
  - Débit primaire (avant split) : 200 à 500 nl/min

### 2.2.2. La contrepression

La viscosité du méthanol (0,55 cP) est plus grande que celle de l'acétonitrile (0,38 cP). Par conséquent, la contrepression générée lorsque le méthanol est utilisé en HPLC comme solvant organique dans la phase mobile est plus importante qu'avec l'acétonitrile. De plus la pression subit des variations très importantes en fonction de la composition méthanol/eau (Figure 4B). En effet, des liaisons hydrogènes se forment entre les deux solvants et augmentent ainsi la viscosité du mélange pour atteindre un maximum vers une composition de 50% de méthanol. Le système nanoLC-Chip/MS ne comporte pas de système de chauffage de la colonne analytique, la pression engendrée par le solvant est donc un facteur critique qui doit être pris en compte.



**Figure 4.** Tracés de la pression mesurée au niveau de la pompe nano-chromatographique (bleu) en fonction du gradient (rouge) pour 2 analyses sur chip de 43 mm avec l'utilisation d'acétonitrile comme solvant organique (A) ou de méthanol (B)

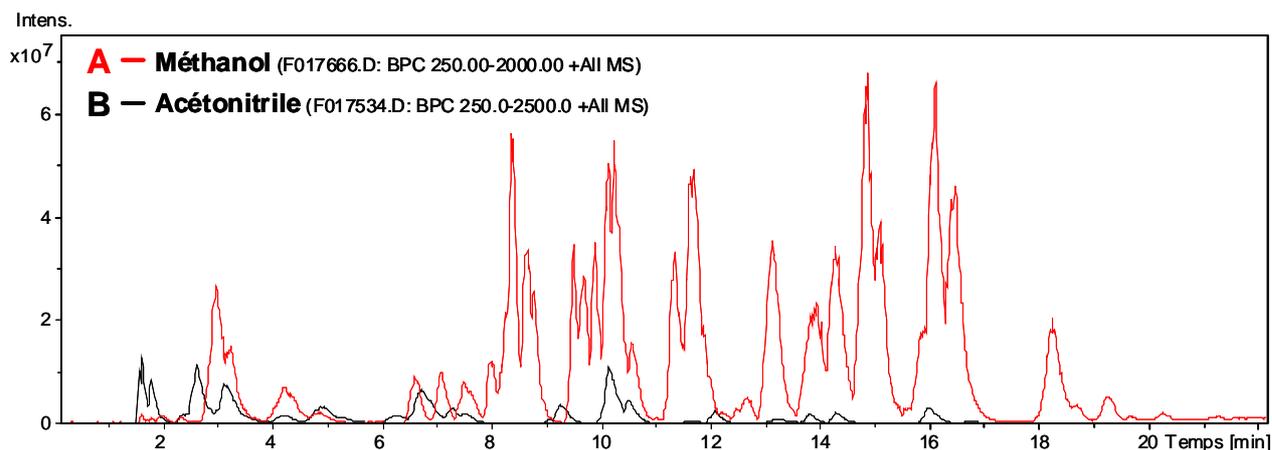
La Figure 4 montre la contrepression générée lorsqu'un gradient avec une pente d'environ 4% est appliqué sur une chip contenant une colonne de 43 mm de long, pour l'acétonitrile (Figure 4A) et pour le méthanol (Figure 4B). En utilisant le méthanol comme solvant organique, la contrepression appliquée au niveau de la pompe nano-chromatographique est plus élevée d'environ 15 bars dans les conditions initiales par rapport à celle obtenue avec l'acétonitrile. De plus, le gradient provoque une amplitude de pression de 30 bars avec le méthanol pour seulement 15 bars avec l'acétonitrile. Cependant, la pression maximale obtenue lorsque la phase mobile est composée de 50% d'eau et de 50% de méthanol est de 75 bars, ce qui reste compatible avec l'utilisation du système nanoLC-Chip/MS.

L'observation du tracé de la pression permet également de constater que l'équilibration de la colonne prend plus de temps lorsque l'on utilise le méthanol. L'augmentation de l'inertie du système doit donc être prise en compte pour l'optimisation du gradient (§ 2.2.4).

### 2.2.3. La sensibilité

Afin de comparer les intensités chromatographiques en fonction du solvant, 50 fmol d'un digestat tryptique de BSA sont analysés sur la même chip (150 mm) et avec la même méthode analytique. La seule différence est la nature du solvant organique (acétonitrile ou méthanol).

Ces analyses permettent d'observer un gain d'intensité des pics chromatographiques d'un facteur de 5 à 10 lorsque l'on utilise le méthanol (Figure 6). Ce phénomène est également observable pour des analyses sur colonne courte (43 mm) et peut-être s'expliquer par une amélioration de la désolvatation des ions au niveau de l'interface nanospray [Giorgianni et al. 2004]. Si l'on compare les deux analyses servant d'exemple pour la Figure 6, lorsque le méthanol est utilisé, le bruit de fond est légèrement plus intense (facteur 2,5) mais les rapports signal/bruit sont largement supérieurs pour les 2 ions de référence.



**Figure 5.** NanoLC-MS de 50 fmol d'un digestat trypsique de BSA avec une colonne de 150 mm; BPC (Base Peak Chromatogramme) en utilisant le méthanol (A); BPC en utilisant l'acétonitrile (B)

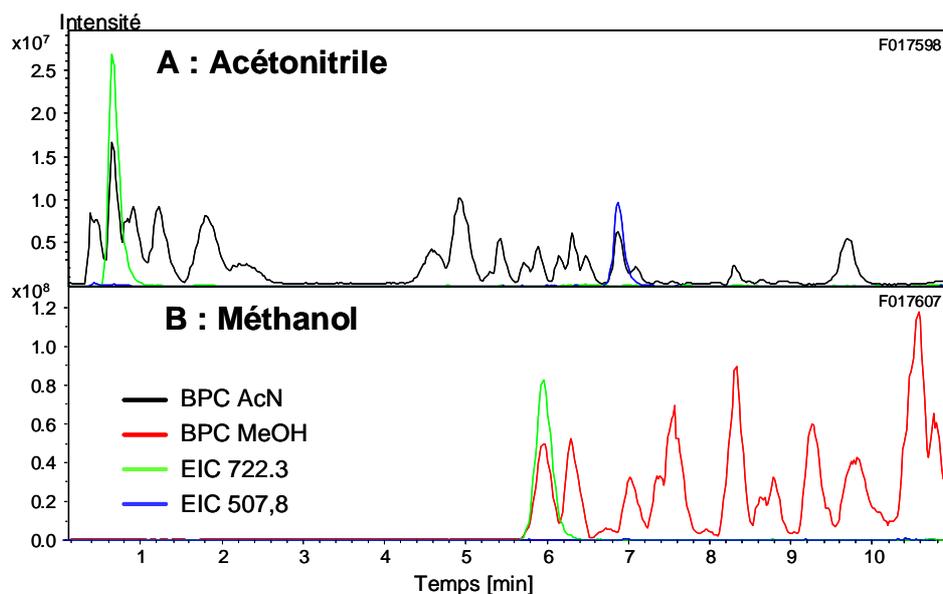
Ce gain d'intensité s'accompagne d'une augmentation du pourcentage de recouvrement lorsque l'on effectue une analyse nanoLC-MS/MS. La recherche dans la banque protéique SwissProt à l'aide du moteur de recherche Mascot permet d'identifier la BSA avec un recouvrement moyen de 75% lorsque le méthanol est utilisé, au lieu de 55% pour une analyse avec de l'acétonitrile (colonne longue, 150 mm).

#### 2.2.4. Pouvoir éluatif

La Figure 6 permet d'observer les chromatogrammes obtenus pour les analyses nanoLC-MS d'un digestat trypsique de BSA sur une colonne courte (43 mm) avec des gradients identiques (8 à 40% de B en 7 min). La Figure 6A correspond à l'utilisation d'acétonitrile comme solvant organique dans la phase mobile, et la Figure 6B à l'utilisation de méthanol. Avec l'acétonitrile le peptide correspondant à l'ion  $m/z$  722,3 a un temps de rétention de 0,7 min alors qu'avec le méthanol, il est élué à 6 min. On observe également que dans ces conditions chromatographiques, l'ion  $m/z$  507,8 n'est pas élué lorsque l'on utilise le méthanol.

Ces décalages importants des temps de rétention peuvent être occasionnés par deux phénomènes :

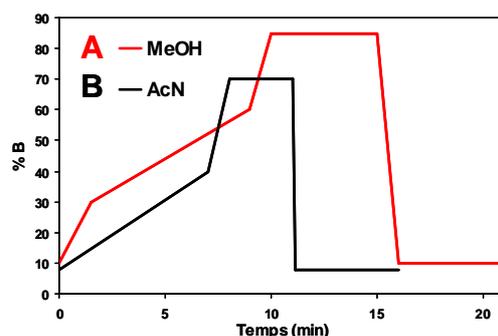
- Le méthanol a un pouvoir éluatif moins important que l'acétonitrile. L'éluion d'un même peptide demande un pourcentage de solvant organique plus grand avec le méthanol qu'avec l'acétonitrile;
- Le système possède une inertie plus importante lorsque l'on utilise le méthanol (§ 2.2.2). Par conséquent, à un temps  $t$ , la consigne donnée ne correspond pas à la composition réelle en solvants.



**Figure 6.** NanoLC-MS de 50 fmol d'un digestat trypsique de BSA avec un gradient allant de 8 à 40 % de B en 7 min sur une colonne de 43mm; en utilisant de l'acétonitrile (A) comme solvant organique ou du méthanol (B)

Le pouvoir éluatif moindre du méthanol ainsi que l'inertie provoquée par son utilisation comme solvant organique m'ont amené à effectuer quelques ajustements au niveau du gradient (Figure 7) :

- Alors qu'avec l'acétonitrile le pourcentage maximal en solvant organique était de 40 % en fin de gradient, il est nécessaire de monter jusqu'à 60% avec le méthanol pour éluer l'ensemble des peptides;
- Afin d'éluer un maximum de contaminants pouvant rester fixés sur la colonne, un palier à 85% de B est établi en fin de gradient au lieu de 70% avec l'acétonitrile;
- La séparation chromatographique étant satisfaisante, nous avons choisi de garder la même pente de gradient pour les colonnes courtes (4,5 %);
- Très peu de peptides sont élués avant 6 min (consigne 30% méthanol). Ainsi, afin de raccourcir le temps d'analyse la pente du gradient est augmentée (13%) pour atteindre rapidement les 30% de méthanol;
- L'inertie du système étant plus importante avec le méthanol (§ 2.2.2), il est nécessaire de rallonger la durée d'enregistrement des données ainsi que le temps d'équilibration de la colonne.

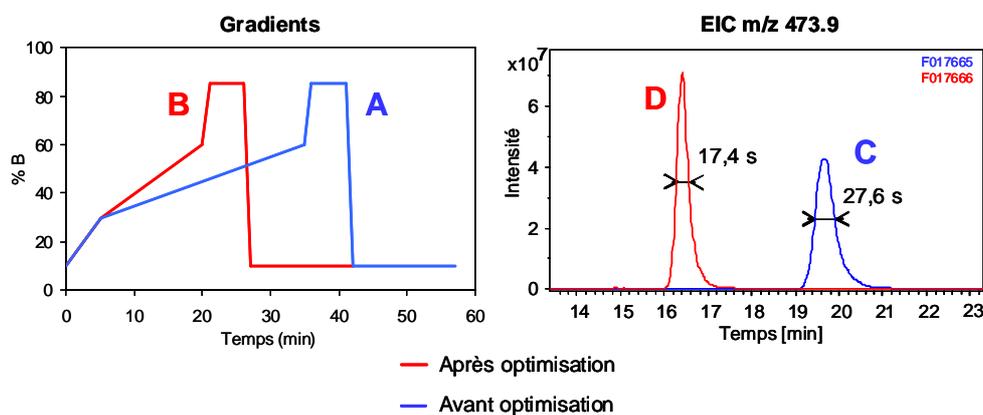


**Figure 7.** Schéma du gradient optimisé pour des colonnes de 43 mm en utilisant le méthanol comme solvant organique (A); gradient utilisé avec l'acétonitrile comme solvant organique (B)

### 2.2.5. Résolution chromatographique

Un gradient est développé pour effectuer des analyses avec une colonne longue (150 mm) en prenant comme modèle celui réalisé pour les colonnes courtes (43 mm) (Figure 7).

Suivant le même principe le gradient débute par une montée rapide en solvant organique suivie d'une pente identique à celle utilisée avec l'acétonitrile (1%) (Figure 8A). L'utilisation de ce gradient provoque un élargissement des pics en fin d'analyse (Figure 8C). Afin d'améliorer la résolution chromatographique, la pente du gradient est accentuée à 2% (Figure 8B). L'augmentation de la pente du gradient permet une diminution de la largeur des pics à mi-hauteur d'un facteur 1,5 (Figure 8D). Ce gradient plus rapide a également pour avantage de raccourcir la durée des analyses.



**Figure 8.** Gradient avant (A) et après optimisation (B); extraction des chromatogrammes de l'ion m/z 473,9 avant optimisation du gradient (C) et après (D)

### 2.3. Discussion et conclusion

Nous avons montré qu'en optimisant les conditions chromatographiques, l'utilisation du méthanol en remplacement de l'acétonitrile comme solvant organique n'altère pas la qualité de la séparation et de la résolution chromatographique et permet même une réduction des temps d'analyse pour les colonnes de 150 mm.

Le méthanol est un solvant moins fort que l'acétonitrile, les peptides sont élués à des compositions en solvant organique plus importantes. Les gouttelettes formées dans la source nano-spray possèdent donc une tension de surface plus faible. L'évaporation du solvant est alors facilitée et la dessolvation des ions est meilleure [Giorgianni et al. 2004], ce qui se traduit par un gain de sensibilité d'un facteur 5 à 10 sans avoir changé les paramètres de la source par rapport à ceux utilisés pour l'acétonitrile.

Dans notre cas, la seule limitation vient de la contrepression provoquée par le méthanol qui est plus importante qu'avec l'acétonitrile tout en restant dans les limites acceptables par le système chromatographique.

Plusieurs études traitent du remplacement de l'acétonitrile par d'autres solvants comme le méthanol ou l'acétone. Leurs conclusions concernant les performances obtenues par l'utilisation du méthanol sont assez discordantes.

Giorgianni et al [Giorgianni et al. 2004] ont analysé par nanoLC-MS et MS/MS différents digestats tryptiques de protéines en comparant l'utilisation du méthanol et de l'acétonitrile comme solvant organique, et leurs conclusions vont dans le même sens que les nôtres. Après optimisation du gradient, la séparation et la résolution chromatographique sont semblables pour les deux solvants. Ils observent également un net gain en sensibilité et un abaissement de la limite de détection d'un facteur 5.

Fritz et al comparent eux, l'acétonitrile, l'acétone et le méthanol en utilisant une colonne de 2,1 mm avec un débit de 150  $\mu\text{L}/\text{min}$  et une source ESI [Fritz et al. 2009]. Ils n'utilisent donc pas comme nous un système nano-chromatographique et une source nano-spray. Ils analysent un digestat tryptique de cinq protéines modèles en utilisant les mêmes conditions chromatographiques pour les trois solvants. Ils observent que l'utilisation du méthanol augmente les temps de rétention des peptides. En effet, le méthanol étant un solvant plus faible que l'acétonitrile, nous avons montré que pour obtenir des performances chromatographiques semblables, les conditions chromatographiques doivent être adaptées en utilisant un gradient plus rapide et une composition plus importante en solvant organique. De plus leur travail montre une diminution du nombre de peptides identifiés et des pourcentages de recouvrement. Étant donné que les conditions chromatographiques ne sont pas optimisées et que la source utilisée est de l'ESI et non du nano-spray, ces résultats sont difficilement comparables aux nôtres.

Cette étude montre en revanche que l'utilisation de l'acétone comme solvant organique permet une augmentation des pourcentages de recouvrement et une diminution des temps de rétentions par rapport à l'acétonitrile. Ces résultats vont dans le même sens que ceux obtenus par Keppel et al [Keppel et al. 2010]. L'acétone possède la même viscosité que l'acétonitrile, son utilisation génère

donc une contrepression comparable. De plus, malgré un léger élargissement des pics chromatographiques, l'acétone apporte un gain en sensibilité.

Ainsi, nous avons montré que pour notre système nanoLC-Chip/MS, le remplacement de l'acétonitrile par le méthanol a permis un gain en sensibilité et une augmentation du pourcentage de recouvrement sur la BSA. L'utilisation de l'acétone pourrait également être une alternative, il est cependant nécessaire de vérifier la compatibilité de la chaîne chromatographique et du polymère constituant les chip, avec ce solvant.

## **3. Optimisation des paramètres de la trappe ionique**

### **3.1. Optimisation de l'interface**

#### **3.1.1. La source**

Le couplage du système nanoLC-Chip/MS avec la trappe HCT Ultra a nécessité l'optimisation de cette nouvelle interface et plus particulièrement au niveau de la génération du spray. En effet, afin d'éviter un encrassement trop rapide de l'interface de la trappe ionique qui provoque une chute importante de sensibilité, le spray doit être parfaitement stable et ne doit pas entrer directement dans le spectromètre de masse. Pour cela, deux facteurs sont à prendre en compte : la position de l'aiguille nano-spray et la tension appliquée au niveau du capillaire de transmission. Ces deux paramètres sont à adapter en fonction de la chip utilisée et de l'état de l'aiguille. En effet, l'aiguille nano-spray est intégrée à la chip et vieillie avec elle.

#### **3.1.2. La transmission**

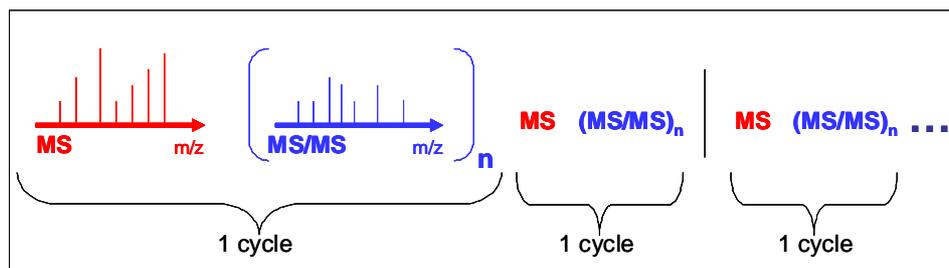
Dans le cadre d'analyses protéomiques, on considère que les peptides ont une masse moyenne de  $m/z$  750. Les paramètres de transmission des ions de la source vers l'analyseur sont optimisés pour cette masse. Pour cela, l'instrument possède une option appelée "Smart Parameter Setting" (SPS). Cette option permet de régler automatiquement les tensions appliquées au niveau des différents octopoles et lentilles et ainsi d'obtenir la transmission optimale pour une masse "cible" donnée. Le mode SPS est particulièrement utile pour optimiser rapidement une méthode afin d'analyser des ions dans une gamme de masse donnée. Par exemple, pour l'analyse de glycopeptides qui possèdent généralement une masse plus élevée que des peptides non glycosylés.

Il existe aussi une option SPS dans le cadre de l'auto-MS/MS. Celle-ci permet d'optimiser automatiquement la transmission pour l'ion parent sélectionné et ainsi d'améliorer la qualité des spectres MS/MS.

## **3.2. Optimisation des paramètres d'acquisition en mode MS/MS automatique**

### **3.2.1. Les cycles d'acquisition en mode MS/MS automatique**

Le mode d'acquisition MS/MS automatique (DDA) est utilisé pour les analyses LC-MS/MS. Il permet de sélectionner des ions parents de manière automatique en fonction de critères optimisés par l'utilisateur. L'acquisition fonctionne par cycles qui comprennent une MS permettant de sélectionner le ou les ions parents à fragmenter suivi d'une ou de plusieurs MS/MS de ces ions (Figure 9).



**Figure 9.** Fonctionnement par cycles du mode d'acquisition MS-MS/MS automatique : acquisition d'un spectre MS sur lequel sont sélectionnés le ou les ions parents avant d'être isolé(s) puis fragmenté(s) pour la MS/MS [Carapito 2006]

En protéomique, le but d'une analyse LC-MS/MS est d'identifier un maximum de protéines avec le meilleur pourcentage de recouvrement possible. Pour cela, il est nécessaire d'obtenir le plus de séquences peptidiques possibles, donc d'avoir le plus de spectres de fragmentation MS/MS avec une qualité suffisante pour leur interprétation.

Dans nos conditions d'analyse, les pics chromatographiques mesurent entre 10 et 15 s de large à mi-hauteur. Afin d'obtenir un maximum d'informations, il est nécessaire d'adapter la durée des cycles à la largeur des pics chromatographiques et à la complexité de l'échantillon. La durée d'un cycle est régie par plusieurs paramètres qui sont décrits dans la suite de ce chapitre.

L'optimisation du mode d'acquisition en MS/MS automatique sur une trappe HCT Ultra a déjà été décrite dans une thèse effectuée au laboratoire [Carapito 2006]. Dans ce chapitre ne sont décrits que les paramètres qui ont du être modifiés suite à l'acquisition du système nanoLC-Chip/MS. Ce système permet d'obtenir des pics chromatographiques plus fins. Par conséquent, la durée des cycles MS-MS/MS a du être optimisée en conséquence.

### 3.2.2. Conditions expérimentales

- Echantillon modèle : 50 fmol d'un digestat de 4 protéines dans des quantités équimolaires :
  - Alcool Deshydrogénase de levure (ADH);
  - Glycogène Phosphorylase B de lapin (GPB);
  - Sérum Albumine Bovine (BSA);
  - Enolase 1 de levure (ENO)

#### Conditions chromatographiques :

- Chip utilisées :
  - Phase stationnaire : ZORBAX 300SB-C18 particules de 5  $\mu\text{m}$
  - Colonne analytique : 75  $\mu\text{m}$  de diamètre, 43 mm de long
  - Pré-colonne de 40nl

- Pompe nano :
  - Solvants :
    - Voie A : 98% H<sub>2</sub>O, 2% AcN, 0,1% HCOOH
    - Voie B : 98% AcN, 2% H<sub>2</sub>O, 0,1% HCOOH
  - Débit : 300 nl/min
- Pompe de chargement :
  - Solvant : 98% H<sub>2</sub>O, 2% AcN, 0,1% HCOOH
  - Débit : 3,75 µl/min

#### Paramètres standards de l'auto MS/MS :

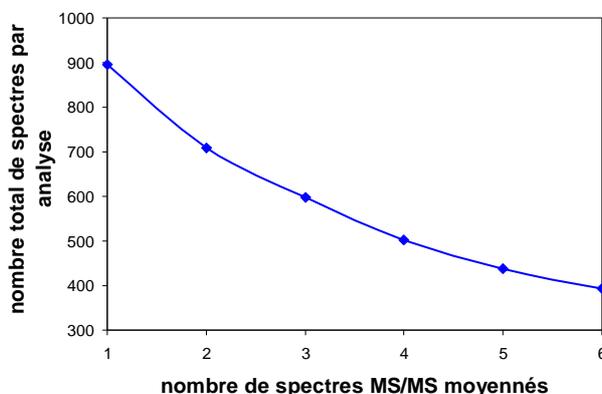
- Nombre de parents par MS sélectionnés pour la fragmentation MS/MS : 3
- Durée de l'exclusion active : 12 secondes
- Nombre de spectres MS/MS moyennés : 3

#### Conditions expérimentales :

Les conditions chromatographiques sont identiques pour l'ensemble des analyses. Chaque analyse a été réalisée 3 fois.

### 3.2.3. Le nombre de spectres moyennés

Sur les instruments de type trappe ionique, chaque spectre apparaissant à l'écran est un spectre moyen de plusieurs spectres dont le nombre est défini par l'utilisateur. Plus le nombre de spectres moyennés est grand et plus le spectre résultant est représentatif du signal étudié (le rapport signal/bruit de chacun des pics est accru). Cependant la multiplication du nombre de spectres moyennés augmente le temps d'acquisition et par conséquent diminue le nombre de spectres par analyse (Figure 10). Il faut donc trouver un compromis satisfaisant à la fois nos conditions d'analyse chromatographique et la qualité de la représentation du signal.



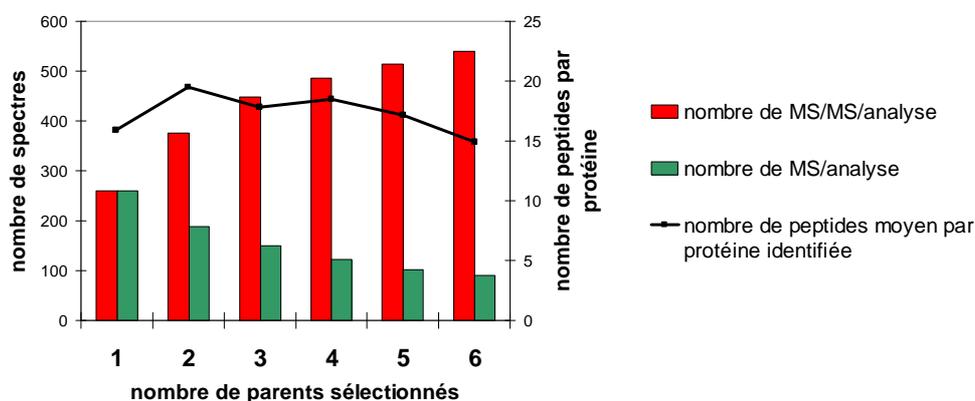
**Figure 10.** Nombre de spectres par analyse en fonction du nombre de spectres MS/MS moyennés pour une analyse nanoLC-MS/MS sur une chip de 43 mm

Le nombre de spectres MS/MS moyennés est adapté en fonction de la complexité de l'échantillon mais également en fonction de la qualité de spectres que l'on souhaite obtenir. Par exemple, une interprétation par *de novo* nécessite des spectres avec un meilleur rapport signal/bruit que si un moteur de recherche est utilisé.

Le but de cette optimisation étant de développer une méthode servant de standard pour l'analyse d'un échantillon peu complexe (digestat de BSA) le nombre de moyenne a été fixé à 3 pour les MS/MS.

### 3.2.4. Nombre de parents sélectionnés pour être fragmentés en MS/MS

La durée d'un cycle dépend du nombre de parents sélectionnés (Figure 9). Comme le montre la Figure 11, augmenter le nombre de parents permet d'augmenter proportionnellement le nombre de spectres MS/MS. Pourtant, le nombre de peptides identifiés par protéine a tendance à diminuer si le nombre de parents sélectionnés est trop important (> 4 parents). Le nombre optimal de parents est dépendant de la largeur des pics chromatographiques et peut être adapté en fonction de la complexité de l'échantillon.



**Figure 11.** Nombre de spectres MS et MS/MS en fonction du nombre de parents sélectionnés pour être fragmentés en MS/MS et nombre moyen de peptides identifiés par protéine; données issues d'analyses nanoLC-MS/MS de 50 fmol d'un digestat trypsique de 4 protéines en quantités équivalentes (ADH, GPB, BSA et ENO)

La vitesse de balayage n'est pas la même en MS et en MS/MS. En effet, avec une trappe ionique lorsque la vitesse de balayage augmente, la sensibilité est accrue mais en contrepartie la résolution diminue. Le mode "Standard Enhanced" (8100 m/z / s) qui est utilisé en MS permet de faire un compromis entre résolution et sensibilité. Il est en effet important d'avoir la masse de parent la plus précise possible pour l'identification d'un peptide. En revanche, en MS/MS l'accent est mis sur la sensibilité c'est pourquoi l'on utilise le mode "Ultra Scan" qui permet une vitesse de balayage plus grande (26000 m/z / s) et donc une meilleure sensibilité.

Ainsi, un spectre MS/MS est généré plus rapidement qu'un spectre MS. Par exemple pour les méthodes développées pour la vérification du couplage, 3 parents sont sélectionnés par spectre MS. Un cycle est donc composé d'une MS et de 3 MS/MS et dure au total environ 1 s. Un tiers du temps est consacré à la MS et les deux tiers restants aux 3 MS/MS. Etant donné qu'un pic chromatographique mesure environ 10 secondes à mi-hauteur, il est possible d'effectuer 10 cycles par pic en moyenne.

### 3.2.5. Exclusion temporaire des ions parents après sélection

Il est possible d'exclure des ions parents de la sélection pour la MS/MS une fois qu'ils ont été fragmentés. Cette option permet d'éviter la sélection multiple d'un ion qui serait largement majoritaire dans une analyse et empêcherait la sélection d'ions minoritaires. Cette option doit être utilisée avec prudence car elle implique toujours l'exclusion d'une fenêtre de 2 à 3 Da par ion exclu. Or plusieurs peptides de séquence différente peuvent avoir des masses très proches et il faut éviter de perdre des informations sur ces peptides parce qu'ils seraient exclus de la sélection. Il est donc conseillé d'utiliser cette exclusion de façon temporaire, c'est-à-dire d'exclure les ions uniquement pendant la durée du pic chromatographique (Figure 12). En effet, même s'il arrive que des peptides aient des masses proches, il est plus rare qu'ils aient en plus le même temps de rétention chromatographique.

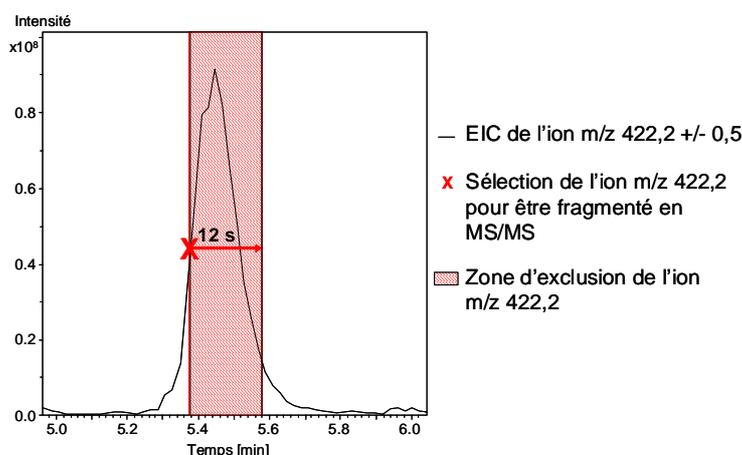


Figure 12. Exemple d'exclusion temporaire d'un ion

## 3.3. Discussion et conclusion

L'acquisition par le laboratoire d'un système nanoLC-Chip/MS (Agilent Technologies) a nécessité l'optimisation des paramètres de la trappe ionique HCT Ultra (Bruker Daltonics) pour s'adapter aux performances chromatographiques de ce nouvel instrument. Le mode d'acquisition en auto MS/MS a de plus du être adapté à l'évolution des méthodes d'interprétation pour l'identification des spectres de fragmentation.

En effet, au moment où ce travail a commencé, l'interprétation par *de novo* était encore assez courante car beaucoup d'organismes n'étaient pas séquencés. Par conséquent, les spectres de

fragmentation devaient avoir un rapport signal/bruit important et pour cela 6 spectres étaient moyennés.

Depuis, les banques de données se sont grandement développées et la majeure partie des interprétations est aujourd'hui effectuée à l'aide d'un moteur de recherche (PFF) comme Mascot ou/et OMSSA. Ces logiciels sont alors capables d'interpréter des spectres de fragmentation de qualité moindre. De plus la tendance actuelle est aux analyses d'échantillons de plus en plus complexes. Ainsi, pour séquencer un maximum de peptides les spectromètres de masse doivent être de plus en plus rapide pour effectuer le plus de MS/MS possible par analyse.

Le but de ce travail a été d'établir une méthode standard permettant de tester les performances du couplage nanoLC-Chip/HCT Ultra en utilisant un digestat trypsique de BSA comme référence. Cet échantillon modèle est peu complexe et parfaitement séquencé dans les banques protéiques. Nous avons donc choisi un compromis entre la qualité des spectres de fragmentation et le nombre de spectres par analyse avec : 3 spectres moyennés par MS/MS et 3 ions parents isolés par spectre MS qui sont exclus pendant 12 s après leur première sélection.

Cette méthode doit bien évidemment être adaptée en fonction de la complexité de l'échantillon, mais aussi en fonction du mode d'interprétation PFF ou *de novo*, et de la question posée. Pour l'analyse d'échantillons provenant d'organismes non séquencés ou pour caractériser des protéines mutées ou avec des modifications post traductionnelles, comme des glycosylations par exemple, la qualité des MSMS doit être optimale. Pour cela, le nombre de spectres MS/MS moyennés est augmenté (6 par exemple) et le paramétrage de l'exclusion temporaire modifié (exclusion après 2 sélections au lieu d'une seule). Sélectionner deux fois le même ion pour la MS/MS permet de s'assurer qu'il sera isolé à son intensité maximale.

En revanche dans le cas d'analyses d'échantillons plus complexes, l'objectif est de séquencer le plus de peptides possibles et donc d'obtenir un maximum de spectres MS/MS par analyses. Pour cela, le nombre de spectres moyenné est diminué au maximum soit 1 moyenne. Dans ce cas, un spectre correspond à une fragmentation.

Ainsi, j'ai optimisé une méthode standard avec des paramètres permettant d'effectuer un compromis entre qualité des spectres de fragmentation et nombre de spectres par analyse. Cette méthode convient parfaitement pour des mélanges peu complexes comme des spots de gel 2D, mais doit bien sûr être adaptée en fonction des échantillons et de la question posée.

## 4. Développement de tests permettant d'évaluer les performances du couplage nanoLC-Chip/HCT Ultra

L'évaluation des performances des instruments utilisés dans le cadre d'analyses protéomiques est un problème récurrent et mal résolu. Une analyse protéomique comporte plusieurs étapes qui nécessitent chacune un paramétrage précis [Aebersold 2009]. Chacune de ces étapes peut induire une erreur. Il est donc très délicat d'obtenir une bonne répétabilité et une bonne reproductibilité. De nombreux travaux ont été publiés à ce sujet [Elias et al. 2005, Van Midwoud et al. 2007, Bell et al. 2009, Tabb et al. 2010]. La plupart de ces études utilisent le nombre de peptides identifiés après l'analyse LC-MS/MS d'un mélange test de protéines comme critère d'évaluation. Cependant, cette seule donnée n'indique pas d'où provient le problème en cas de panne (injecteur, pompe, colonne, spectromètre de masse, etc.). Des travaux récents décrivent de nouvelles méthodes d'évaluation en mettant en avant les paramètres critiques lors d'analyses protéomiques par LC-MS/MS [Rudnick et al. 2010], ou en développant un logiciel permettant de tester de manière automatisée les performances d'un couplage LC-MS [Sigmon et al. 2010].

Jusqu'alors au laboratoire, notre seul moyen pour vérifier le bon fonctionnement du couplage nanoLC-Chip/HCT Ultra était la mesure du pourcentage de recouvrement après une analyse nanoLC-MS/MS d'un digestat trypsinique de BSA. Les optimisations précédemment décrites ont alors permis de développer un ensemble de tests complémentaires permettant de vérifier les performances de chacune des composantes du couplage tant au niveau chromatographique que du spectromètre de masse, mais permettant également de faciliter le diagnostic en cas de panne.

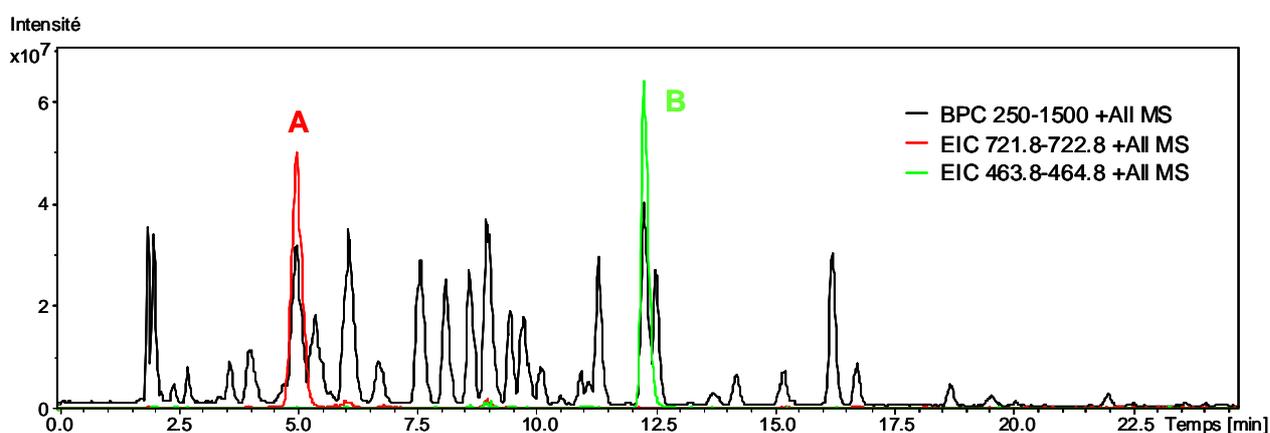
Le système nanoLC-Chip/MS ne pouvant pas être utilisé de manière autonome mais uniquement dans le cadre d'un couplage nanoLC-MS, ces performances ne peuvent être vérifiées indépendamment de celles du spectromètre de masse. La trappe ionique quant à elle, peut être testée en infusion, mais cette opération n'est réalisable qu'en électrospray. Le paramétrage de la source en ESI étant différent de celui utilisé en nanospray, l'infusion ne constitue pas un test suffisamment relevant pour vérifier les performances de la trappe ionique dans le cadre du couplage nanoLC-Chip/HCT Ultra.

Enfin, les tests de performances doivent pouvoir être effectués par tous les utilisateurs avant chaque série d'analyse (environ une fois par jour). Ils doivent donc être rapides et simples à réaliser tout en permettant une vérification complète du couplage. Il a donc été nécessaire de déterminer des grandeurs permettant d'obtenir des données significatives afin de diagnostiquer de manière précise une panne éventuelle.

Les grandeurs choisies sont les suivantes :

- L'intensité des pics chromatographiques,
- La largeur à mi-hauteur des pics chromatographiques,
- La différence entre les temps de rétention de deux ions de référence,
- La résolution spectrale,
- L'identification de la protéine par nanoLC-MS/MS (pourcentage de recouvrement).

Le digestat tryptique de BSA est conservé comme échantillon modèle. Deux ions correspondant à des peptides de la BSA (Figure 13) sont choisis pour servir de référence pour effectuer les mesures ( $m/z$  722,33 et  $m/z$  464,24). Des valeurs seuils sont déterminées expérimentalement pour chaque paramètre et servent de référence. Par exemple, pour une analyse sur une colonne de 150 mm, il doit y avoir au minimum 7 min entre les temps de rétention des deux ions de référence. Chacun des paramètres de mesure sont décrits dans la suite de ce chapitre.



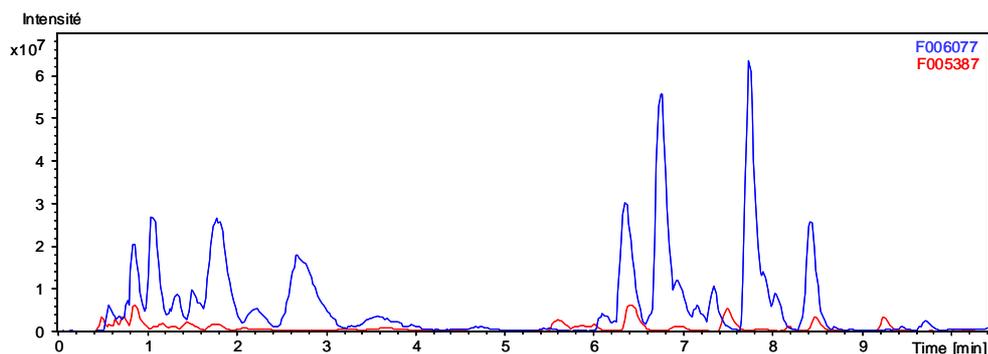
**Figure 13.** LC/MS de 50 fmol d'un digestat tryptique de BSA analysés avec un système nanoLC-Chip/HCT en utilisant une colonne de 150 mm de long. (A) et (B) chromatogrammes (EIC) des 2 ions ( $m/z$  722.3 et 464.2) servant de référence. Pour chacun de ces EIC, on mesure l'intensité du pic chromatographique et sa largeur à mi-hauteur. Ils permettent également de calculer la différence de temps de rétention entre les 2 ions

#### 4.1. Intensité des pics chromatographiques

Cette mesure donne une indication sur la sensibilité du système. En protéomique, les échantillons analysés sont généralement en faible quantité (de l'ordre du femtomole) et nécessitent des instruments très sensibles. Ce paramètre est donc particulièrement critique.

Une perte de sensibilité peut avoir plusieurs origines :

- L'encrassement du spectromètre de masse (Figure 14);
- Une phase stationnaire (colonne analytique) dégradée provoquant un élargissement des pics chromatographiques qui sont donc moins intenses;
- Une pré-colonne dégradée ne retient plus l'échantillon, les peptides sont alors élués avant d'arriver sur la colonne analytique d'où une perte d'intensité des pics chromatographiques.



**Figure 14.** NanoLC-MS de 200 fmol d'un digestat trypsique de BSA; avant (rouge) et après (bleu) nettoyage complet de la trappe ionique (transmission et analyseur)

## 4.2. Largeur des pics chromatographiques à mi-hauteur

Un élargissement des pics chromatographiques provoque une perte d'intensité et donc une perte de sensibilité. De plus, la mesure de la largeur des pics à mi-hauteur permet d'observer la résolution chromatographique. En effet, l'objectif d'une analyse nanoLC-MS/MS en protéomique est de séquencer un maximum de peptides afin d'identifier le plus de protéines possibles avec le meilleur recouvrement possible. Pour qu'un peptide soit séquencé, il doit être sélectionné pour être fragmenté en auto-MS/MS, et meilleure sera la séparation chromatographique plus grand sera le nombre de peptides sélectionnés pour être fragmentés.

Une augmentation de la largeur des pics chromatographiques peut être due à :

- Une colonne analytique non équilibrée ou avec une phase stationnaire dégradée;
- Un dysfonctionnement au niveau du gradient qui indiquerait un problème au niveau des pompes chromatographiques ou de la composition des solvants;
  - Un volume mort dans le système, qui peut venir d'un capillaire mal connecté par exemple. A 300 nL/min le moindre volume mort a des conséquences immédiates;
- Une fuite qui est indétectable à de tels débits.

## 4.3. Différence entre les temps de rétention des deux ions de référence

Le suivi de la différence entre les temps de rétention des ions de référence permet de déterminer la qualité de la séparation chromatographique. Une diminution de cette valeur, c'est-à-dire un "rapprochement" des pics chromatographiques mes en évidence le même type de panne qu'un élargissement des pics (§ 2.2.).

#### 4.4. Résolution spectrale

Une baisse de la résolution spectrale entraîne l'augmentation du taux de faux positifs lors de l'identification des protéines. En effet, si la résolution n'est pas suffisante, il devient impossible d'attribuer les états de charges des ions et donc d'identifier correctement les peptides correspondants. La résolution moyenne pour l'ion  $m/z$  722, 3 est 3000 et pour l'ion  $m/z$  464.2, 2000.

La diminution de la résolution spectrale peut être due à :

- Un encrassement de l'analyseur (trappe ionique);
- Un dysfonctionnement du détecteur (Figure 15).

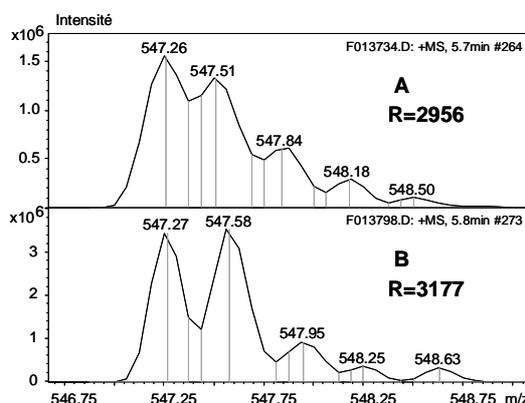


Figure 15. Spectres de l'ion  $m/z$  547,27 avec un détecteur défectueux (A) et avec un détecteur neuf (B)

#### 4.5. Identification de la protéine

L'ensemble des mesures décrites précédemment est effectué à partir d'une analyse nanoLC-MS. L'identification de la protéine nécessite une analyse nanoLC-MS/MS. Cette dernière mesure correspond au but de toute analyse protéomique, soit connaître les protéines présentes dans un échantillon biologique.

L'analyse nanoLC-MS/MS du digestat tryptique de BSA permet de générer une "peak list" (\*.mgf) qui est soumise au moteur de recherche Mascot pour effectuer une recherche dans la banque protéique SwissProt. Deux données sont alors relevées :

- L'erreur sur l'identification du parent en ppm, qui indique la justesse de l'étalonnage de l'instrument;
- Le pourcentage de recouvrement de la protéine, qui permet de donner une idée rapide du nombre de peptides sélectionnés en auto-MS/MS et ayant une qualité de spectre suffisante pour participer à l'identification de la protéine. Cette valeur peut légèrement varier en fonction du lot de BSA mais aussi en fonction de l'analyse car la sélection des parents en auto-MS/MS n'est pas parfaitement reproductible.

L'identification de la protéine donne donc une indication sur l'état du spectromètre de masse mais aussi sur l'état général de l'ensemble du système nanoLC-MS/MS.

#### **4.6. Discussion et conclusion**

En analyse protéomique, les échantillons fournis par les biologistes sont la plupart du temps en très petite quantité et/ou ne peuvent être fournis qu'une fois (purification complexe, patient décédé, etc.). Ce type d'échantillon ne permet qu'une seule injection et doit donc être analysé avec un instrument au maximum de ces performances. J'ai développée une série de tests simples et rapides permettant à chaque utilisateur de vérifier les performances du couplage nanoLC-Chip/HCT Ultra avant chaque série d'analyses.

Ces tests sont basés sur les analyses nanoLC-MS et MS/MS d'un digestat tryptique de sérum albumine bovine (BSA) qui sert de standard. Néanmoins, les échantillons biologiques ne contenant qu'une seule protéine sont très rares. L'utilisation comme standard d'un digestat issu du mélange de plusieurs protéines purifiées, à des concentrations différentes, permettrait de tester le couplage dans des conditions plus proches de celles utilisées expérimentalement. Pour cela il existe des mélanges commerciaux de protéines purifiées [Tabb et al. 2010]. L'identification d'une protéine présente en faible concentration dans ce mélange peut constituer un bon test de sensibilité. Il est aussi possible d'utiliser un mélange très complexe comme un digestat total de levure (*Saccharomyces cerevisiae*). Dans ce cas le test peut consister à identifier un nombre minimal de protéines. Ce test peut aussi être utilisé pour comparer les performances de différents couplages LC-MS [Elias et al. 2005].

Cependant des études ont montré que la seule identification d'une ou de plusieurs protéines en calculant leur pourcentage de recouvrement ne constitue pas un test suffisant pour vérifier les performances de l'ensemble du couplage [Rudnick et al. 2010, Sigmon et al. 2010]. Ainsi, j'ai établi une série de tests complémentaires permettant de vérifier plus finement l'ensemble des composantes du couplage nanoLC-Chip/HCT Ultra. Chacun des paramètres de mesure est significatif (sensibilité, résolution chromatographique, résolution spectrale) et permet de prévenir les pannes mais également d'en faciliter le diagnostic, ce qui n'est pas possible lorsque l'on recherche uniquement le nombre de peptides ou de protéines identifiés dans un mélange complexe.

Ces tests ont été adaptés à l'ensemble des couplages utilisés au laboratoire dans le cadre d'analyses protéomiques. Pour cela, les valeurs seuils ont été redéfinies en fonction des performances de chaque instrument.

## 5. Nouveauté instrumentale : la trappe amaZon

Au cours de ce travail, la nanoLC-chip a été couplée avec une nouvelle génération de trappe ionique (amaZon, Bruker Daltonics). Les principales améliorations apportées sur ce spectromètre de masse sont : la sensibilité, la vitesse de balayage, la résolution, la qualité des spectres de fragmentation.

Nous avons donc adapté les réglages du spectromètre de masse aux nouvelles capacités de cet instrument et redéfini les valeurs seuils pour les tests de performance du couplage nanoLC-Chip

### 5.1. Géométrie de l'instrument

La trappe amaZon (Bruker Daltonics) se différencie de la trappe HCT Ultra principalement par le remplacement du skimmer par deux séries de lentilles appelées funnels et par l'ajout d'une autre série de lentilles et d'un troisième octopole juste avant l'analyseur (Figure 16). Ces modifications permettent d'améliorer la transmission mais également de filtrer plus finement les ions arrivant jusqu'à la trappe ionique.

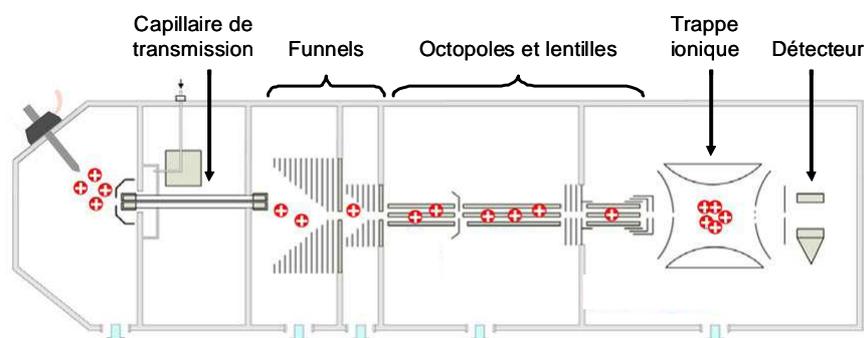


Figure 16. Schéma de la trappe ionique amaZon

### 5.2. Vitesse de balayage

L'une des améliorations apportées par la trappe amaZon est l'augmentation des vitesses de balayages (Figure 17). En mode "UltraScan" qui est utilisé pour la génération des spectres MS/MS la vitesse de balayage est passée de 26000 (m/z)/s avec l'HCT Ultra à 32500 (m/z)/s avec l'amaZon. L'augmentation de la vitesse de balayage permet d'améliorer la sensibilité de l'instrument.

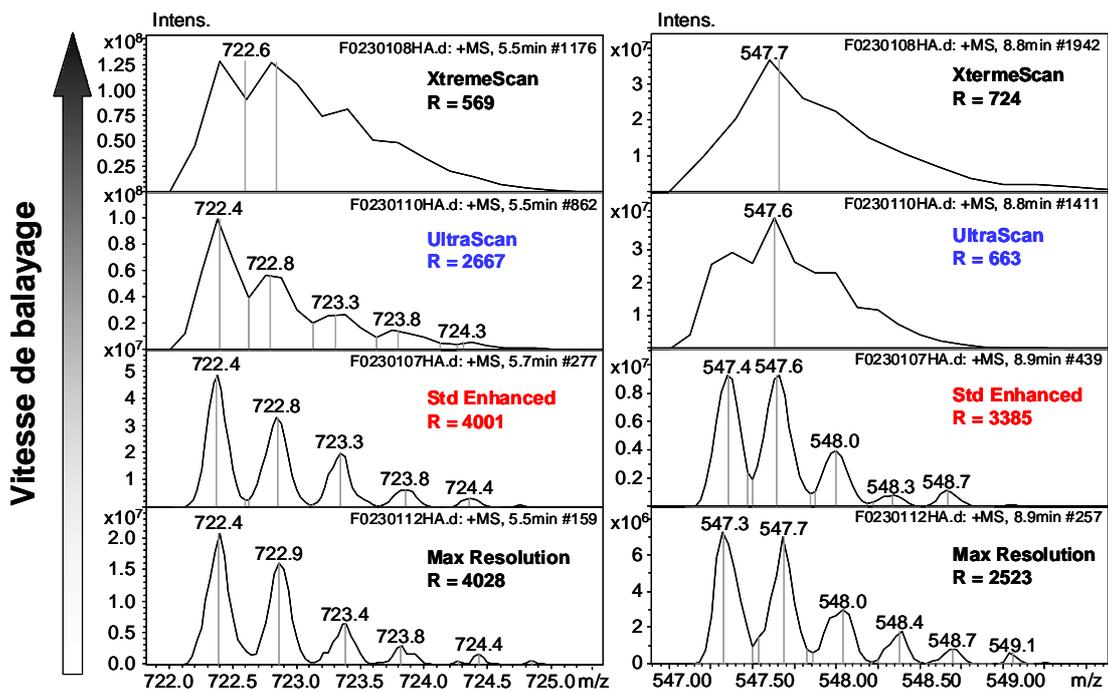
		MS/MS	MS	
(m/z)/s		UltraScan	Standard Enhanced	Standard Maximum
HCT Ultra		26000	8100	800
	XtremeScan	UltraScan	Standard Enhanced	Maximum Resolution
AmaZon	52000	32500	8100	4650

**Figure 17.** Modes et vitesses de balayage en (m/z)/s pour les trappes ioniques HCT Ultra et AmaZon

De plus, un nouveau mode a été créé, le mode "XtremeScan". Il permet un balayage encore plus rapide mais n'est pas utilisable en mode auto MS/MS, ce qui le rend peu intéressant dans le cadre d'analyses protéomiques où l'utilisation de l'auto MS/MS est indispensable. Enfin, le mode "Standard Maximum" est devenu le mode "Maximum Resolution" en multipliant par 6 sa vitesse de balayage. La sensibilité de ce mode est donc grandement améliorée mais n'est toujours pas suffisante pour des analyses protéomiques en nanoLC-MS/MS.

### 5.3. Résolution spectrale

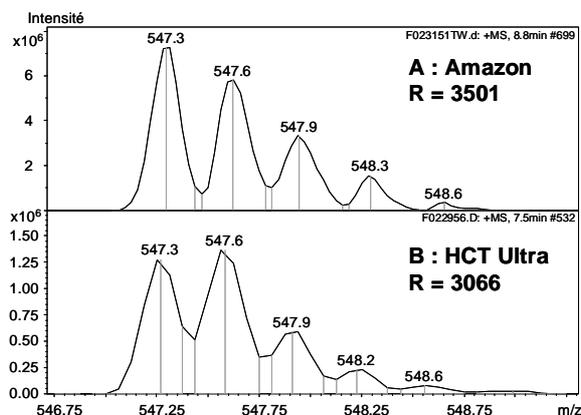
Avec les instruments de type trappe ionique, la vitesse de balayage est directement liée à la résolution spectrale. La Figure 18 montre la résolution obtenue pour deux ions de la BSA, un dichargé (m/z 722,4) et un trichargé (m/z 547,3) en utilisant les différents modes de balayage. Ces deux ions ont été choisis car ils sont représentatifs du type d'ions analysés généralement en protéomique. On observe que la résolution diminue quand la vitesse de balayage augmente. Le mode "XtremeScan" n'est pas assez résolutif pour distinguer les états de charges pour les deux ions choisis. En mode "UltraScan" il est possible de distinguer un ion doublement chargé mais pas un triplement chargé pour cette gamme de masse. Enfin, le mode "Maximum Resolution" ne génère pas un gain de résolution significatif pour les ions sélectionnés. Ainsi, pour le type d'ions générés dans le cadre d'analyses protéomiques les modes "XtremeScan" et "Maximum Resolution" ne sont pas adaptés.



**Figure 18.** Résolution (R) en fonction du mode de balayage pour un ion dichargé (m/z 722,4) et un ion trichargé (m/z 547,3)

Avec cette nouvelle génération la vitesse de balayage en mode "UltraScan" a été augmentée (32500 au lieu de 26000 (m/z)/s). Pourtant, on a pu observer une amélioration de la résolution. Ce mode étant utilisé pour la génération des spectres MS/MS, il est maintenant possible de déterminer l'état de charge de fragments dichargés ce qui n'était pas le cas avec la génération précédente. L'interprétation des spectres MS/MS est donc plus facile et surtout plus précise.

De plus comme le montre la Figure 19, pour un même ion, la trappe AmaZon permet d'obtenir une meilleure résolution que l'HCT Ultra en mode "Standard Enhanced" qui est le mode utilisé en MS/MS, bien que la vitesse de balayage n'ait pas changé entre les deux générations.



**Figure 19.** Résolutions obtenues pour un ion trois fois chargé correspondant à un peptide de BSA en mode "Standard Enhanced" pour les trappes AmaZon (A) et HCT Ultra (B)

## 5.4. Sensibilité

L'un des autres apports majeurs de cette nouvelle génération de trappe ionique est le gain de sensibilité de l'ordre d'un facteur 10. Ceci est essentiellement dû à la nouvelle géométrie de l'interface, avec la mise en place de funnels, et à une meilleure transmission des ions au travers des octopoles. L'augmentation de la vitesse de balayage en mode "UltraScan" permet également de gagner en sensibilité en MS/MS. Ainsi, avec la trappe amaZon, le test de performance est effectué en analysant 5 fmol d'un digestat tryptique de BSA et non plus 50 fmol comme avec les trappes HCT Plus et Ultra.

## 5.5. Conclusion

La nouvelle génération de trappe ionique amaZon a permis un net gain en sensibilité, en résolution et en vitesse de balayage par rapport à la génération précédente. L'ensemble de ces améliorations permet d'obtenir des spectres MS/MS de meilleure qualité et en plus grand nombre. Ainsi, cette instrument donne la possibilité d'analyser des échantillons plus complexes et à des concentrations plus faibles.

L'ensemble des méthodes utilisées pour évaluer les performances établies pour le couplage nanoLC-Chip avec la trappe HCT Ultra ont été adaptés pour ce nouvel instrument.

## 6. Discussion et conclusion

L'objectif de ce travail était de développer une méthode permettant d'évaluer les performances d'un couplage nanoLC-Chip/HCT Ultra et d'en tirer partie pour mieux maîtriser les différents paramètres chromatographiques pour l'analyse des échantillons précieux, sur lesquels une seule analyse pouvait être faite avec le meilleur compromis entre la sensibilité et la qualité des données MS/MS.

Notre travail a porté sur l'utilisation de l'acétonitrile ainsi que du méthanol comme modificateur de la phase mobile. Nous avons montré que l'utilisation du méthanol permettait d'obtenir une meilleure sensibilité qu'avec l'acétonitrile. De plus, nous avons montré, que dans le cas du système nanoLC-Chip/MS que nous avons utilisé, le remplacement de l'acétonitrile par le méthanol permettait d'augmenter le taux de recouvrement sur un digeste trypsique de protéines modèles, comme la BSA. Ce dernier résultat est important parce que l'augmentation du taux de recouvrement est l'objectif final de toute optimisation pour une analyse protéomique.

Nous avons déterminé les paramètres chromatographiques, (le débit, gradient et remplacement de l'acétonitrile par le méthanol) pour obtenir la meilleure résolution chromatographique et la finesse des pics chromatographique adéquate pour des cycles d'acquisition MS/MS en mode automatique. Le paramétrage de la MS/MS automatique a également été optimisé pour s'adapter aux performances du système nanoLC-Chip/MS. Il ressort de notre étude que les choix des paramètres chromatographiques et d'acquisition des spectres MS/MS sont très interdépendants. C'est pourquoi, nous avons considéré que le critère d'optimisation ultime pour un système nanoLC-MS/MS était le taux de recouvrement sur un digeste trypsique de protéines modèles.

Nous avons proposé un ensemble de tests pour vérifier les performances du couplage nanoLC-Chip/HCT Ultra, qui comprend bien sûr le taux de recouvrement sur un digeste trypsique de protéines modèles, mais également d'autres critères complémentaires d'évaluation du système. J'ai développé un ensemble de tests basés sur le suivi de 2 peptides lors d'analyses nanoLC-MS et nanoLC-MS/MS d'un digestat trypsique de BSA. Ces deux peptides (ions de  $m/z$  722,33 et  $m/z$  464,24) ont été choisis pour mesurer différentes grandeurs qui permettront de vérifier l'intégrité du système. Ces grandeurs sont : l'intensité et la largeur à mi-hauteur des pics chromatographiques, la différence entre les temps de rétention des deux ions de référence, la résolution spectrale et le pourcentage de recouvrement de la BSA. Des valeurs seuils ont été déterminées expérimentalement pour chaque grandeur et servent de critère pour déterminer l'état du système. Les tests utilisés en complément de la seule mesure du taux de recouvrement nous permettent d'avoir une vision très fine de l'état du système et de détecter des dérives (légère fuite, problème de pompage, légère perte à l'injection, dégradation de la résolution chromatographique). Ce suivi du système complet permet donc de mettre toutes les chances de notre côté pour prévenir une panne qui pourrait survenir pendant une série d'analyse d'échantillons précieux.

Cet ensemble de tests peut être utilisé pour tous les systèmes nanoLC-MS/MS en ajustant- les valeurs seuils expérimentales en fonction des performances chromatographiques et de celles du spectromètre de masse.

En 2009, HUPO (Human Proteome Organisation) a publié une étude impliquant 27 laboratoires [Bell et al. 2009]. Chacune des équipes a du analyser une mélange de 20 protéines recombinantes en quantités équimolaires. Les résultats ont ensuite été comparés. Bien que cela semble être une évidence, cette étude montre que l'optimisation des méthodes analytiques et la parfaite maîtrise de l'instrument et de l'interprétation des données sont indispensables pour obtenir des résultats fiables. Plusieurs types de spectromètres de masses (géométrie, génération, constructeur, etc) et des moteurs de recherches différents ont été utilisés. Pourtant ce ne sont pas ces facteurs qui ont influencés les résultats mais l'expertise des manipulateurs. Cette étude montre bien à quel point l'ensemble des étapes qui composent une analyse protéomique doit être optimisé et maîtrisé [Aebersold 2009].

## Chapitre 2

### Développements méthodologiques pour la détection et la caractérisation des protéines N-glycosylées

---

Au cours de mon travail de thèse, j'ai traité plusieurs projets de recherche qui portaient sur la caractérisation de protéines N-glycosylées présentes dans des mélanges complexes et avec des quantités d'échantillons relativement faibles. Pour obtenir les données de protéomique recherchées, nous avons développé des méthodologies spécifiques à l'étude des glycoprotéines rencontrées. Nous avons choisi d'adapter des méthodes de protéomique classiques (digestion enzymatique, LC-MS et LC-MS/MS) pour développer une méthodologie en trois étapes :

- La **digestion** par une enzyme spécifique de la glycoprotéine d'intérêt qui génère un mélange de peptides et de glycopeptides;
- L'analyse **LC-MS** du digestat de la glycoprotéine dans des conditions permettant de former les **ions oxoniums** spécifiques de la fragmentation des glycosylations;
- L'analyse **LC-MS/MS** du digestat afin d'obtenir des spectres de fragmentation des glycopeptides et ainsi de les caractériser.

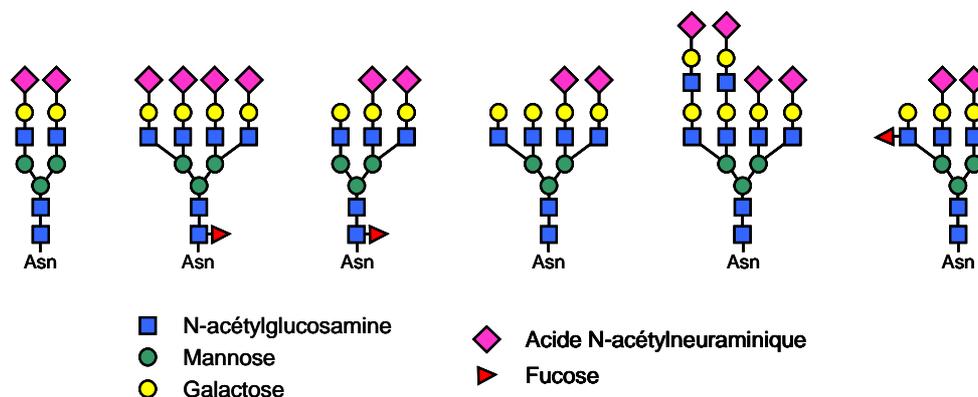
Cette méthode a été optimisée et testée sur plusieurs spectromètres de masse disponibles au laboratoire afin de comparer leurs performances et de choisir le plus adapté pour nos analyses. L'objectif était bien sûr de parvenir à la sensibilité la plus élevée possible pour la détection de tous les glycopeptides présents.

## 1. Introduction

### 1.1. Une glycoprotéine modèle : l'alpha-1 glycoprotéine acide humaine (AGP)

Pour optimiser notre stratégie, nous avons utilisé une protéine modèle appelée alpha-1 glycoprotéine acide humaine que nous appellerons AGP dans la suite de ce manuscrit dans un but de simplification. Cette protéine plasmatique possède deux rôles principaux qui sont le transport de substance endogènes ou exogènes et une fonction de modulation immunologique. Sa concentration dans le plasma augmente fortement en cas d'inflammation ou de diverses maladies chroniques [Fournier et al. 2000]. Cette glycoprotéine est purifiée à partir du sérum et est disponible commercialement.

L'AGP est une glycoprotéine particulièrement hétérogène de 37 kDa dont 45% de la masse correspond à des glycosylations. En effet, elle possède 5 sites (Asn-15, 38, 54, 75 et 85) qui portent des N-glycosylations de type complexe. Les chaînes glycaniques peuvent être bi-, tri- ou tétra-antennées et porter des acides sialiques, des fucoses et des motifs lactosamine (Figure 1) [Treueheit et al. 1992, Fournier et al. 2000, Imre et al. 2005].



**Figure 1.** Exemples de structure de glycosylations de l'AGP

L'AGP possède également un polymorphisme génétique. Elle existe sous de nombreuses isoformes [Budai et al. 2009] dont les quantités relatives varient en fonction des pathologies. L'AGP commerciale est purifiée à partir de sérum sain et correspond au mélange de deux isoformes principales (AGP1 et AGP2). Les séquences et les masses des peptides tryptiques portant des sites consensus de glycosylation et correspondants à ces isoformes sont décrites dans le Tableau 1.

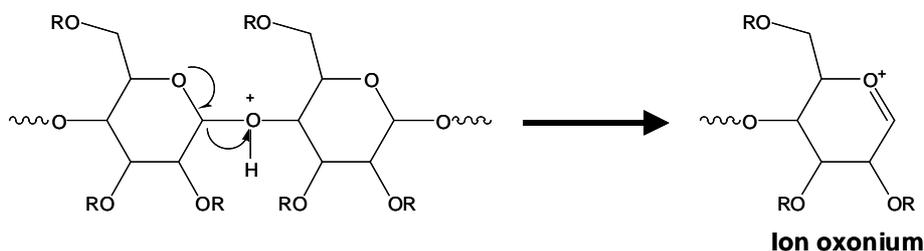
numéro du peptide	position du site	Séquences (Isoformes)	Masse (Da)	CAM (+57)
1	Asn <sup>15</sup>	LVPVPIT <b>N</b> ATLDR	1407,81	
2		CANLVPVPIT <b>N</b> ATLDR	1695,90	1752,90
3		LVPVPIT <b>N</b> ATLDQITGK	1779,01	
4		CANLVPVPIT <b>N</b> ATLDQITGK	2067,10	2124,10
5	Asn <sup>38</sup>	NEEY <b>N</b> K	795,34	
6	Asn <sup>54</sup>	SVQEIQATFFYFTP <b>N</b> K	1918,95	
7		SVQEIQATFFYFTP <b>N</b> KTEDTIFLR	2894,44	
8	Asn <sup>75</sup>	QDQCIY <b>N</b> TTYLNVQR	1857,87	1914,87
9		QNQCFY <b>N</b> SSYLVNQR	1862,84	1919,84
10	Asn <sup>85</sup>	<b>E</b> NGTISR	775,38	
11		<b>E</b> NGTVSR	761,37	

**Tableau 1.** Séquences et masses monoisotopiques des peptides tryptiques correspondants aux isoformes les plus courantes de l'AGP et portant un des cinq sites de glycosylation ; les peptides sont numérotés de 1 à 11 ; la dernière colonne correspond aux masses des peptides réduits/alkylés qui portent une cystéine

L'hétérogénéité des glycosylations, mais également la grande diversité de séquence font de l'AGP une glycoprotéine de choix en tant que modèle pour le développement de méthodes d'analyses de protéines N-glycosylées [Yu et al. 2007, Zhang et al. 2008, Ivancic et al. 2010].

## 1.2. Les ions diagnostiques

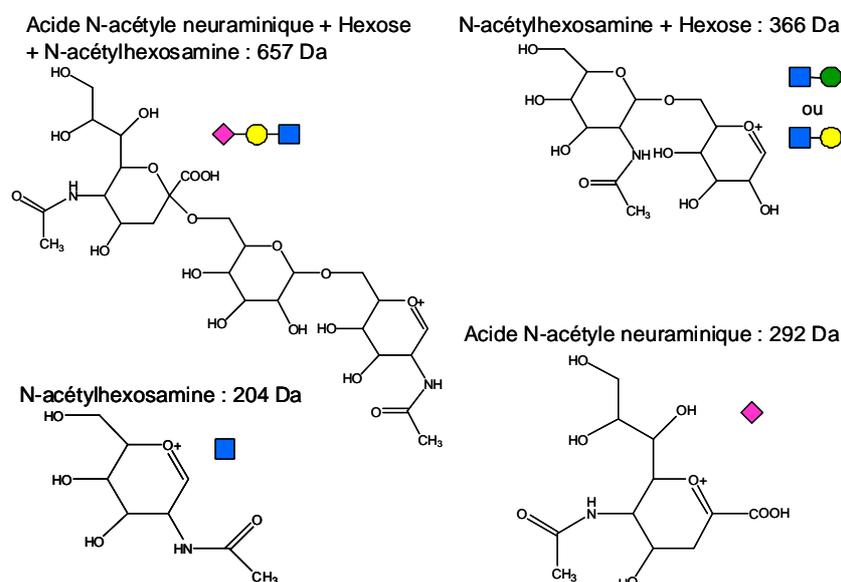
Au début des années 90, Conboy et Henion [Conboy et Henion 1992], ainsi que l'équipe de Carr [Carr et al. 1993, Huddleston et al. 1993], ont montré que la partie glycanique des glycopeptides se fragmente en générant des ions oxoniums spécifiques (Figure 2) qui peuvent servir de marqueurs. Nous appellerons ces ions "ions diagnostiques" dans la suite de ce manuscrit.



**Figure 2.** Mécanisme de fragmentation pour la génération d'ions oxoniums qui servent d'ions diagnostiques

Bien que de nombreux travaux décrivent l'utilité des ions diagnostiques, aucune méthodologie permettant de les générer à partir de mélanges complexes et en très petites quantités (extrait de bande de gel 1D par exemple) n'a encore été publiée.

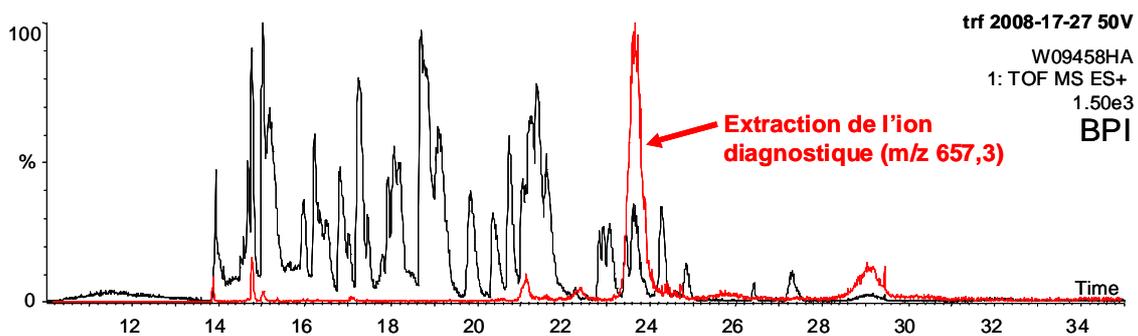
Les principaux ions diagnostiques générés par des glycosylations de type complexes et sialylées sont décrits ci-dessous dans la Figure 3.



**Figure 3.** Structures et symboles des principaux ions oxoniums générés par la fragmentation des glycosylations de type complexes et sialylées utilisés comme ions diagnostiques.

Ces ions permettent de déterminer les temps de rétention des glycopeptides lors d'une analyse LC-MS ou LC-MS/MS (cf. Biblio, Chap 1).

L'exemple ci-dessous illustre l'intérêt d'utiliser le courant d'ions correspondant aux ions diagnostiques des glycopeptides pour détecter les zones d'élution des glycopeptides dans un mélange des peptides de digestion d'une glycoprotéine. La Figure 4 présente l'analyse nanoLC/MS d'un digestat tryptique de Transferrine. Cette glycoprotéine est moins hétérogène et donc moins complexe que l'AGP qui est utilisée comme modèle pour le développement de notre méthodologie. La transferrine a une masse d'environ 77 kDa et possède 2 sites de glycosylations. Les deux glycopeptides sont donc "noyés" au milieu de la cinquantaine de peptides tryptiques générés par la digestion enzymatique de la glycoprotéine. La recherche d'ions diagnostiques a permis de déterminer facilement la zone d'élution de l'un d'eux à 23,7 min. Dans cette analyse, réalisée sans optimisation particulière, le deuxième glycopeptide n'a pu être détecté. Ceci est dû à sa masse trop élevée (environ 5000 Da) pour les conditions utilisées comme le montreront les études décrites dans la suite de ce chapitre.



**Figure 4.** NanoLC-MS d'un digestat de transferrine sur un Q-TOF (Synapt G1, Waters); extraction de l'ion m/z 657,3 qui correspond à NeuAcHexHexNAc

## 2. Comparaison de différentes géométries d'instruments pour la génération d'ions diagnostiques

Les ions diagnostiques peuvent être générés de deux manières : soit par CID (Collision Induced Dissociation) en source, soit par CID dans la cellule de collision du spectromètre de masse. Une partie de mon travail a consisté à comparer et optimiser ces deux modes de fragmentation sur plusieurs spectromètres de masse disponibles au laboratoire afin de déterminer la méthode la plus efficace pour générer des ions diagnostiques.

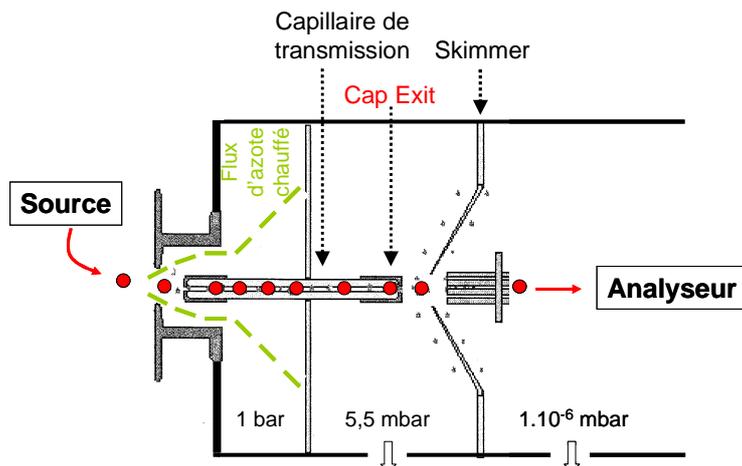
### 2.1. Génération d'ions diagnostiques par CID en source

Afin de déterminer quel est le spectromètre de masse le plus adapté pour effectuer de la CID en source, la capacité de chaque interface à générer des ions diagnostiques a été testée. J'ai ainsi comparé et optimisé des méthodes de fragmentation en source pour trois types d'interfaces différentes :

- Interface comprenant un capillaire de transmission et un skimmer (HCT Plus et Ultra, Bruker Daltonics (trappes ioniques));
- Interface comprenant un capillaire de transmission et des funnels (MicrOTOF-Q et maXis, Bruker Daltonics (Q-TOF));
- Interface de type Z-spray (Synapt G1, Waters (Q-TOF)).

#### 2.1.1. Interface comprenant un capillaire de transmission et un skimmer

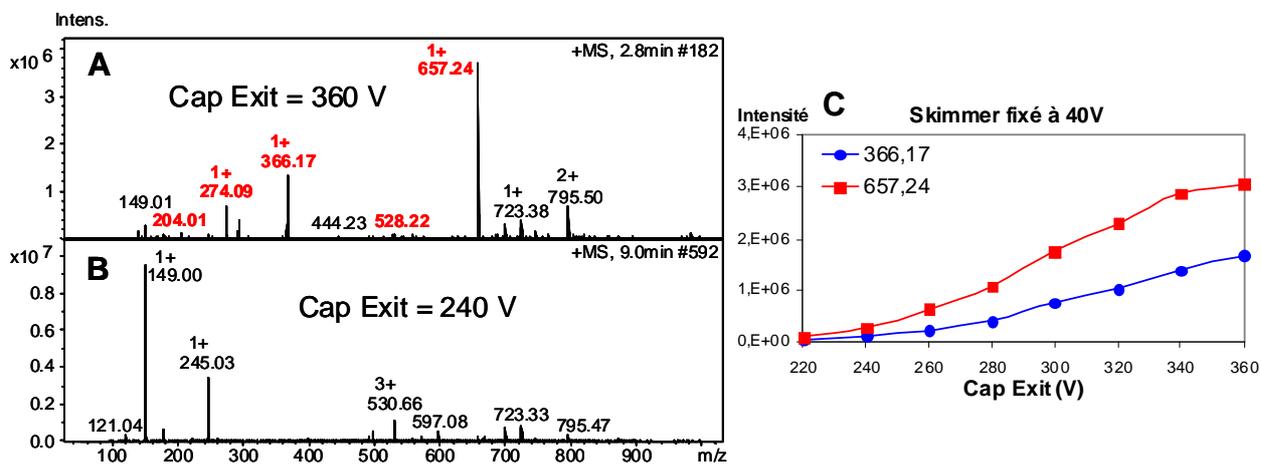
Ce type d'interface est composé d'un capillaire de transmission en quartz métallisé à ses extrémités suivi d'un skimmer (Figure 5). L'entrée du capillaire se trouve à pression atmosphérique et la sortie à environ 5,5 mbar. En nanospray positif (nanoLC-Chip/MS) la tension appliquée sur l'entrée du capillaire est de l'ordre de -1750 V. Dans les conditions standards utilisées en analyse protéomique, la tension appliquée à l'extrémité du capillaire de transmission (Cap Exit) est de +140 V et de +40 V sur le skimmer. Malgré une différence de potentiel défavorable entre l'entrée et la sortie du capillaire, les ions sont tout de même transmis vers le skimmer et ce, grâce à la différence de pression gazeuse.



**Figure 5.** Schéma de l'interface composée d'un capillaire de transmission et d'un skimmer

En appliquant une tension plus importante au niveau de l'extrémité du capillaire de transmission (Cap Exit), il est possible de fragmenter les ions et ainsi d'effectuer de la CID en source. Dans le cas des glycopeptides, cela permet de générer des ions diagnostiques.

J'ai développé une méthode pour générer et détecter ces ions diagnostiques sur une trappe ionique HCT Plus (Bruker Daltonics). Les paramètres critiques à optimiser sont : la tension appliquée au niveau du Cap Exit, la gamme de balayage et les paramètres régissant la transmission des ions. Une première expérience a été effectuée en infusant une fraction purifiée de glycopeptides tryptiques d'AGP. L'échantillon infusé est issu d'une collecte LC-MS (C18) basée sur les ions diagnostiques.



**Figure 6.** Spectres obtenus avec une tension de 360 V (A) et 240 V (B) appliquée au niveau du Cap Exit; Intensité des ions diagnostiques majoritaires (m/z 366,17 et 657,24) en fonction de la tension appliquée au niveau du Cap Exit, avec la tension du skimmer fixée à 40 V (C)

Afin d'obtenir une meilleure sensibilité, la gamme de balayage est restreinte à m/z 50 à 1000. Les paramètres de transmission sont réglés pour une masse de m/z 500, qui correspond à une moyenne entre les deux principaux ions diagnostiques de l'AGP (m/z 366 et 657). Différentes tensions

de Cap Exit sont testées, de 220 à 360 V (Figure 6). L'intensité maximale pour les deux ions diagnostiques majoritaires ( $m/z$  366,17 et  $m/z$  657,24) est obtenue à 360 V qui est la tension maximale applicable sur le Cap Exit.

Cependant j'ai observé qu'une augmentation trop importante de la tension au niveau du Cap Exit entraîne une défocalisation des ions, ce qui provoque un encrassement très rapide de l'interface et se traduit par une chute importante du signal. J'ai effectué des tests pour optimiser la transmission et ainsi éviter ce problème sans résultats probants. La seule solution reste le nettoyage du capillaire de transmission et du skimmer après chaque série d'analyse, ce qui est fastidieux et risque d'endommager le spectromètre de masse. En effet, les démontages et remontages répétés de l'interface finissent par entraîner des dommages mécaniques (connexions, micro-soudure, capillaires, contacts électriques,...).

Ce type d'interface n'est donc pas utilisable en routine à haut débit pour générer des ions diagnostiques par fragmentation en source.

### 2.1.2. Interface comprenant un capillaire de transmission et des funnels

Cette interface possède une géométrie relativement proche de celle décrite dans le paragraphe précédent. En effet, elle comprend un capillaire de transmission similaire à celui décrit ci-dessus. La principale différence se situe après le capillaire. Dans cette interface, le skimmer est remplacé par deux séries de lentilles qui sont appelées funnels (Figure 7) sur lesquelles sont appliqués des potentiels décroissants et qui permettent une meilleure focalisation des ions à la sortie du capillaire de transmission. De plus, les orifices des lentilles sont légèrement décentrés par rapport à la sortie du capillaire de transmission. Ainsi, seuls les ions qui sont focalisés par les champs électriques générés par les lentilles sont transmis. Tous les ions négatifs ou les molécules neutres vont s'écraser sur le funnel 1. Ce système a pour avantage de diminuer le rapport signal/bruit et d'améliorer la sensibilité de l'instrument.

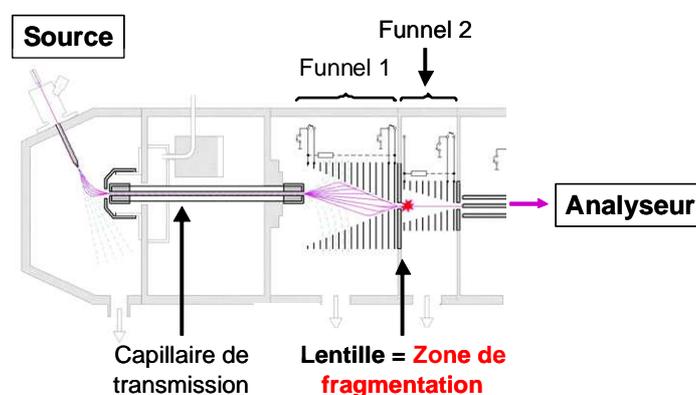
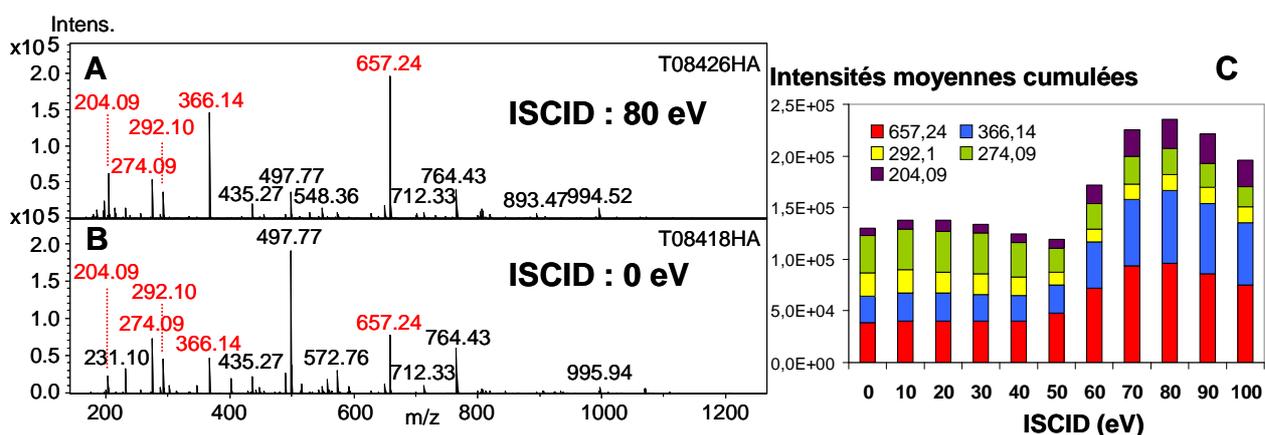


Figure 7. Schéma de l'interface composée d'un capillaire de transmission et de funnels

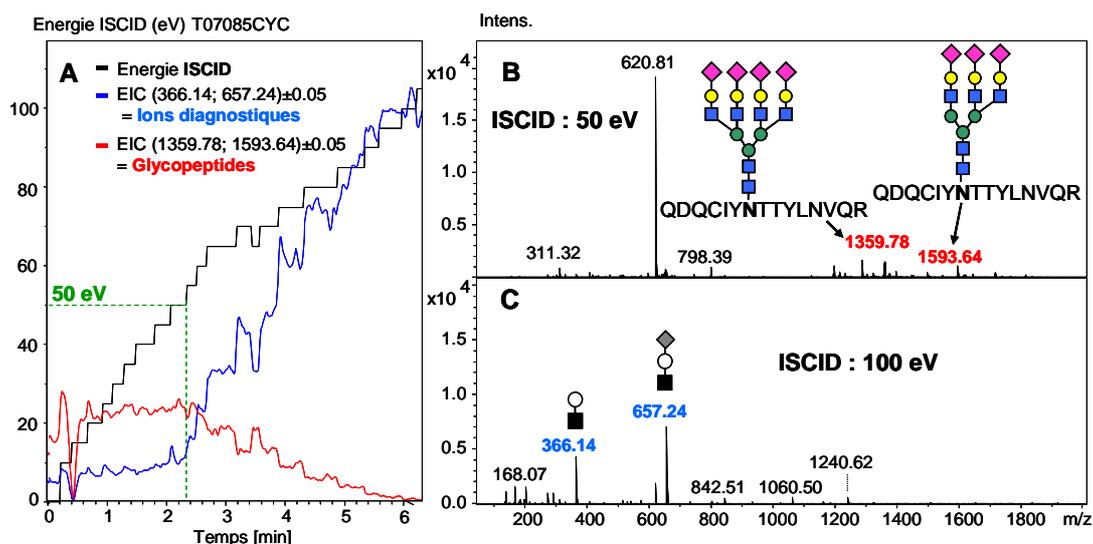
En augmentant l'énergie appliquée sur la lentille située entre les deux funnels, il est possible de fragmenter les ions en source. Sur les instruments de type Q-TOF de Bruker Daltonics (maXis et MicrOTOF-Q) cette option est appelée ISCID (In Source Collision Induced Dissociation). Le funnel 2 permet ensuite de refocaliser les ions et donc de limiter une perte de sensibilité due à la fragmentation.

Afin de déterminer l'énergie optimale à appliquer au niveau de l'ISCID pour la génération d'ions diagnostiques, un digestat trypsique d'AGP est analysé en nanoLC/MS avec un MicrOTOF-Q (Bruker Daltonics). La gamme de scan utilisée est  $m/z$  100-1300 en favorisant la transmission des ions diagnostiques. Plusieurs énergies sont testées (0 à 100 eV). Les intensités indiquées dans le graphique de la Figure 8C correspondent à des valeurs moyennes relevées sur l'ensemble de l'analyse. Si l'on additionne les intensités de chaque ion diagnostique, l'intensité maximale est obtenue entre 70 et 90 eV (Figure 8A) appliquée au niveau de l'ISCID. Cependant, les ions diagnostiques sont générés même lorsque qu'aucune énergie n'est appliquée (Figure 8B).



**Figure 8.** Spectres obtenus avec une énergie de 80 eV (A) et 0 eV (B) appliquée en ISCID; Intensités moyennes cumulées des ions diagnostiques en fonction de l'énergie appliquée (C)

Le MicrOTOF-Q est un analyseur de type Q-TOF et sa gamme de masse permet de voir sur un même spectre les ions des glycopeptides (entre  $m/z$  1000 et 2000) et les ions diagnostiques (entre  $m/z$  204 et 657) issus de leur fragmentation. Afin d'observer le comportement des glycopeptides et la génération des ions diagnostiques en fonction de l'énergie ISCID, j'ai infusé une fraction de glycopeptides de l'AGP en mode électrospray sur le micrOTOF-Q avec une gamme de balayage allant de  $m/z$  100 à  $m/z$  3000. L'énergie ISCID est augmentée par palier de 5 eV pour gamme allant de 0 à 100 eV.



**Figure 9.** Infusion d'une fraction de glycopeptides de l'AGP en électrospray sur un MicroTOF-Q; courants d'ions extraits (EIC) pour les ions diagnostiques et les glycopeptides en fonction de l'énergie ISCID (A); spectres pour 50 eV (B) et 100 eV (C)

Ainsi, comme le montre la Figure 9, l'intensité des ions diagnostiques est constante de 0 à 50 eV. Pour ces énergies, l'intensité des glycopeptides reste également constante. Ensuite, entre 50 et 100 eV, l'intensité des ions diagnostiques augmente tandis que celle des glycopeptides diminue. Ceci indique qu'en dessous de 50 eV l'énergie appliquée n'est pas suffisante pour fragmenter les glycopeptides en source, mais qu'une fragmentation parasite se produit à un autre endroit du spectromètre de masse, entre l'interface et le détecteur. Néanmoins, comme le montre la Figure 9B, à basse énergie, les ions diagnostiques restent faibles et les glycopeptides sont suffisamment intenses pour être identifiés. Dans la Figure 9C, on observe la disparition de ces glycopeptides due à leur fragmentation qui se traduit par la génération d'ions oxoniums de  $m/z$  366,14 et 657,24.

Grâce au réglage facile de la tension dans la zone de fragmentation, ce type d'interface permet de générer et transmettre efficacement les ions diagnostiques issus de la fragmentation en source de glycopeptides. Associé à un analyseur de type TOF, la résolution et la précision de masse de ce type d'instrument permettent une détection très spécifique des ions diagnostiques.

### 2.1.3. Interface de type Z-spray

Avec l'interface de type Z-spray, les ions subissent deux changements directionnels (Figure 10) induits par des différences de potentiel au niveau du cône d'échantillonnage et du cône d'extraction, d'où son nom de "Z"-spray. Ainsi, les espèces neutres ou chargées négativement entrent en collision avec les parois de l'interface. Ce système permet de filtrer finement les ions et ainsi d'obtenir un bruit de fond de faible intensité ainsi qu'un encrassement plus limité du spectromètre de masse.

L'énergie appliquée au niveau du cône d'échantillonnage est déterminante pour une bonne transmission des ions. Dans des conditions standards pour des analyses protéomiques, elle est fixée à 35 V. En augmentant cette tension il est possible de fragmenter les ions et ainsi d'obtenir des ions diagnostiques lors de l'analyse de glycopeptides.

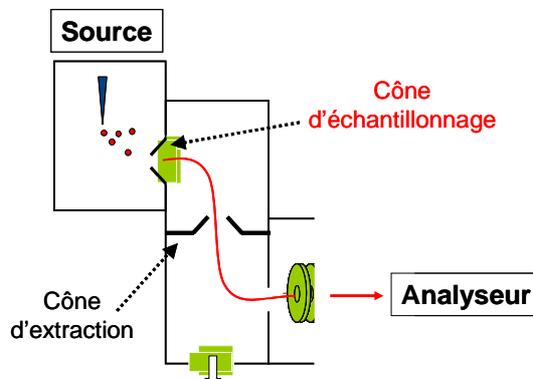


Figure 10. Schéma de l'interface Z-spray

Afin d'optimiser l'intensité des ions diagnostiques, j'ai analysé un digestat trypsique d'AGP en nanoLC-MS avec un Synapt G1 (Waters) pour différentes tensions de cône (5 à 90 V). La gamme de balayage est de  $m/z$  50-1000 avec des paramètres de transmission adaptés à la détection des ions diagnostiques. Les résultats obtenus sont illustrés dans la Figure 11.

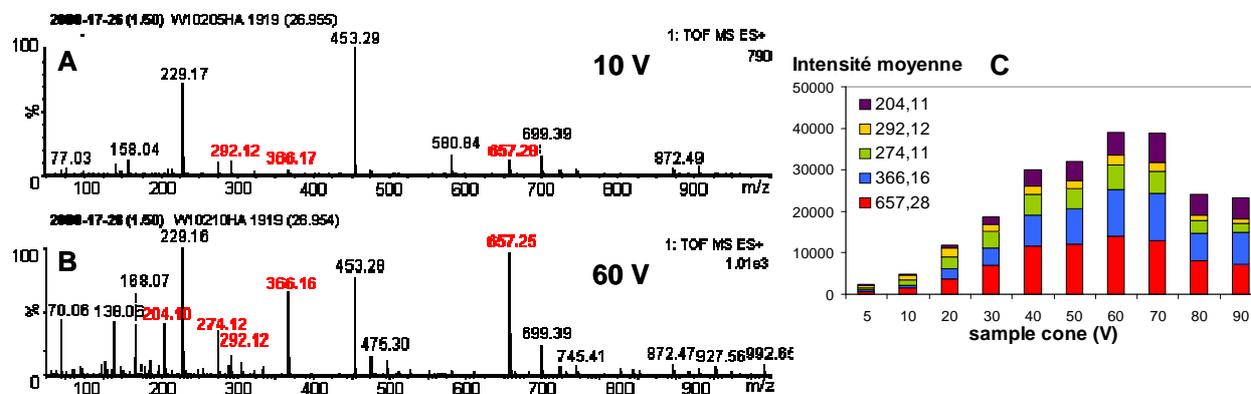
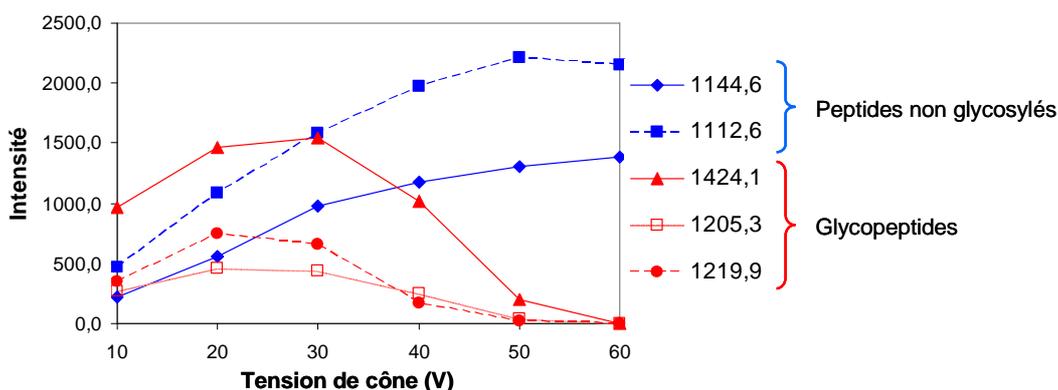


Figure 11. Spectres obtenus avec une tension de 10 V (A) et 60 V (B) appliquée au niveau du cône d'échantillonnage; Intensités moyennes cumulées des ions diagnostiques en fonction de la tension de cône appliquée (C)

Les intensités des ions diagnostiques indiquées dans le graphique de la Figure 11C correspondent à des valeurs moyennes relevées sur l'ensemble de l'analyse. L'intensité des ions diagnostiques augmente de façon constante pour des tensions allant de 5 à 70 V. Ainsi le phénomène de fragmentation en source se manifeste même pour des tensions de cône très basses (Figure 11A). Bien que leur intensité maximale soit obtenue aux alentours de 60 V (Figure 11B), les ions diagnostiques sont parfaitement détectables dès 20 V.

Ce type d'interface est très énergétique et permet donc d'obtenir facilement des ions diagnostiques par fragmentation en source. Si l'on cherche à identifier des glycopeptides entiers, il est nécessaire de faire un compromis entre une bonne transmission des ions et un minimum de fragmentation en source. En effet, comme le montre la Figure 12, l'augmentation de la tension de cône favorise la transmission des ions comme c'est le cas pour les peptides non glycosylés. En revanche, les glycopeptides réagissent différemment. Au delà de 30 V leur intensité chute car ils se fragmentent.



**Figure 12.** Intensités de peptides tryptiques non glycosylés et de glycopeptides de l'AGP en fonction de la tension appliquée au niveau du cône d'échantillonnage sur un Synapt G1 (Waters)

L'interface de type Z-spray est particulièrement bien adaptée pour la génération d'ions diagnostiques. Pour analyser des glycopeptides intacts, la tension au niveau du cône d'échantillonnage doit être inférieure à 30 V.

#### 2.1.4. Discussion et conclusion

J'ai développées des méthodes permettant de générer des ions diagnostiques en fragmentant des glycopeptides par CID en source pour trois types d'interfaces :

- L'interface comprenant un capillaire de transmission et un skimmer est montée sur les trappes ioniques HCT Plus et Ultra de Bruker Daltonics. J'ai réussi à générer et à détecter des ions diagnostiques obtenus par CID en source avec ces instruments, mais la fragmentation des ions dans l'interface provoque un encrassement rapide et une perte de signal qui rendent cette technique inutilisable à haut débit;
- L'interface comprenant un capillaire de transmission et des funnels est montée sur des Q-TOF MicroTOF-Q et maXis de Bruker Daltonics. L'application d'une énergie au niveau de la lentille qui se situe entre les deux funnels grâce à l'option ISCID permet d'obtenir facilement des ions diagnostiques par fragmentation en source. La nouvelle génération de trappe ionique amaZon

(Bruker Daltonics) possède également ce type d'interface mais l'option ISCID n'est pas incrémentée au niveau du logiciel de contrôle de l'instrument. Les premiers tests de CID en source effectués sur cet appareil n'ont pas permis de détecter d'ions diagnostiques;

- L'interface de type Z-spray a été testée sur une Q-TOF Synapt G1 de Waters. La tension de cône pour obtenir une intensité optimale des ions diagnostiques est de 60 à 70 V. Ce type de source permet également d'alterner haute et basse tension au niveau du cône d'échantillonnage, il est alors possible d'obtenir à la fois des glycopeptides entiers à basse tension et des ions diagnostiques issus de leur fragmentation à haute tension. Ivanic et al utilisent cette technique sur un ESI-TOF (LCT, Waters) pour identifier et quantifier les glycopeptides de l'AGP [Ivanic et al. 2010].

La génération et la transmission d'ions diagnostiques par CID en source de glycopeptides est très dépendante de la géométrie de l'interface, mais la détection des ions diagnostiques est également étroitement liée à la géométrie de l'analyseur utilisé.

## 2.2. La génération d'ions diagnostiques par CID dans la cellule de collision

Les ions diagnostiques peuvent également être obtenus par CID des glycopeptides dans la cellule de collision. Afin de déterminer l'instrument le plus adapté à la génération et à la détection des ions diagnostiques, j'ai comparé deux types d'analyseurs : les trappes ioniques et les Q-TOF.

### 2.2.1. Les trappes ioniques

Au cours de ce travail, deux types de trappes ioniques ont été utilisées : des trappes HCT (Plus et Ultra, Bruker Daltonics) et la dernière génération, une trappe amaZon (Bruker Daltonics).

Avec ce type d'analyseur, les ions sont piégés dans la trappe par un puits de potentiel. La profondeur du puits augmente avec la taille des ions à analyser et il est difficile de piéger efficacement à la fois les ions de bas et de haut poids moléculaire. Lors d'analyses MS/MS, cela se traduit par ce que l'on appelle le "cut off" (Introduction, Chapitre 2, 1.2.2). La profondeur du puits de potentiel est optimisée pour piéger l'ion parent puis l'énergie de fragmentation est appliquée. Au moment de l'éjection des ions, les ions de petite taille (< 30% masse du parent) adoptent une trajectoire instable et s'écrasent sur les parois de la trappe. Ainsi, pour un ion parent de  $m/z$  1500, les ions fragments inférieurs à  $m/z$  450 sont peu intenses voire totalement absents du spectre. L'option PAN qui est disponible sur la nouvelle génération de trappe amaZon permet en théorie de diminuer le "cut off" à 17%.

Les glycopeptides sont généralement des ions relativement lourds (>  $m/z$  1000) tandis que les ions diagnostiques sont de petite taille (<  $m/z$  700). Par conséquent, les plus petits ions oxoniums ( $m/z$  163 et 204) issus de la fragmentation de glycopeptides sont très rarement observés dans les spectres de fragmentation de glycopeptides avec des analyseurs de type trappe ionique.

De plus, les expériences effectuées sur les trappes testées au laboratoire ont permis de montrer qu'il est difficile d'analyser des glycopeptides ayant des masses supérieures à  $m/z$  1500 avec ce type d'instrument. Les spectres générés montrent des ions qui ne sont pas suffisamment intenses pour être isolés puis fragmentés en MS/MS pour générer des ions diagnostiques.

Enfin, la précision de mesure de masse ( $\pm 0,25$  Da) et la résolution des analyseurs de type trappe ionique est assez basse (700 à 1500 de résolution pour les ions monochargés 366 et 657 sur la trappe HCT Plus) et peut provoquer une ambiguïté dans l'identification des ions diagnostiques particulièrement dans le cas de mélanges complexes. En effet, il est possible d'observer des interférences isobariques avec des ions issus de la fragmentation de peptides non glycosylés. Dans ce cas la présence de plusieurs ions diagnostiques dans un même spectre de fragmentation peut permettre de confirmer qu'il s'agit bien d'ions issus de la fragmentation de glycopeptides.

Récemment, le laboratoire a fait l'acquisition d'une nouvelle génération de trappe ionique : la trappe amaZon (Bruker Daltonics) qui permet d'obtenir une meilleure résolution et une meilleure précision de mesure de masse. Cet instrument est également plus sensible d'un facteur 10 par rapport à la génération précédente (HCT Ultra, Bruker Daltonics). Le "cut off" peut également être diminué à 17% de la masse du parent sans perte de sensibilité grâce à l'option PAN et à la "Fast eXcitation Dissociation". L'ensemble de ces améliorations devrait permettre une meilleure détection des glycopeptides et des ions diagnostiques issus de leur fragmentation par MS/MS.

### 2.2.2. Les Q-TOF

Plusieurs instruments de type Q-TOF sont disponibles au laboratoire. Les expériences présentées ici ont été effectuées sur trois instruments différents : un Synapt G1 (Waters), un micrOTOF-Q et un maXis (Bruker Daltonics).

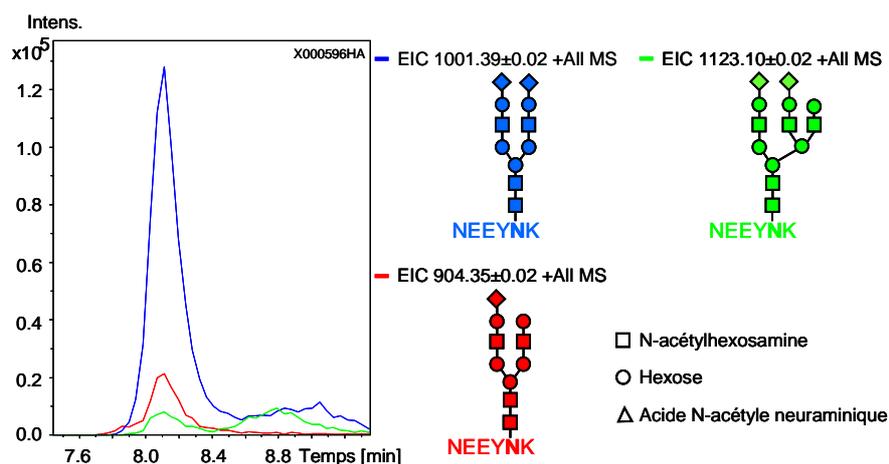
Sur les Q-TOF la gamme dynamique au niveau de la mesure de masse est uniquement limitée par le quadripôle car en théorie, les analyseurs TOF ont une gamme de masse illimitée. Ainsi, contrairement aux trappes ioniques, il n'y a pas de phénomène de "cut off" lorsque l'on effectue une analyse MS/MS. Les ions diagnostiques sont donc parfaitement visibles dans les spectres de fragmentation. De plus, les Q-TOF sont généralement très résolutifs et très précis en mesure de masse, ce qui permet de lever toute ambiguïté dans la reconnaissance des ions oxoniums en évitant les interférences isobariques.

La géométrie des analyseurs de type Q-TOF permet d'effectuer trois types d'acquisitions :

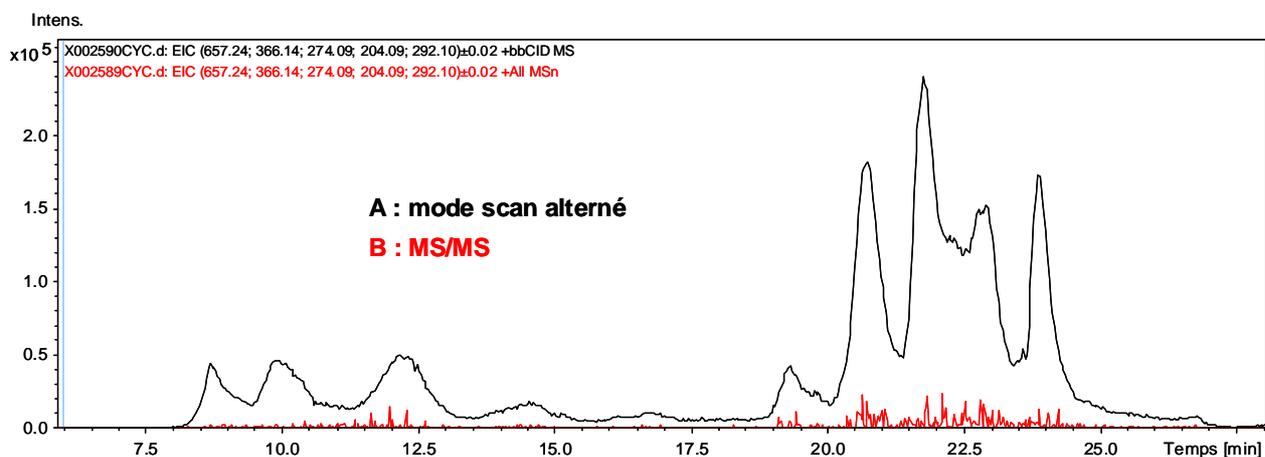
- La **MS** : le quadripôle est en mode transmission et l'énergie dans la cellule de collision est généralement basse et fixe afin de mesurer la masse des ions sans les fragmenter;

- La **MS/MS** : le quadripôle isole un parent qui est ensuite fragmenté dans la cellule de collision à haute énergie afin d'obtenir un spectre de fragmentation par ion sélectionné;
- La mode **scan alterné** : le quadripôle est en mode transmission et deux énergies (une haute et une basse) sont appliquées alternativement dans la cellule de collision, on obtient alors un spectre avec les ions intacts et un spectre avec les fragments de l'ensemble des ions (Figure 15).

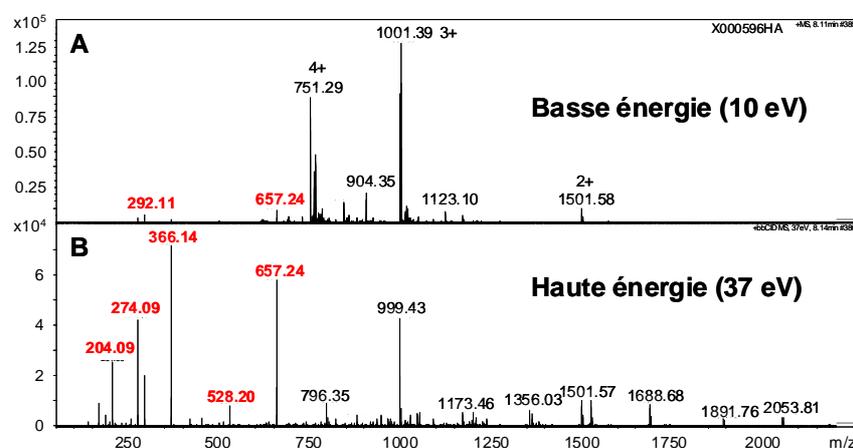
La plupart des ions oxoniums sont communs à l'ensemble des glycosylations et donc, généralement, toutes les glycoformes portées par le même peptide génèrent les mêmes ions diagnostiques. Lors de l'analyse d'un digestat d'une glycoprotéine par HPLC en phase inverse (C18), les glycoformes d'un même glycopeptide sont très faiblement séparées chromatographiquement (Figure 13). En effet, l'élution d'un glycopeptide est essentiellement fonction de sa séquence en acides aminés. Ainsi, en appliquant une énergie suffisante pour fragmenter toutes les glycoformes simultanément, l'intensité de l'ensemble des ions oxoniums générés est cumulée. Les ions diagnostiques sont donc plus facilement détectables que lorsqu'ils sont extraits des spectres de fragmentation d'une analyse LC-MS/MS (Figure 14).



**Figure 13.** Extraction des chromatogrammes des ions de trois glycoformes d'un même glycopeptide, provenant de l'analyse nanoLC-MS d'un digestat total d'AGP



**Figure 14.** Courants des ions diagnostiques ( $m/z$  657,24; 366,14; 292,10; 274,09; 204,09) extraits d'une nanoLC-MS en mode scan alterné (A) et d'une nanoLC-MS/MS (B), d'un digestat trypsique d'AGP, sur un Q-TOF maXis (Bruker Daltonics)



**Figure 15.** Spectres d'un glycopeptide de l'AGP à basse et haute énergie au niveau de la cellule de collision sur un maXis (Bruker Daltonics)

La Figure 15 correspond à l'analyse LC-MS en scan alterné d'un digestat d'AGP. Dans ce cas, basse et haute énergie dans la cellule de collision sont alternées au sein d'une même analyse. Chaque spectre correspond à un temps d'accumulation de 1 s, et toutes les secondes l'énergie au sein de la cellule de collision change. On obtient alors un spectre avec les glycopeptides intacts à basse énergie (Figure 15A) et un spectre avec les ions diagnostiques issus de la fragmentation de ces glycopeptides à haute énergie (Figure 15B). Ceci permet d'obtenir au sein d'une seule analyse, à la fois les ions diagnostiques et les ions des glycopeptides entiers correspondants. Néanmoins, les ions diagnostiques ( $m/z$  292,11 et 657,24) visibles sur le spectre de la Figure 15A montrent que malgré une énergie basse dans la cellule de collision les glycopeptides subissent tout de même une légère fragmentation.

### 2.2.3. Discussion et conclusion

Au cours de cette étude, j'ai comparé les analyseurs de type trappe ionique et Q-TOF pour la génération d'ions diagnostiques par CID de glycopeptides dans la cellule de collision. Les résultats obtenus montrent clairement que la gamme de masse, la résolution et la précision de mesure de masse des Q-TOF testés ici, en font des instruments plus adaptés que les trappes ioniques pour effectuer ces expériences. Cependant, bien que très efficace pour obtenir des ions diagnostiques, particulièrement en scan alterné, les Q-TOF testés génèrent parfois des spectres MS/MS de moindre qualité par rapport à ceux obtenus avec une trappe ionique. Ces spectres de fragmentation sont alors moins informatifs pour la caractérisation des glycopeptides.

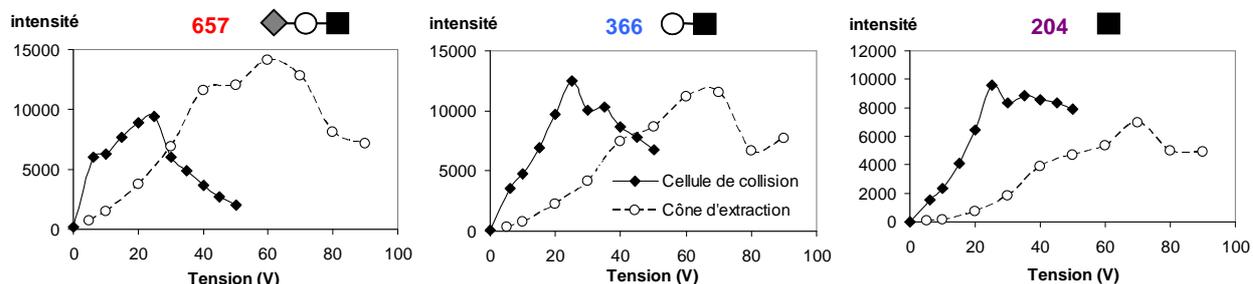
Les expériences réalisées sur les trappes ioniques HCT Plus et Ultra (Bruker Daltonics) ont permis de montrer que ces instruments ne permettent pas de piéger efficacement à la fois les ions de masses élevées (plus de 3000 Da environ) et les ions de basses masses (moins de 400 Da) qui sont les ions diagnostiques. Ce phénomène de "cut off" limite donc grandement la détection des ions diagnostiques en MS/MS. Cependant, la nouvelle génération amaZon (Bruker Daltonics) pourrait permettre de limiter le phénomène de "cut off" grâce à l'option PAN. Des expériences devront être menées afin de vérifier son efficacité.

### 2.3. CID en source vs CID dans la cellule de collision pour un Q-TOF

Nous avons montré que, dans l'état actuel de l'instrumentation, les géométries Q-TOF testées permettent de générer plus efficacement des ions diagnostiques. Nous allons maintenant comparer, pour le Q-TOF Synapt G1 de Waters, les résultats obtenus par collisions en source et ceux obtenus par CID dans la cellule de collision.

#### 2.3.1. Intensité des ions diagnostiques en fonction du mode de fragmentation

Dans la Figure 16, les intensités mesurées des principaux ions diagnostiques générés lors de la fragmentation de glycopeptides de l'AGP pour la CID en source et dans la cellule de collision sont reportés en fonction de la tension appliquée. Les résultats sont différents en fonction de l'ion. Pour l'ion  $m/z$  657 (NeuAc-Hex-HexNAc) qui est constitué de trois sucres, l'intensité maximale est obtenue par CID en source. Pour l'ion  $m/z$  366 (Hex-HexNAc) qui est constitué de deux sucres, les intensités obtenues sont équivalentes pour les deux modes de fragmentation, tandis que pour l'ion  $m/z$  204 (HexNAc) qui est un monosaccharide, c'est la CID dans la cellule de collision qui permet d'obtenir le signal optimal.



**Figure 16.** Intensités des principaux ions diagnostiques issus de la fragmentation des glycopeptides de l'AGP par CID en source ou dans la cellule de collision en fonction de la tension appliquée (Synapt G1, Waters)

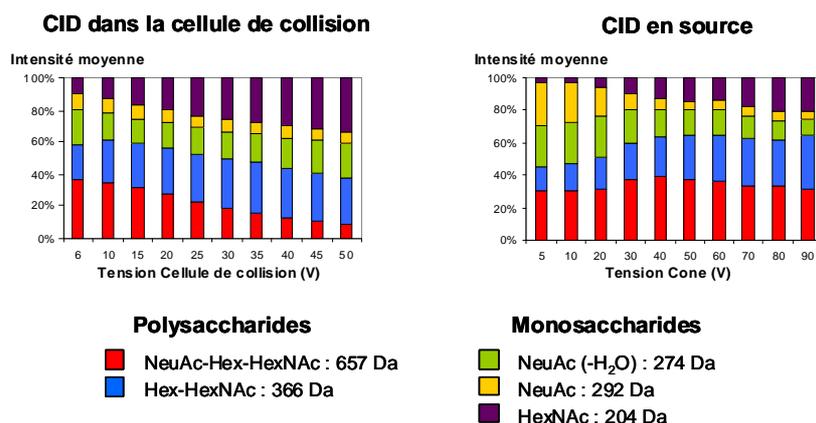
Les résultats indiquent que des ions composés de plusieurs sucres comme l'ion  $m/z$  657 (NeuAc-Hex-HexNAc) peuvent eux même générer des ions plus petits comme l'ion  $m/z$  204 (HexNAc)

par fragmentation secondaire. La littérature montre que ce phénomène a également pu être observé sur un autre type de Q-TOF avec la fragmentation secondaire dans la cellule de collision de l'ion  $m/z$  366 (Hex-HexNAc) en ion  $m/z$  204 (HexNAc) [Jebanathirajah et al. 2003].

Afin d'explorer cette hypothèse nous avons suivi l'évolution de la nature des ions oxoniums générés en fonction de l'énergie appliquée tant au niveau de la source que de la cellule de collision.

### 2.3.2. Type d'ion diagnostique généré en fonction du mode de fragmentation

La Figure 17 décrit les intensités relatives des ions diagnostiques générés par CID en source et dans la cellule de collision en fonction de la tension appliquée. Pour une énergie de fragmentation donnée, les intensités de tous les ions diagnostiques générés sont cumulées et le total est normalisé à 100%. Chaque graphique permet donc de montrer les ions générés en fonction de l'énergie appliquée pour un mode de fragmentation donné, mais aussi de comparer la nature des ions générés en fonction du type de fragmentation, en source ou dans la cellule de collision.



**Figure 17.** Intensités relatives des ions diagnostiques en fonction de l'énergie de fragmentation et du mode de fragmentation (CID dans la cellule de collision/CID en Source). Expériences effectuées sur un Q-TOF (Synapt G1, Waters)

Ainsi lorsque les ions diagnostiques sont générés dans la cellule de collision, l'augmentation de l'énergie de fragmentation provoque la diminution des polysaccharides et l'augmentation des monosaccharides et plus particulièrement de l'ion  $m/z$  204. Ce résultat indique que l'augmentation de l'énergie dans la cellule de collision provoque la fragmentation secondaire des polysaccharides en monosaccharides.

Dans le cas de la CID en source, l'intensité relative de l'ion  $m/z$  204 augmente aussi avec la tension appliquée sur le cône d'extraction, mais au détriment des ions correspondant aux acides sialiques, tandis que l'intensité de l'ion  $m/z$  366 augmente et celle de l'ion  $m/z$  657 reste stable. L'ion  $m/z$  204 n'est donc pas issu dans ce cas de la fragmentation des polysaccharides.

Ces résultats montrent qu'avec ce type de Q-TOF (Synapt G1, Waters), la CID dans la cellule de collision est plus énergétique que la CID en source et permet de générer plus d'ions monosaccharides.

### 2.3.3. Discussion et conclusion

Les expériences réalisées sur le Q-TOF Synapt G1 (Waters) ont permis de montrer que la nature des ions diagnostiques et leur intensité varient selon le mode de fragmentation. La CID dans la cellule de collision permet de générer plus d'ions monosaccharides que la CID en source.

Néanmoins, l'utilisation de la CID en source nécessite moins d'optimisations que la CID dans la cellule de collision. En effet, pour fragmenter des glycopeptides qui sont des ions plus lourds que des peptides classiques, l'ensemble de la transmission doit être réglé en conséquence. Tandis que pour faire de la CID en source, il est possible de conserver des paramètres classiques utilisés dans le cadre de l'analyse de peptides. Le seul paramètre à modifier est la tension du cône d'extraction.

Ces conclusions sur le mode de fragmentation le plus efficace ne sont valables que pour le Q-TOF Synapt G1 et cette étude comparative devra être poursuivie avec les autres géométries de Q-TOF disponibles au laboratoire (micrOTOF-Q, maXis, Bruker Daltonics et Synapt G2, Waters).

## 2.4. Discussion et conclusion

L'ensemble des expériences réalisées ont permis de comparer les différents instruments disponibles au laboratoire en tenant compte à la fois de l'interface et de l'analyseur de chaque spectromètre de masse. Ainsi, ces résultats montrent qu'en fonction de l'instrument utilisé le mode de fragmentation permettant d'obtenir une intensité optimale des ions diagnostiques est différent :

- Dans le cas des instruments **HCT Plus et Ultra** (Bruker Daltonics) qui sont de type trappe ionique, l'interface (capillaire + skimmer) n'est pas adaptée pour effectuer de la fragmentation en source à haut débit. Ainsi, le meilleur moyen pour obtenir des ions diagnostiques est de fragmenter les glycopeptides par MS/MS. La détection des ions diagnostiques est cependant limitée par le phénomène de "cut off". L'option PAN incrémentée sur la nouvelle génération de trappe amaZon (Bruker Daltonics) doit permettre de diminuer ce "cut off" sans perte d'intensité et devrait être testée afin de compléter ce travail;
- Les expériences menées sur le **Q-TOF Synapt G1** (Waters) ont permis de montrer qu'avec cet instrument, les ions diagnostiques peuvent être obtenus à la fois par CID en source et dans la cellule de collision. Cependant la génération d'ions oxoniums par CID en source nécessite moins d'optimisation et est donc plus facile à mettre en place;
- Avec les **Q-TOF Bruker (MicrOTOF-Q et maXis)** il est également possible d'effectuer de la CID en source sur les glycopeptides (ISCID). Néanmoins, de très bons résultats ont été obtenus en effectuant du scan alterné au niveau de la cellule de collision ce qui permet d'obtenir à la fois des

ions diagnostiques et des glycopeptides intacts. Cette méthode a été utilisée dans la cadre d'une collaboration décrite dans la Partie 2 des Résultats Chapitre 3.

La glycoprotéine utilisée comme modèle (AGP) porte des glycosylations de type complexes sialylées. L'étude pourrait être poursuivie en utilisant une autre glycoprotéine possédant des glycosylations de type high-mannose ou hybride afin d'étudier et d'optimiser la génération d'ions diagnostiques pour ces autres types de glycosylations. Néanmoins, au cours de l'étude décrite dans la Partie 2 des Résultats Chapitre 3 de ce manuscrit, la méthode du scan alterné sur le maXis a été utilisée avec de très bons résultats sur la glycoprotéine GAP50 de *Toxoplasma gondii* qui porte des glycosylations de type high-mannose.

Des méthodes alternatives existent pour palier au problème de détection des ions diagnostiques générés par MS/MS sur les trappes ioniques. Par exemple, Wuhrer et al caractérisent des glycopeptides à l'aide d'une trappe ionique Esquire (Bruker Daltonics), en détectant leurs zones d'élution en extrayant le chromatogramme correspondant aux pertes de neutres [Wuhrer et al. 2005]. En effet, lorsque les glycosylations se fragmentent, les ions majoritaires du spectre MS/MS correspondent à des pertes de neutres de sucres. Ainsi, la recherche de ces pertes de neutres est une alternative à l'utilisation des ions oxoniums comme marqueurs. Cependant, les glycopeptides se présentent souvent sous la forme d'ions multichargés. Les pertes de neutres génèrent donc des fragments multichargés. Ainsi, afin d'être exhaustif, il faut extraire le chromatogramme correspondant aux pertes de neutres pour l'ensemble des sucres constituant les différents types de glycosylations (Hex, dHex, HexNAc, NeuAc, etc.) et à tous les états de charge observables (mono, di, tri-chargés). La recherche d'ions oxoniums est nettement plus aisée. Les ions diagnostiques sont monochargés et la simple recherche des ions  $m/z$  204 (HexNAc) et  $m/z$  366 (Hex-HexNAc) suffit car ils sont communs à tous les types de glycosylations [Huddleston et al. 1993].

En conclusion, le meilleur choix pour obtenir les courants d'ions diagnostiques les plus intenses, tout en ayant des spectres MS/MS de bonne qualité, est actuellement le Q-TOF MaXis de Bruker, parmi les instruments testés au laboratoire.

Sur les trappes ioniques testées, les spectres MS/MS des glycopeptides sont de très bonne qualité. Malheureusement, l'intensité du courant d'ions diagnostiques, qu'ils soient générés en source ou par MS/MS, est souvent insuffisante pour détecter l'élution des glycopeptides.

### 3. Stratégie développée pour la caractérisation des protéines N-glycosylées

Grâce aux optimisations réalisées pour l'obtention d'ions diagnostiques par fragmentation de glycopeptides, une méthodologie a été développée qui comporte trois étapes principales : la digestion enzymatique de l'échantillon suivi des analyse LC-MS et LC-MS/MS du digestat (Figure 18).

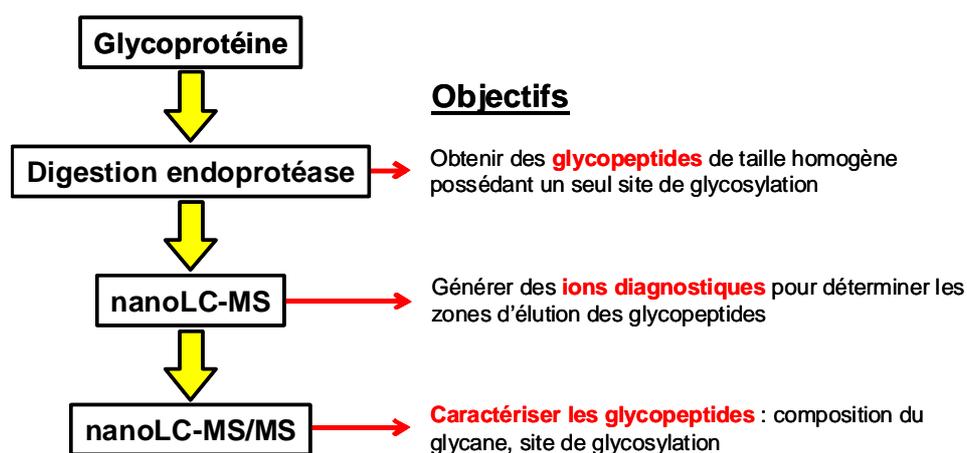


Figure 18. Schéma décrivant la méthodologie développée pour la caractérisation des glycoprotéines

#### 3.1. Digestion enzymatique : obtention de glycopeptides

Le choix de l'endoprotéase est critique. Cette enzyme doit permettre de générer des glycopeptides ne portant qu'un seul site de glycosylation, ayant une taille compatible avec les analyses de masse tout en conservant une séquence suffisamment informative.

La trypsine est l'enzyme la plus couramment utilisée, il arrive cependant que les peptides générés soient de trop grande taille (> 10 acides aminés) ou portent plusieurs sites de glycosylation. L'encombrement stérique provoqué par les glycosylations peut empêcher l'accès aux sites de clivage. De plus, lorsque la glycoprotéine possède plusieurs sites de glycosylation, les glycopeptides générés peuvent avoir des tailles très hétérogènes ce qui complique l'optimisation de la méthode d'analyse par spectrométrie de masse. L'utilisation d'une ou de plusieurs autres endoprotéases est alors nécessaire (GluC, Chymotrypsine, AspN, etc.) pour remplacer ou compléter la digestion trypsique.

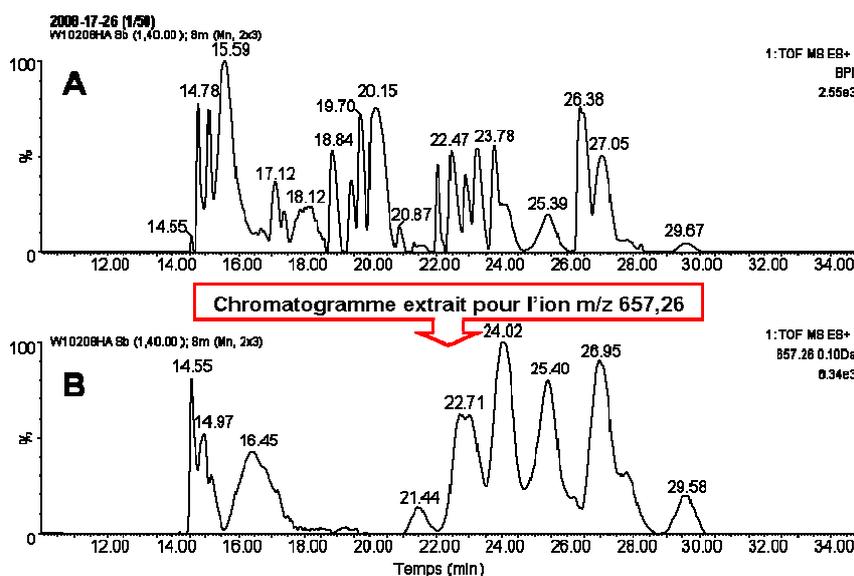
Pour déterminer précisément la position du site de glycosylation dans la glycoprotéine, la séquence en acides aminés du glycopeptide doit être spécifique. Pour cela un peptide de minimum 5 acides aminés est généralement nécessaire.

Ainsi, le glycopeptide "idéal" possède entre 5 et 10 acides aminés et un seul site de glycosylation. Afin d'obtenir ce type de peptide, une ou plusieurs endoprotéases peuvent être utilisées. Une digestion *in silico* de la glycoprotéine permet d'effectuer le bon choix.

### 3.2. Analyse LC-MS : génération d'ions diagnostiques

Le mélange de peptides et de glycopeptides obtenu après digestion enzymatique de la glycoprotéine est analysé par LC-MS en utilisant une méthode permettant de générer des ions diagnostiques. La détection de ces ions permet de localiser les zones d'élution des glycopeptides (§ 1.2).

La méthode utilisée pour générer les ions diagnostiques dépend de l'instrument choisi (§ 2). Par exemple, dans le cas d'une analyse nanoLC-MS sur un Q-TOF Synapt G1 (Waters), la manière la plus simple et la plus efficace d'obtenir des ions diagnostiques est d'effectuer de la CID en source en appliquant une tension de 40 à 60 V au niveau du cône d'échantillonnage. Les tensions appliquées sur le quadripôle sont optimisées pour transmettre les ions diagnostiques et la gamme de masse est réduite à  $m/z$  50-1000, afin d'améliorer la sensibilité pour leur détection.



**Figure 19.** (A) BPI (Base Peak Ion) de l'analyse nanoLC-MS sur un Q-TOF Synapt G1 d'un digestat tryptique d'AGP; (B) Chromatogramme extrait de l'ion  $m/z$  657,26

La Figure 19 montre l'analyse nanoLC-MS d'un digestat tryptique d'AGP sur un Synapt G1 en utilisant la méthode décrite ci-dessus. L'extraction de l'ion diagnostique majoritaire ( $m/z$  657,26) permet de visualiser les zones du chromatogramme qui contiennent des glycopeptides. Cette méthode permet de montrer clairement la présence de glycopeptides dans un mélange complexe.

### 3.3. Analyse LC-MS/MS : caractérisation des glycopeptides

Après avoir déterminé les zones d'élution à l'aide des ions diagnostiques, le digestat est analysé par LC-MS/MS afin de caractériser les glycopeptides. Les spectres MS/MS des glycopeptides permettent de déterminer la composition du glycane, le site de glycosylation et la séquence en acides aminés.

Les spectres MS/MS des glycopeptides sont recherchés au niveau des zones d'élution des ions diagnostiques. Par exemple, dans la Figure 19B, l'ion  $m/z$  657 est intense à 14,5 min, ce qui indique que des glycopeptides sont élués dans cette zone du chromatogramme. Le spectre MS/MS de l'un d'entre eux ( $m/z$  1501,83) est présenté dans la Figure 20. Plusieurs ions diagnostiques sont observables ( $m/z$  204,13, 274,14, 366,21, 528,30 et 657,36), ce spectre correspond donc bien à la fragmentation d'un glycopeptide. L'interprétation du spectre MS/MS en assignant les ions fragments est effectuée manuellement. En effet, actuellement, il n'existe pas de logiciel permettant d'interpréter automatiquement les spectres MS/MS de glycopeptides. Les fragments majoritaires permettent de déterminer la structure du glycane ( $\text{HexNAc}_4\text{Hex}_5\text{NeuAc}_2$ ) et la masse du peptide non glycosylé ( $(M+H)^+ = 796,5$  Da). Afin d'identifier le site de glycosylation, la glycoprotéine (AGP) est digérée *in silico* pour prédire les séquences des glycopeptides potentiels. Les masses théoriques des peptides contenant un site de glycosylation (Tableau 1, § 1.1) sont comparées aux masses mesurées dans le spectre MS/MS. Dans le cas de l'ion  $m/z$  1501,83, la masse du peptide non glycosylé étant  $(M+H)^+ = 796,5$  Da, sa séquence est NEEYNK et correspond au site de glycosylation Asn<sup>38</sup> de la glycoprotéine AGP. Les résultats sont validés en tenant compte de l'interprétation des trois types d'ions fragments classiquement générés par les glycopeptides (perte de neutre de sucre, fragments internes du peptide et fragments internes du glycane).

Ainsi, l'ion  $m/z$  1501,83 correspond au glycopeptide <sup>34</sup>NEEYNK<sup>39</sup> portant une N-glycosylation de type complexe bi-antennée avec deux acides sialiques, en position Asn<sup>38</sup>.

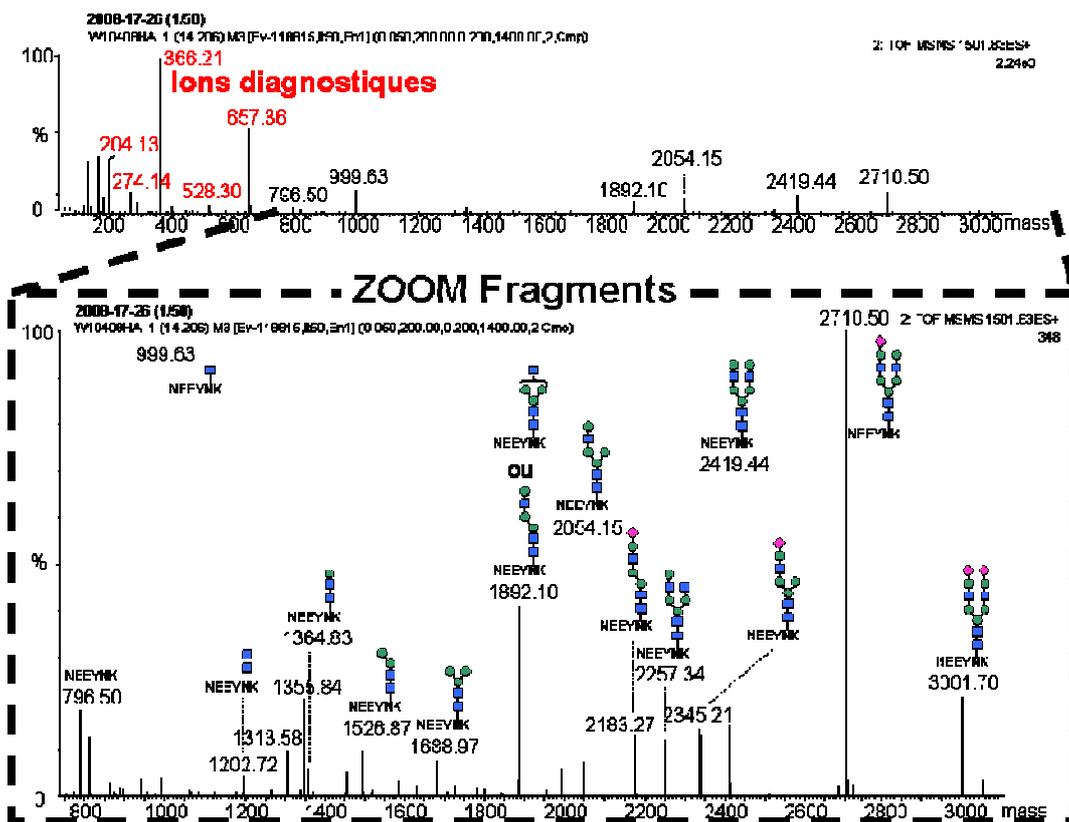


Figure 20. Spectre MS/MS déconvolué de l'ion dichargé m/z 1501,83

La masse de l'ion correspondant au peptide non glycosylé n'est pas toujours présent sur le spectre MS/MS mais peut être calculée à partir de la masse de l'ion correspondant au peptide portant un GlcNAc qui est souvent le fragment majoritaire. Les ions correspondants à la fragmentation du peptide non glycosylé peuvent être observés mais cela reste rare.

Les glycopeptides ont souvent des masses moléculaires qui dépassent 2000 Da et donc leur massif isotopique est plus complexe que pour les peptides tryptiques non glycosylés qui ont souvent des masses comprises entre 500 et 1500 Da. La détermination de l'état de charge est donc, de façon générale, plus difficile pour les glycopeptides. Il est donc important d'utiliser un spectromètre de masse ayant une résolution suffisante pour déterminer les états de charge des ions pour faciliter l'interprétation des spectres et ainsi limiter le nombre de faux-positifs dans l'identification des glycopeptides.

## 4. Conclusion

L'objectif de ce travail était de développer une méthodologie permettant de caractériser des protéines N-glycosylées présentes en petites quantités dans des échantillons biologiques complexes. Pour cela, nous avons mis au point une stratégie permettant d'obtenir le maximum d'informations sur les glycoprotéines présentes dans un échantillon inconnu et disponible en quantité si faible que seulement 1 ou 2 analyses peuvent être réalisées. Nous avons adapté les méthodes de protéomique classique comme la digestion enzymatique, la LC-MS et la LC-MS/MS en tirant partie au maximum de la détection d'ions diagnostiques issus de la fragmentation spécifique des glycosylations par CID en source ou dans la cellule de collision.

L'utilisation de ces ions diagnostiques étaient déjà décrites dans la littérature depuis une vingtaine d'années mais étaient très peu utilisés, par manque de sensibilité des instruments. Nous avons déterminé les meilleurs modes opératoires pour plusieurs types d'instrument afin d'obtenir des courants d'ions diagnostiques intenses qui permettent de localiser les zones d'élution des glycopeptides dans les chromatogrammes. Nous illustrerons l'intérêt d'utiliser ces ions diagnostiques pour la caractérisation de glycoprotéines dans le chapitre 3.

Une étude comparative de différents spectromètres de masse a été réalisée pour la génération des ions diagnostiques. Nous avons ainsi pu montrer que le mode de fragmentation pour obtenir la meilleure intensité pour les ions diagnostiques dépend à la fois de la géométrie de l'interface et de celle de l'analyseur. Ainsi pour les trappes ioniques testées, la méthode la plus efficace pour générer des ions diagnostiques est de fragmenter les glycopeptides en MS/MS. Pour le Q-TOF Synapt G1 (Waters), il est préférable d'effectuer de la CID en source, et pour les Q-TOF MicroTOF-Q et maXis (Bruker Daltonics), l'intensité optimale des ions diagnostiques est obtenue par CID dans la cellule de collision et plus précisément en effectuant du scan alterné à basse et haute énergie.

Grâce à ces optimisations, j'ai pu développer une stratégie, permettant de caractériser des glycoprotéines dans des mélanges complexes, qui comporte trois étapes :

- La **digestion enzymatique** de la glycoprotéine permet d'obtenir un ensemble de peptides et de glycopeptides. Une ou plusieurs endoprotéase peuvent être utilisée afin de générer des glycopeptides portant un seul site de glycosylation, ayant une taille compatible avec les analyses par spectrométrie de masse et une séquence en acides aminés spécifique permettant de déterminer précisément le site de glycosylation;
- Le digestat est analysé par **LC-MS** avec une méthode de masse permettant de générer des **ions diagnostiques**. La détection de ces ions permet de déterminer les zones d'élution des glycopeptides;

- Le digestat est analysé par **LC-MS/MS** afin de fragmenter les glycopeptides et de les caractériser. L'interprétation des spectres de fragmentation permet de déterminer la séquence du glycane, la séquence du peptide et le site de glycosylation.

Cette méthodologie peut être améliorée si l'on combine les analyses LC-MS et LC-MS/MS. Par exemple, Mechref et al utilisent le scan alterné dans la cellule de collision sur un Q-TOF Global Micromass (Waters) [Mechref et al. 2005] associé au mode PID (Parent Ion Discovery) pour caractériser les glycopeptides en une seule analyse. Avec le mode PID, lorsque des ions diagnostiques sont détectés, le Q-TOF passe automatiquement en mode MS/MS pour isoler et fragmenter les glycopeptides. L'utilisation de cette option, disponible sur les Synapt G1 et G2 (Waters) du laboratoire, pourrait permettre d'effectuer les étapes de détection de glycosylations à l'aide des ions diagnostiques et la MS/MS en une seule analyse. Cela permettrait non seulement de gagner du temps d'analyse mais surtout d'utiliser moins d'échantillon. Cependant, les premiers tests effectués dans ce sens sur le Synapt G1 n'ont pas permis d'obtenir une fragmentation MS/MS de qualité suffisante pour caractériser les glycopeptides.

L'automatisation des interprétations des spectres de fragmentation MS/MS des glycopeptides constitue l'un des facteurs limitant pour la caractérisation des glycoprotéines par LC-MS/MS. En effet, les outils bioinformatiques disponibles pour la protéomique classique ne permettent pas, pour l'instant, de gérer des PTMs complexes telles que les glycosylations. Il est donc indispensable de disposer d'algorithmes qui permettraient d'automatiser l'interprétation des données MS/MS de tous les types de spectromètres de masse (et pas seulement ceux qui génèrent des données avec une précision de mesure de masse élevée). Plusieurs algorithmes ont été développés les dernières années [Goldberg et al. 2007, Deshpande et al. 2010, Wu et al. 2010]. Cependant, il n'existe pas, pour le moment, de programme performant permettant d'automatiser entièrement l'interprétation des spectres de fragmentation des glycopeptides

La stratégie que nous avons développée pour la caractérisation des protéines N-glycosylées à été appliquée dans la cadre d'une collaboration pour la caractérisation de la glycoprotéine GAP50 du parasite *Toxoplasma gondii*. Ce travail est détaillé dans la Partie 2 Chapitre 3.

# **BIBLIOGRAPHIE**



- Aebersold, R. (2009). "A stress test for mass spectrometry-based proteomics." *Nat Methods* 6(6): 411-412.
- Bell, A. W., Deutsch, E. W., Au, C. E., Kearney, R. E., Beavis, R., Sechi, S., Nilsson, T. et Bergeron, J. J. (2009). "A HUPO test sample study reveals common problems in mass spectrometry-based proteomics." *Nat Methods* 6(6): 423-430.
- Budai, L., Ozohanics, O., Ludányi, K., Drahos, L., Kremmer, T., Krenyacz, J. et Vékey, K. (2009). "Investigation of genetic variants of  $\alpha$ -1 acid glycoprotein by ultra-performance liquid chromatography–mass spectrometry." *Analytical and Bioanalytical Chemistry* 393(3): 991-998.
- Bynum, M. A., Yin, H., Felts, K., Lee, Y. M., Monell, C. R. et Killeen, K. (2009). "Characterization of IgG N-Glycans Employing a Microfluidic Chip that Integrates Glycan Cleavage, Sample Purification, LC Separation, and MS Detection." *Anal Chem*.
- Carapito, C. (2006). "Vers une meilleure utilisation des données de spectrométrie de masse en analyse protéomique." Université Louis Pasteur de Strasbourg.
- Carr, S. A., Huddleston, M. J. et Bean, M. F. (1993). "Selective identification and differentiation of N- and O-linked oligosaccharides in glycoproteins by liquid chromatography-mass spectrometry." *Protein Sci* 2(2): 183-196.
- Conboy, J. J. et Henion, J. D. (1992). "The determination of glycopeptides by liquid chromatography/mass spectrometry with collision-induced dissociation." *Journal of the American Society for Mass Spectrometry* 3(8): 804-814.
- Deshpande, N., Jensen, P. H., Packer, N. H. et Kolarich, D. (2010). "GlycoSpectrumScan: fishing glycopeptides from MS spectra of protease digests of human colostrum sIgA." *J Proteome Res* 9(2): 1063-1075.
- Elias, J. E., Haas, W., Faherty, B. K. et Gygi, S. P. (2005). "Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations." *Nat Methods* 2(9): 667-675.
- Fournier, T., Medjoubi, N. N. et Porquet, D. (2000). "Alpha-1-acid glycoprotein." *Biochim Biophys Acta* 1482(1-2): 157-171.
- Fritz, R., Ruth, W. et Kragl, U. (2009). "Assessment of acetone as an alternative to acetonitrile in peptide analysis by liquid chromatography/mass spectrometry." *Rapid Commun Mass Spectrom* 23(14): 2139-2145.
- Giorgianni, F., Cappiello, A., Beranova-Giorgianni, S., Palma, P., Trufelli, H. et Desiderio, D. M. (2004). "LC-MS/MS analysis of peptides with methanol as organic modifier: improved limits of detection." *Anal Chem* 76(23): 7028-7038.
- Goldberg, D., Bern, M., Parry, S., Sutton-Smith, M., Panico, M., Morris, H. R. et Dell, A. (2007). "Automated N-Glycopeptide Identification Using a Combination of Single- and Tandem-MS." *Journal of Proteome Research* 6(10): 3995-4005.
- Gstaiger, M. et Aebersold, R. (2009). "Applying mass spectrometry-based proteomics to genetics, genomics and network biology." *Nat Rev Genet* 10(9): 617-627.
- Huddleston, M. J., Bean, M. F. et Carr, S. A. (1993). "Collisional fragmentation of glycopeptides by electrospray ionization LC/MS and LC/MS/MS: methods for selective detection of glycopeptides in protein digests." *Anal Chem* 65(7): 877-884.
- Imre, T., Schlosser, G., Pocsfalvi, G., Siciliano, R., Molnar-Szollosi, E., Kremmer, T., Malorni, A. et Vekey, K. (2005). "Glycosylation site analysis of human alpha-1-acid glycoprotein (AGP) by capillary liquid chromatography-electrospray mass spectrometry." *J Mass Spectrom* 40(11): 1472-1483.

Ivancic, M. M., Gadgil, H. S., Halsall, H. B. et Treuheit, M. J. (2010). "LC/MS analysis of complex multiglycosylated human alpha(1)-acid glycoprotein as a model for developing identification and quantitation methods for intact glycopeptide analysis." *Anal Biochem* 400(1): 25-32.

Jebanathirajah, J., Steen, H. et 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(7): 777-784.

Keppel, T., R. , Jacques, M., E. et Weis, D., D. (2010). "The use of acetone as a substitute for acetonitrile in analysis of peptides by liquid chromatography/electrospray ionization mass spectrometry." *Rapid Communications in Mass Spectrometry* 24(1): 6-10.

Mechref, Y., Muzikar, J. et Novotny, M. V. (2005). "Comprehensive assessment of N-glycans derived from a murine monoclonal antibody: a case for multimethodological approach." *Electrophoresis* 26(10): 2034-2046.

Rudnick, P. A., Clauser, K. R., Kilpatrick, L. E., Tchekhovskoi, D. V., Neta, P., Blonder, N., Billheimer, D. D., Blackman, R. K., Bunk, D. M., Cardasis, H. L., Ham, A. J., Jaffe, J. D., Kinsinger, C. R., Mesri, M., Neubert, T. A., Schilling, B., Tabb, D. L., Tegeler, T. J., Vega-Montoto, L., Variyath, A. M., Wang, M., Wang, P., Whiteaker, J. R., Zimmerman, L. J., Carr, S. A., Fisher, S. J., Gibson, B. W., Paulovich, A. G., Regnier, F. E., Rodriguez, H., Spiegelman, C., Tempst, P., Liebler, D. C. et Stein, S. E. (2010). "Performance metrics for liquid chromatography-tandem mass spectrometry systems in proteomics analyses." *Mol Cell Proteomics* 9(2): 225-241.

Sigmon, I., Lee, L. W., Chang, D. K., Krusberski, N., Cohen, D., Eng, J. K. et Martin, D. B. (2010). "ChromEval: a software application for the rapid evaluation of HPLC system performance in proteomic applications." *Anal Chem* 82(12): 5060-5068.

Steen, H. et Mann, M. (2004). "The ABC's (and XYZ's) of peptide sequencing." *Nat Rev Mol Cell Biol* 5(9): 699-711.

Tabb, D. L., Vega-Montoto, L., Rudnick, P. A., Variyath, A. M., Ham, A. J., Bunk, D. M., Kilpatrick, L. E., Billheimer, D. D., Blackman, R. K., Cardasis, H. L., Carr, S. A., Clauser, K. R., Jaffe, J. D., Kowalski, K. A., Neubert, T. A., Regnier, F. E., Schilling, B., Tegeler, T. J., Wang, M., Wang, P., Whiteaker, J. R., Zimmerman, L. J., Fisher, S. J., Gibson, B. W., Kinsinger, C. R., Mesri, M., Rodriguez, H., Stein, S. E., Tempst, P., Paulovich, A. G., Liebler, D. C. et Spiegelman, C. (2010). "Repeatability and reproducibility in proteomic identifications by liquid chromatography-tandem mass spectrometry." *J Proteome Res* 9(2): 761-776.

Treuheit, M. J., Costello, C. E. et Halsall, H. B. (1992). "Analysis of the five glycosylation sites of human alpha 1-acid glycoprotein." *Biochem J* 283 ( Pt 1): 105-112.

Tullo, A. (2008). "A Solvent Dries Up : Acetonitrile is in short supply, and chemists are concerned." *Chemical & Engineering News* 86(47): 27.

Van Midwoud, P. M., Rieux, L., Bischoff, R., Verpoorte, E. et Niederlander, H. A. (2007). "Improvement of recovery and repeatability in liquid chromatography-mass spectrometry analysis of peptides." *J Proteome Res* 6(2): 781-791.

Wu, Y., Mechref, Y., Klouckova, I., Mayampurath, A., Novotny, M. V. et Tang, H. (2010). "Mapping site-specific protein N-glycosylations through liquid chromatography/mass spectrometry and targeted tandem mass spectrometry." *Rapid Communications in Mass Spectrometry* 24(7): 965-972.

Wuhrer, M., Koeleman, C. A., Hokke, C. H. et Deelder, A. M. (2005). "Protein glycosylation analyzed by normal-phase nano-liquid chromatography--mass spectrometry of glycopeptides." *Anal Chem* 77(3): 886-894.

Yin, H. et Killeen, K. (2007). "The fundamental aspects and applications of Agilent HPLC-Chip." *Journal of Separation Science* 30(10): 1427-1434.

Yin, H., Killeen, K., Brennen, R., Sobek, D., Werlich, M. et van de Goor, T. (2004). "*Microfluidic Chip for Peptide Analysis with an Integrated HPLC Column, Sample Enrichment Column, and Nanoelectrospray Tip.*" *Analytical Chemistry* 77(2): 527-533.

Yu, Y. Q., Fournier, J., Gilar, M. et Gebler, J. C. (2007). "*Identification of N-linked glycosylation sites using glycoprotein digestion with pronase prior to MALDI tandem time-of-flight mass spectrometry.*" *Anal Chem* 79(4): 1731-1738.

Zhang, Y., Go, E. P. et Desaire, H. (2008). "*Maximizing Coverage of Glycosylation Heterogeneity in MALDI-MS Analysis of Glycoproteins with Up to 27 Glycosylation Sites.*" *Anal Chem*.



# **Partie 2 : Applications**

**Application des développements méthodologiques  
pour l'identification et la caractérisation  
des protéines et des glycoprotéines  
à différentes thématiques biologiques**





## Chapitre 3

# Identification et caractérisation de protéines et de glycoprotéines du parasite *Toxoplasma gondii*

---

Cette étude a fait l'objet d'une collaboration avec le Docteur Stanislas Tomavo de l'Unité de Glycobiologie Structurale et Fonctionnelle, CNRS UMR 8576, Institut Pasteur, Université des Sciences et Technologies de Lille, Villeneuve d'Ascq.

## 1. Contexte général de l'étude

### 1.1. Introduction

Au cours de ce travail, nous avons étudié le parasite *Toxoplasma gondii*. Plusieurs expériences protéomiques qui présentent des difficultés analytiques communes ont été abordées, à savoir l'analyse de protéines difficiles à purifier et en très petites quantités. C'est pourquoi nous avons utilisé des méthodologies protéomiques adaptées qui ont été développées dans le Chapitre 1 des résultats de ce manuscrit.

Une première étude protéomique globale a été réalisée dans le but d'identifier les protéines impliquées dans le mécanisme d'invasion (le glideosome) de la cellule hôte par le parasite *Toxoplasma gondii*. L'objectif de ce travail a également été de prouver la présence et l'implication de glycoprotéines dans ce complexe macromoléculaire. Par la suite, nous avons caractérisé les

glycosylations et la micro-hétérogénéité de l'une de ces glycoprotéines (GAP50) en utilisant la méthodologie développée dans le Chapitre 2. Enfin, lors d'une deuxième étude protéomique, nous avons étudié les facteurs de transcription impliqués dans l'interconversion du parasite de sa forme virulente (tachyzoïte) en sa forme dormante (bradyzoïte).

## 1.2. Le parasite *Toxoplasma gondii*

*T. gondii* est un protozoaire responsable de la toxoplasmose et appartenant au phylum des Apicomplexa qui inclut de nombreux autres pathogènes comme *Plasmodium falciparum*, responsable de la malaria chez l'homme. Il s'agit d'un parasite intracellulaire obligatoire, c'est-à-dire qu'il doit vivre à l'intérieur d'une cellule pour survivre.

*T. gondii* est un eucaryote unicellulaire haploïde et sa prolifération est obligatoirement intracellulaire car il dépend de la cellule hôte pour un certain nombre de nutriments essentiels pour lesquels il ne dispose pas de voie métabolique propre. Néanmoins, comme toute cellule eucaryote, il possède toute la machinerie nécessaire à la synthèse et au transport de protéines : noyau, réticulum endoplasmique, appareil de Golgi, mitochondries [Dubey et al. 1998].

*T. gondii* peut potentiellement infecter tous les animaux à sang chaud. Son mode de transmission est varié et peut s'effectuer par ingestion orale de kystes présents dans une viande contaminée insuffisamment cuite ou d'oocystes provenant des excréments de félins par exemple.

Les personnes contaminées développent alors la toxoplasmose. Cette infection est la plupart du temps bénigne ou asymptomatique, mais des formes graves peuvent apparaître chez des patients immunodéprimés ou dans des cas de transmission fœtale si la femme enceinte est contaminée en cours de grossesse [Tenter et al. 2000].

*T. gondii* sert de modèle pour l'étude des parasites de la famille des Apicomplexa [Kim et Weiss 2004].

### 1.2.1. Cycle d'infection et de reproduction de *T. gondii*

Le cycle parasitaire (Figure 1) de *T. gondii* comporte une multiplication asexuée, qui s'effectue dans différents tissus chez les homéothermes (mammifères, oiseaux), appelés hôtes intermédiaires et une multiplication sexuée, qui s'effectue dans l'épithélium digestif du chat et de quelques autres félinés (hôtes définitifs).

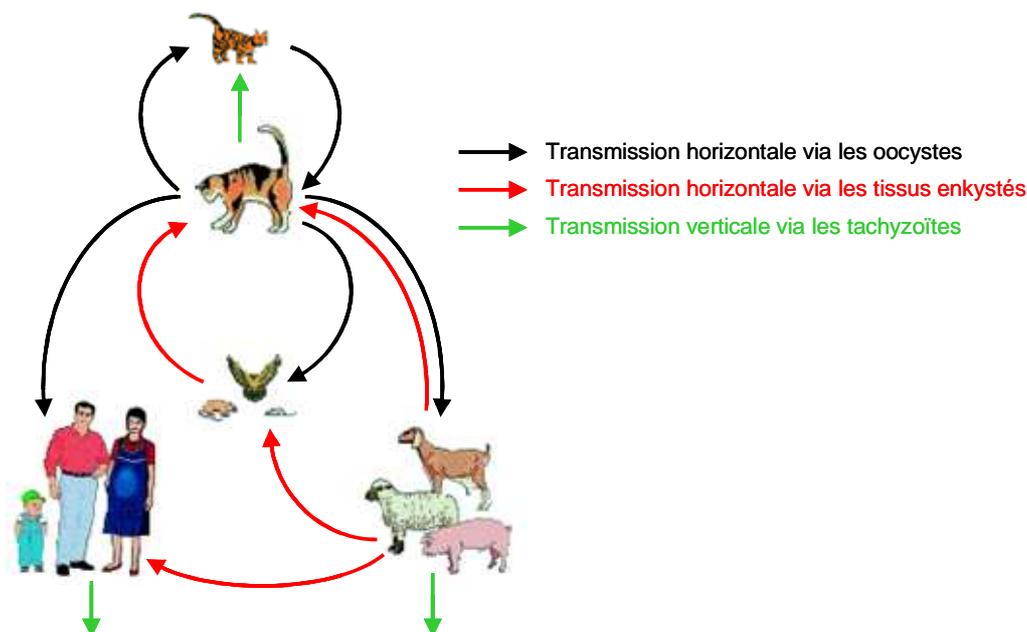


Figure 1. Cycle parasitaire de *Toxoplasma gondii* [Tenter et al. 2000]

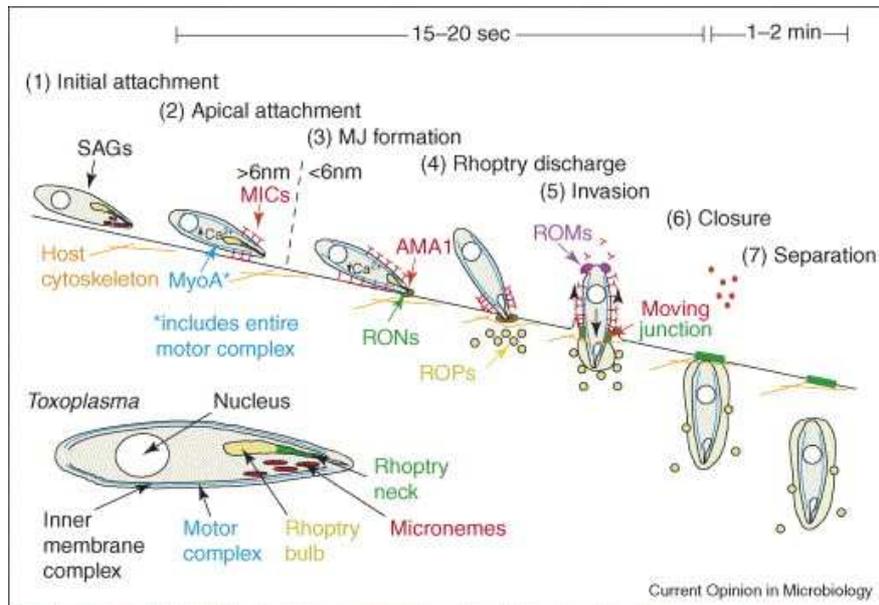
Ce cycle complexe comprend plusieurs stades au cours desquelles le parasite prend différentes formes [Dubey et al. 1998, Tenter et al. 2000] :

- Le stade **tachyzoïte**, forme proliférative infectieuse chez l'hôte intermédiaire, se développe dans des vacuoles transitoires qui peuvent contenir jusqu'à 128 parasites;
- Le stade **bradyzoïte**, forme dormante chez l'hôte intermédiaire, est contenu dans des **kystes** intracellulaires (muscles, système nerveux centrale) pouvant contenir plusieurs milliers de parasites;
  - Le stade mérozoïte, chez l'hôte définitif, est le seul stade capable de reproduction sexuée;
  - Les sporozoïtes, résultat de la reproduction sexuée chez l'hôte définitif, sont libérés dans l'environnement avec les déjections du chat dans des **oocystes**. Les oocystes une fois ingérés par l'Homme, sont lysés dans l'intestin pour libérer les sporozoïtes qui se disséminent dans la circulation sanguine après conversion en tachyzoïtes.

### 1.2.2. Reconnaissance et invasion de la cellule hôte

Le parasite *T. gondii* est capable d'infecter tout type de cellules mammifères. Pourtant, comme l'ensemble des Apicomplexa, il ne possède ni cils ni flagelles, et utilise une machinerie moléculaire unique : le "Glideosome". Il s'agit d'un complexe macromoléculaire composé de protéines adhésives qui sont libérées par différentes organites (les micronèmes, les rhoptries et les granules denses) et transloquées à un endroit donné par l'action d'un système composé de filaments d'actine et de myosine ancrés dans le complexe intermembranaire [Keeley et Soldati 2004] (Figure 2). Une fois à l'intérieur 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 émergent de la cellule hôte en utilisant le même mécanisme que pour l'envahir.



**Figure 2.** Schéma représentant le parasite et son mécanisme d'invasion dans la cellule hôte; d'après [Carruthers et Boothroyd 2007]

## **2. Etude des glycoprotéines impliquées dans le mécanisme de reconnaissance et d'infection de la cellule hôte par le parasite *T. gondii***

### **2.1. Contexte de l'étude**

Ce travail s'inscrit dans le cadre de l'étude du parasite *T. gondii*, et plus particulièrement des interactions hôte-pathogène qui permettent au parasite d'envahir les cellules mammifères. Avec nos collaborateurs, nous avons voulu prouver que les glycosylations jouent un rôle majeur dans le mécanisme d'invasion (glideosome) de la cellule hôte par le parasite. Jusqu'à présent, la présence et le rôle des glycosylations chez les Apicomplexas étaient assez controversés [Handman et al. 1980, Odenthal-Schnittler et al. 1993, Gaskins et al. 2004].

Ainsi, ce travail a consisté en une étude globale ayant pour objectif d'une part d'identifier les fonctions des glycoprotéines impliquées dans le glideosome en utilisant une stratégie protéomique adaptée, et d'autre part de caractériser leurs glycosylations. Dans ce but, nos collaborateurs ont étudié les glycanes libérés obtenus après déglycosylation des glycoprotéines purifiées, et en parallèle nous avons caractérisé leur micro-hétérogénéité en analysant les glycopeptides intacts obtenus après digestion enzymatique des glycoprotéines. Ce travail a donné lieu à une publication présentée en Annexe 1.

### **2.2. Méthodologie**

#### **2.2.1. Préparation d'échantillon et analyse protéomique**

La préparation d'échantillon et l'ensemble des expériences effectuées par nos collaborateurs (culture et isolation du parasite, chromatographie d'affinité, microscopie confocale, immunofluorescence, traitement à la tunicamycine, western-blot, etc), sont détaillés dans la partie matériel et méthode de la publication en Annexe 1.

Pour les études de protéomique, nos collaborateurs ont purifié les glycoprotéines par chromatographie d'affinité en utilisant de l'agarose greffé avec de la concanavaline A. Les protéines éluées ont ensuite été déposées sur deux pistes de gel d'électrophorèse monodimensionnel SDS-PAGE en quantités équivalentes pour être séparées. Le gel a été révélé avec une coloration au bleu de Coomassie. Les bandes des deux pistes ont ensuite été découpées. Les protéines ont été réduites et alkylées puis digérées à la trypsine. Les peptides extraits des bandes ont été analysés par nanoLC-MS/MS avec un couplage nanoLC-Chip/HCT Ultra.

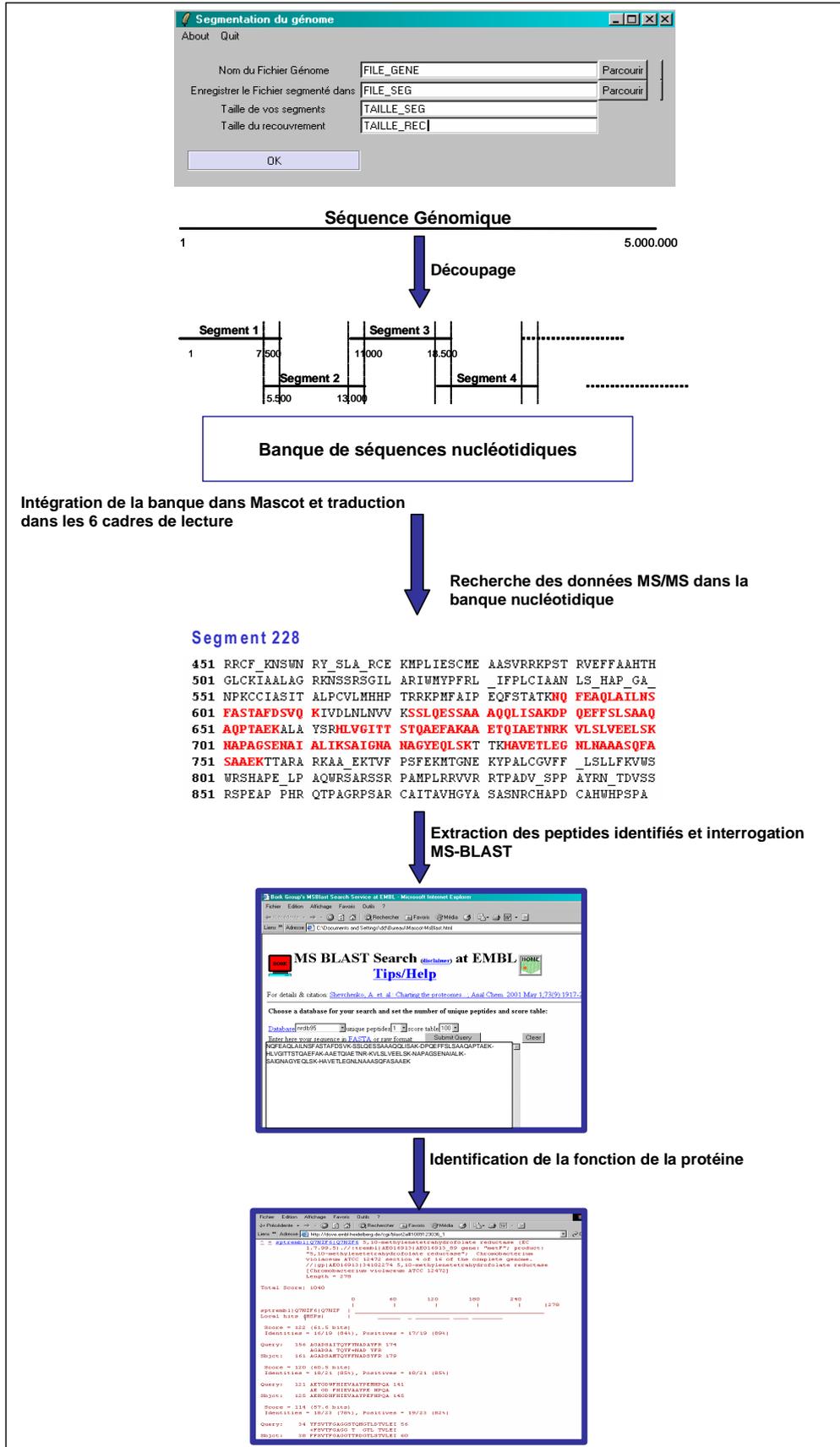
### **2.2.2. Identification des glycoprotéines par analyse nanoLC-MS/MS et recherche dans le génome complet non annoté de *T. gondii***

Les bandes de la première piste ont été analysées en nanoLC-MS/MS. Les données ont ensuite été interprétées en recherchant dans la banque de données NCBIInr à l'aide du moteur de recherche Mascot.

Lorsque ce projet a débuté, le génome de *T. gondii* bien qu'entièrement séquencé n'était pas encore totalement annoté. Nous avons alors choisi d'adopter une stratégie originale d'identification de protéines en recherchant directement dans le génome complet non annoté de *T. gondii* en complément de la recherche classique dans la banque de données NCBIInr. Cette méthode a été décrite dans une thèse effectuée au laboratoire [Carapito 2006] et est résumée dans la Figure 3.

Le génome complet de *T.gondii* a été téléchargé sur le site ToxoDB (<http://toxodb.org/toxo/>) [Gajria et al. 2007]. Il a ensuite été découpé en segments de 7500 paires de bases (bp) avec un recouvrement de séquence de 2500 bp pour chaque segment. Une fois le génome formaté, la banque de segments nucléotidiques a été importée dans le programme Mascot et traduit automatiquement dans les 6 cadres de lecture. Les requêtes Mascot ont alors été lancées dans cette banque génomique en utilisant les critères habituels : la trypsine spécifiée comme enzyme de coupure, un clivage manqué toléré, 0,25 Da d'erreur toléré en mode MS et MS/MS et la carbamidomethylation des cystéines et l'oxydation des méthionines définies comme modifications variables. Une fois la séquence de la protéine déterminée, sa fonction a été identifiée en utilisant le programme MS-BLAST qui permet d'effectuer des recherches par homologie de séquence.

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 dans la partie matériel et méthode de la publication en annexe 1.



**Figure 3.** Les principales étapes de la stratégie de recherche de données MS/MS dans les génomes complets développée au laboratoire. D'après [Carapito 2006, Gallien 2009]

### 2.2.3. Caractérisation de protéines N-glycosylées

Après réduction/alkylation et digestion trypsique, les peptides extraits de la 2<sup>ème</sup> piste ont été analysés par nanoLC-MS/MS afin de caractériser les glycopeptides intacts. Une colonne carbone graphite a été utilisée. Ce type de phase stationnaire permet de mieux retenir les glycopeptides qui sont plus hydrophiles que les peptides non glycosylés [Wagner-Rousset et al. 2008]. Les spectres MS/MS contenant les ions diagnostiques 366,14 et 528,19 ont été interprétés manuellement et permettent d'identifier les glycopeptides. Ce travail a déjà été décrit dans une thèse effectuée au laboratoire [Bednarczyk 2008].

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 dans la partie matériel et méthode de la publication en annexe 1.

## 2.3. Résultats publiés

Cette étude a permis de mettre en évidence plusieurs protéines qui contribuent à la prolifération et à la survie du parasite *T. gondii*, durant sa phase infectieuse. Parmi elles, 15 sont connues comme étant glycosylées. Certaines de ces glycoprotéines font directement partie du mécanisme macromoléculaire, appelé glideosome, comme la Myosine A, la Myosine Light Chain, l'Actine, IMC2A et GAP 50 [Gaskins et al. 2004, Keeley et Soldati 2004, Daher et Soldati-Favre 2009]. D'autres de ces protéines sont sécrétées par les organites du parasite comme les Rhoptries par exemples. Ces protéines sont impliquées dans l'adhésion du parasite à la cellule hôte et à la formation de la vacuole parasitaire [Alexander et al. 2005, Zhou et al. 2005].

Après déglycosylation à la PNGase F des extraits de glycoprotéines nos collaborateurs ont identifié plusieurs glycanes libérés correspondant à des glycosylations de type high-mannose et pauci-mannoside ( $\text{Man}_3\text{GlcNAc}_2$ ,  $\text{Man}_4\text{GlcNAc}_2$ ,  $\text{Man}_5\text{GlcNAc}_2$ ,  $\text{Man}_6\text{GlcNAc}_2$ ,  $\text{Man}_7\text{GlcNAc}_2$ ,  $\text{Man}_8\text{GlcNAc}_2$ ,  $\text{Man}_9\text{GlcNAc}_2$ ). La distribution de ces glycanes montre que les trois glycanes majoritaires sont  $\text{Man}_6\text{GlcNAc}_2$ ,  $\text{Man}_7\text{GlcNAc}_2$  et  $\text{Man}_8\text{GlcNAc}_2$ . Cependant, ces résultats ne permettent pas de déterminer les sites d'attachement de ces glycanes, ni la micro-hétérogénéité au niveau de chaque site de glycosylation.

C'est pourquoi, nous avons réalisé des analyses nanoLC-MS/MS des digestats des glycoprotéines afin de caractériser les glycopeptides intacts. Ces expériences ont permis d'identifier un des trois sites de glycosylation de la protéine GAP50 :  $\text{Asn}^{136}$ . Deux glycoformes ont été caractérisées au niveau de ce site :  $\text{Man}_7\text{GlcNAc}_2$  et  $\text{Man}_8\text{GlcNAc}_2$ .

## 2.4. Conclusion et perspectives

L'ensemble de ce travail a permis de démontrer que des glycoprotéines sont impliquées dans la reconnaissance et l'invasion de la cellule hôte par le parasite. Les analyses protéomiques par nanoLC-MS/MS ont permis d'identifier 15 glycoprotéines.

L'ensemble des protéines identifiées est impliqué dans la survie du parasite et dans son mécanisme d'infection ce qui en fait des cibles thérapeutiques potentielles. Cependant avant d'arriver à cette étape, les protéines doivent être caractérisées afin de déterminer leur rôle dans les mécanismes biologiques impliqués.

Seul un site de glycosylation (Asn<sup>136</sup>) et deux glycoformes (Man<sub>7</sub>GlcNAc<sub>2</sub> et Man<sub>8</sub>GlcNAc<sub>2</sub>) ont pu être caractérisés pour la protéine GAP50 sur l'ensemble des glycoprotéines identifiées alors que l'étude des glycanes libérés après déglycosylation a mis en évidence plusieurs structures oligosaccharidiques (Man<sub>3-9</sub>GlcNAc<sub>2</sub>). En effet, l'analyse de glycopeptides intacts a été réalisée sur une très faible quantité de matériel, qui n'a pas permis d'optimiser une méthode pour les caractériser en détail.

Ainsi, suite à cette première étude protéomique, nous avons ciblé le projet sur la caractérisation d'une protéine particulièrement importante dans le complexe macromoléculaire qu'est le glideosome : GAP50. Pour cela, nous avons appliqué la méthodologie développée au cours de ce travail de thèse et qui est décrite dans le Chapitre 2. Les résultats obtenus sont détaillés dans la partie suivante (§ 3).



# **ANNEXE 1: 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\*<sup>§</sup>

Sylvain Fauquenoy‡, Willy Morelle‡, Agnès Hovasse§, Audrey Bednarczyk§, Christian Slomianny¶, Christine Schaeffer§, Alain Van Dorselaer§, and Stanislas Tomavo‡||

The apicomplexan parasite *Toxoplasma gondii* recognizes, binds, and penetrates virtually any kind of mammalian cell using a repertoire of proteins released from late secretory organelles and a unique form of gliding motility (also named glideosome) that critically depends on actin filaments and myosin. How *T. gondii* glycosylated proteins mediate host-parasite interactions remains elusive. To date, only limited evidence is available concerning *N*-glycosylation in apicomplexans. Here we report comprehensive proteomics and glycomics analyses showing that several key components required for host cell-*T. gondii* interactions are *N*-glycosylated. Detailed structural characterization confirmed that *N*-glycans from *T. gondii* total protein extracts consist of oligomannosidic (Man<sub>5-8</sub>-GlcNAc)<sub>2</sub> and paucimannosidic (Man<sub>3-4</sub>(GlcNAc)<sub>2</sub>) sugars, which are rarely present on mature eukaryotic glycoproteins. *In situ* fluorescence using concanavalin A and *Pisum sativum* agglutinin predominantly stained the entire parasite body. Visualization of *Toxoplasma* glycoproteins purified by affinity chromatography followed by detailed proteomics and glycan analyses identified components involved in gliding motility, moving junction, and other additional functions implicated in intracellular development. Importantly tunicamycin-treated parasites were considerably reduced in motility, host cell invasion, and growth. Collectively these results indicate that *N*-glycosylation probably participates in modifying key proteins that are essential for host cell invasion by *T. gondii*. *Molecular & Cellular Proteomics* 7:891–910, 2008.

Apicomplexans are protozoan parasites defined by the presence of a complex of specialized organelles at their apical end (1). The phylum Apicomplexa contains many important

pathogens of humans and animals, including the causative agents of malaria (*Plasmodium* spp.) and chicken coccidiosis (*Eimeria* spp.) as well as some of the opportunistic infections associated with AIDS patients such as *Cryptosporidium parvum* and *Toxoplasma gondii*. *T. gondii* is uniquely adapted to infect a wide range of hosts, including virtually all warm blooded animals and up to 50% of the world's human population. The primary transmission route in humans is via ingestion of undercooked contaminated meat, particularly lamb, and contact with feces from infected domestic cats. The parasite causes a serious opportunistic disease (toxoplasmosis) in congenitally infected infants that can lead to severe syndromes including congenital malformations such as blindness, mental retardation, and hydrocephaly in children exposed *in utero*. More attention has been paid to *T. gondii* because toxoplasmosis is a common opportunistic infection in immunocompromised individuals with AIDS and in transplant patients (2, 3). Although the parasite has a sexual cycle occurring in cats, in mammalian non-feline hosts, *T. gondii* is found in two haploid asexual forms, the rapidly replicating tachyzoites and the slowly dividing, quiescent encysted bradyzoites. Infection during the acute phase is rapidly established in the host by the fast replicating tachyzoites, which can invade a broad range of cell types. Unlike other protozoan pathogens that hijack the pre-existing host cell uptake machinery, *T. gondii* actively gains entry into host cells. Penetration is of paramount importance for survival and relevant to its pathogenesis during acute infection. In response to immune defense, tachyzoites differentiate into encysted bradyzoites that remain in the brain and other organs throughout the life of chronically infected hosts. The reactivation of encysted bradyzoites into actively replicating and cytolytic tachyzoites is the cause of fatal toxoplasmic encephalitis in AIDS patients (4). Like all apicomplexan parasites, *T. gondii* lacks cilia or flagella throughout most of its life cycles and enters into host cells using an unusual form of gliding motility (5). The process of host cell invasion also requires contact with the host cell plasma membrane followed by reorientation and generation of a motive force ensured by the actomyosin motor, which drives gliding and penetration into a novel, parasite-induced structure called the parasitophorous vacuole (6–9). Success-

From the ‡Unité de Glycobiologie Structurale et Fonctionnelle, CNRS UMR 8576, Université des Sciences et Technologies de Lille, 59655 Villeneuve d'Ascq, France, §Laboratoire de Spectrométrie de Masse Bioorganique, CNRS UMR 7178, Université Louis Pasteur, 67087 Strasbourg, France, and ¶Laboratoire de Physiologie Cellulaire, INSERM U 800, Université des Sciences et Technologies de Lille, 59655 Villeneuve d'Ascq, France

Received, August 17, 2007, and in revised form, January 9, 2008  
Published, MCP Papers in Press, January 9, 2008, DOI 10.1074/mcp.M700391-MCP200

ful invasion of host cells requires a highly regulated release of proteins from several parasite organelles, namely micronemes, rhoptries, and dense granules (Ref. 7 and see details in Fig. 9C). Microneme discharge occurs first. Their contents are involved in the attachment to the host cell membrane and the formation of a connection with the parasite's actomyosin system, thereby providing the motive force that drives motility and invasion (8, 9). We speculate that glycosylation of parasite proteins may be important for host cell recognition and invasion by *T. gondii*.

Protein *N*-glycosylation serves a wide variety of functions including signaling through interaction with lectins, protein stabilization, protease resistance, endosome sorting, protein folding, and secretion (10). In eukaryotes, the precursor for *N*-glycans is built up on the lipid carrier dolichol (Dol)<sup>1</sup> located in the endoplasmic reticulum (ER) membrane. Current evidence has established that the biosynthetic pathway is defined by the synthesis of the protein *N*-glycosylation precursor oligosaccharide Glc<sub>3</sub>Man<sub>9</sub>(GlcNAc)<sub>2</sub>-PP-Dol. In most eukaryotes, this biosynthetic pathway of Glc<sub>3</sub>Man<sub>9</sub>(GlcNAc)<sub>2</sub>-PP-Dol is highly conserved, and the glycan portion is transferred en bloc via the action of oligosaccharyltransferase enzymes to Asn residues within Asn-X-(Ser/Thr) sequons during protein translation and sequestration into the lumen of the ER (10, 11). However, it has been shown that other eukaryotes can transfer structures other than the largest lipid-linked oligosaccharide precursor, Dol-PP-Glc<sub>3</sub>Man<sub>9</sub>(GlcNAc)<sub>2</sub> (12–14). Processing of the precursor structure by glycosidase and glycosyltransferase enzymes within the ER and Golgi apparatus generates the final set of mature glycan structures.

Although studies have clearly highlighted *N*-glycosylation as an essential modification for protein folding, stability, half-life, secretion, and other relevant biological functions in most eukaryotic cells, very little is known concerning this issue in *T. gondii*. The presence of *N*-glycoproteins is indeed still controversial (15–19). In general, *N*-glycosylation is considered a rare post-translational modification in apicomplexan parasites. For example, the causative agent of malaria, *Plasmodium falciparum*, lacks most genes involved in lipid (dolichylpyrophosphate)-linked oligosaccharide donors (20). We applied proteomics and glycomics analyses as well as pharmacological approaches to elucidate the parasite's *N*-glyco-

sylation biosynthetic pathway. This led to the elucidation of the major structures of protein *N*-linked glycans that suggests the presence of an almost complete early *N*-glycosylation biosynthetic pathway in *T. gondii*. In particular, we showed that several key proteins involved in gliding motility, moving junction, and other functions that are essential for host cell invasion by *T. gondii* are modified by *N*-glycosylation. Consistently inhibition of this post-translational modification dramatically impaired the parasite's motility, host cell invasion, and intracellular growth.

#### EXPERIMENTAL PROCEDURES

**Growth and Isolation of Parasites**—Human foreskin fibroblasts (HFFs) were maintained in Dulbecco's modified Eagle's medium (DMEM, BioWhittaker, Verviers, Belgium) supplemented with 10% fetal calf serum, 2 mM glutamine (Sigma), and 0.05% gentamycin (Sigma). Tachyzoites from RH or 76K strains were grown in monolayers of HFF cells, harvested, and purified as described previously (21). Encysted bradyzoites were purified from brains of mice chronically infected by *T. gondii* 76K for 2 months (22). These *in vivo* bradyzoites were freed by pepsin digestion (0.05 mg·ml<sup>-1</sup> pepsin in 170 mM NaCl, 60 mM HCl) for 5–10 min at 37 °C. For some experiments, we generated transgenic lines of *T. gondii* 76K expressing  $\beta$ -galactosidase using the plasmid containing the promoter 5'GRA1 and untranslated 3'SAG1 that drives  $\beta$ -galactosidase expression as described previously (23). This plasmid also contains TUB1-CAT-3'SAG1, which allowed the expression of chloramphenicol acetyltransferase for selection of *T. gondii* stable transgenic  $\beta$ -galactosidase-expressing clones. One of the  $\beta$ -galactosidase-expressing clones of *T. gondii* was used to quantify parasite numbers during host cell infection and growth.

**Detergent Extraction of Toxoplasma Proteins, Reduction, and Carboxyamidomethylation**—Proteins were extracted on ice in a buffer of 0.5% (w/v) SDS in 100 mM Tris-HCl (pH 7.4). Insoluble material was removed by centrifugation at 3000 rpm for 15 min. Detergent was removed by extensive dialysis against 50 mM ammonium bicarbonate at 4 °C. The protein extracts were dissolved in 500  $\mu$ l of 600 mM Tris-HCl (pH 8.3) and denatured by guanidine hydrochloride (6 M final concentration). The sample was incubated at 45 °C for 90 min, and the number of disulfide bridges was reduced using 300  $\mu$ g of DTT. The sample was flushed with argon and incubated at 45 °C for 5 h. After addition of 1.8 mg of iodoacetamide, the sample was flushed with argon and incubated at room temperature overnight in the dark. The sample was then extensively dialyzed against 50 mM ammonium bicarbonate at 4 °C and lyophilized.

**Tryptic, PNGase F, PNGase A, or  $\alpha$ -Mannosidase Digestions**—The reduced carboxymethylated proteins were digested with L-1-tosylamide-2-phenylethylchloromethylketone bovine pancreas trypsin (EC 3.4.21.4; Sigma) with an enzyme-to-substrate ratio of 1:20 (by mass), and the mixture was incubated for 24 h at 37 °C in 50 mM ammonium bicarbonate. The reaction was terminated by boiling for 5 min before lyophilization.

PNGase F (peptide-N<sup>4</sup>-(*N*-acetyl- $\beta$ -glucosaminyl)asparagine amidase, EC 3.5.1.52; Roche Applied Science) digestion was carried out in ammonium bicarbonate buffer for 16 h at 37 °C. The reaction was terminated by lyophilization, and the products were purified on a Sep-Pak C<sub>18</sub> (Waters Ltd.) to separate the *N*-glycans from the peptides. After conditioning the Sep-Pak C<sub>18</sub> by sequential washing with methanol (5 ml) and 5% acetic acid (2  $\times$  5 ml), the sample was loaded onto the Sep-Pak, and the glycans were eluted with 3 ml of 5% acetic acid. Peptides were eluted with 3 ml of 80% acetonitrile containing 5% acetic acid. Acetonitrile was evaporated under a stream of nitrogen, and the samples were freeze-dried. Glycopeptides remaining

<sup>1</sup> The abbreviations used are: Dol, dolichol; ER, endoplasmic reticulum; PP, pyrophosphate; HFF, human foreskin fibroblast; DMEM, Dulbecco's modified Eagle's medium; ConA, concanavalin A; PSA, *P. sativum* agglutinin; GC, gas chromatography; SAG, surface antigen; MIC, micronemal protein; GRA, dense granule protein; ROP, rhoptry protein; PNGase, peptide-N<sup>4</sup>-(*N*-acetyl- $\beta$ -glucosaminyl)asparagine amidase; PMAA, partially methylated alditol acetate; BLAST, Basic Local Alignment Search Tool; PFA, paraformaldehyde; mAb, monoclonal antibody; Hex, hexose; HexNAc, *N*-acetylhexosamine; Tg, *T. gondii*; MyoA, myosin A; MLC, myosin light chain; GAP, gliding-associated protein; RON, rhoptry neck protein; AMA, apical membrane antigen.

after PNGase F digestion were further digested with PNGase A (peptide-*N*<sup>4</sup>-(*N*-acetyl- $\beta$ -glucosaminyl)asparagine amidase, EC 3.5.1.52; Roche Applied Science) in ammonium acetate buffer (50 mM, pH 5.0) for 16 h at 37 °C using 0.5 milliunit of the enzyme. The reaction was terminated by lyophilization, and the products were purified on a Sep-Pak C<sub>18</sub> (Waters Ltd.) as described above.

$\alpha$ -Mannosidase digestion was carried out on PNGase F-released glycans using the following conditions: 0.5 unit of enzyme (jack bean  $\alpha$ -mannosidase, EC 3.2.1.24; Sigma) in 200  $\mu$ l of 50 mM ammonium acetate buffer, pH 4.5. The enzyme digestion was incubated at 37 °C for 48 h with a fresh aliquot of enzyme added after 24 h and terminated by boiling for 10 min before lyophilization. An aliquot corresponding to 25% of the total sample was taken after the digestion and permethylated for MALDI-TOF-MS and gas chromatography (GC)-MS analysis.

**Permethylation of Glycans**—Permethylation of the freeze-dried glycans was performed according to the procedure developed by Ciucanu and Kerek (24). The reaction was terminated by adding 1 ml of ultrapure water followed by three extractions with 500  $\mu$ l of chloroform. The pooled chloroform phases (1.5 ml) were then washed eight times with ultrapure water. The methylated derivative-containing chloroform phase was finally dried under a stream of nitrogen, and the extracted products were further purified on a Sep-Pak C<sub>18</sub>. The Sep-Pak C<sub>18</sub> was sequentially conditioned with methanol (5 ml) and water (2  $\times$  5 ml). The derivatized glycans dissolved in methanol were applied on the cartridge, washed with 3  $\times$  5 ml of water and 2 ml of 10% (v/v) acetonitrile in water, and eluted with 3 ml of 80% (v/v) acetonitrile in water. Acetonitrile was evaporated under a stream of nitrogen, and the samples were freeze-dried.

**MALDI-TOF-MS**—MALDI-TOF-MS experiments were carried out on a Voyager Elite DE-STR Pro instrument (PerSeptive Biosystems, Framingham, MA) equipped with a pulsed nitrogen laser (337 nm) and a gridless delayed extraction ion source. The spectrometer was operated in the positive reflectron mode by delayed extraction with an accelerating voltage of 20 kV, a pulse delay time of 200 ns, and a grid voltage of 66%. All spectra shown represent accumulated spectra obtained by 400–500 laser shots. Permethylated glycans were co-crystallized with 2,5-dihydroxybenzoic acid as matrix (10 mg/ml 2,5-dihydroxybenzoic acid in methanol/water solution (50:50)). The instrument was externally calibrated with a maltose ladder encompassing the *m/z* values of the analyzed samples.

**Linkage Analysis**—The permethylated native *N*-glycans were hydrolyzed in 300  $\mu$ l of 4 M TFA at 100 °C for 4 h. After removing TFA by drying *in vacuo*, the permethylated compounds were then reduced at room temperature overnight by adding 200  $\mu$ l of 2 M ammonia solution containing sodium borodeuteride (4 mg/ml). The reduction was terminated by adding acetic acid, and borates were eliminated under a stream of nitrogen in the presence of methanol containing 5% (v/v) acetic acid. After adding 20  $\mu$ l of pyridine and 200  $\mu$ l of acetic anhydride, peracetylation was carried out at 100 °C for 2 h. After evaporation under a stream of nitrogen, the partially methylated alditol acetates (PMAAs) were dissolved in chloroform, and the chloroform phase was washed 10 times with water. This PMAA-containing phase was finally dried under a stream of nitrogen, and the PMAAs were dissolved in methanol before GC-MS analysis. GC separation of PMAAs was performed using a Carbo Erba GC 8000 gas chromatograph fitted with a 25-m  $\times$  0.32-mm CP-Sil5 CB low bleed capillary column, 0.25- $\mu$ m film phase (Chrompack France, Les Ullis, France). The temperature of the Ross injector was 260 °C. Samples were analyzed using a temperature program starting by a gradient of 2 °C/min from 130 to 180 °C after 2 min at 130 °C followed by a gradient of 4 °C/min until 240 °C. The column was coupled to a Finnigan Automass II mass spectrometer. PMAA analyses were performed in the electron impact mode using an ionization energy of 70 eV. Quan-

tification of the various PMAA derivatives was carried out using total ion current of the MS detector in the positive ion mode.

**Affinity Chromatography Using Concanavalin A-linked Agarose**—All purification steps were carried out at 4 °C. Purified tachyzoites of RH strain (10<sup>9</sup> parasites) were lysed with 1 ml of buffer A containing 10 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 mM CaCl<sub>2</sub>, 1 mM MnCl<sub>2</sub>, 1% Triton X-100, and protease inhibitor mixture (Roche Diagnostics). After 1 h at 4 °C, insoluble material was removed by centrifugation at 10,000  $\times$  *g* for 15 min. The supernatant was incubated with concanavalin A (ConA)-agarose beads (from *Canavalin ensiformis*, jack bean, Type VA, Sigma) for 3 h and washed five times with buffer A without protease inhibitors. After a final wash with 62.5 mM Tris-HCl (pH 6.8), the bound proteins were eluted with 2 volumes of 2 $\times$  SDS sample buffer and boiled at 95 °C for 5 min or with 0.5 M  $\alpha$ -methyl-D-mannoside. Competition assays were also performed using ConA beads incubated with protein extract in the presence of 0.5 M  $\alpha$ -methyl-D-mannoside. Large scale purification of glycoproteins for proteomics analyses was also performed according to the protocol described above except that protein extracts from 10<sup>10</sup> tachyzoites were used. The eluted samples were analyzed by SDS-PAGE followed by Coomassie Blue staining.

**Protein Identification by Mass Spectrometry**—In-gel digestion was performed with an automated protein digestion system, MassPREP station (Waters, Milford, MA). The gel plugs were washed twice with 50  $\mu$ l of 25 mM NH<sub>4</sub>HCO<sub>3</sub> and 50  $\mu$ l of ACN. The cysteine residues were reduced by 50  $\mu$ l of 10 mM DTT at 57 °C and alkylated by 50  $\mu$ l of 55 mM iodoacetamide at room temperature. After dehydration of the gel bands with ACN, the proteins were digested overnight in gel with 50  $\mu$ l of 12.5 ng/ $\mu$ l modified pig trypsin (Promega, Madison, WI) in 25 mM NH<sub>4</sub>HCO<sub>3</sub> at room temperature. The peptides released were extracted twice with 60% ACN in 5% formic acid and 100% ACN followed by excess ACN removal. Tryptic digests were analyzed by nano-LC-MS/MS using an Agilent 1100 series HPLC-Chip/MS system with a 43-mm reversed phase C<sub>18</sub> column (Agilent Technologies, Palo Alto, CA) coupled to an HCTUltra ion trap (Bruker Daltonics, Bremen, Germany). The voltage applied to the capillary cap was optimized to -1750 V. For tandem MS experiments, the system was operated with automatic switching between MS and MS/MS modes. The three most abundant peptides and preferentially doubly charged ions were selected on each MS spectrum for further isolation and fragmentation. The MS/MS scanning was performed in the ultrascan resolution mode at a scan rate of 26,000 *m/z* per second. A total of six scans were averaged to obtain an MS/MS spectrum. The complete system was fully controlled by ChemStation (Agilent Technologies) and EsquireControl (Bruker Daltonics) softwares.

Mass data collected during nano-LC-MS/MS analysis were processed, converted into \*.mgf files, and searched first against the National Center for Biotechnology Information non-redundant (NCBI) database using a local Mascot™ server (Matrix Science, London, UK). In addition, an in-house genome database was constructed using the complete genome sequence, downloaded as a single text file from The Institute for Genomic Research, and segmented into regular segments of well chosen length (7500 bp) with a determined overlap length (2500 bp). This nucleic acid database was imported into the local Mascot server, and the segments were translated in all possible reading frames.<sup>2</sup> Once expressed coding regions were identified thanks to the matching peptides on the genome sequence, the protein functions were determined by MS-BLAST sequence similarity searches. Searches were also performed using the annotated genome sequence available at ToxoDB. In both cases,

<sup>2</sup> S. Fauquenoy, W. Morelle, A. Hovasse, A. Bednarczyk, C. Slomianny, C. Schaeffer, A. Van Dorsseleer, and S. Tomavo, unpublished data.

protein identification was validated according to the established guidelines for proteomics data publication (25–27).

**N-Glycosylated Protein Identification by Mass Spectrometry**—Tryptic digests were analyzed by nano-LC-MS/MS using an Agilent 1100 series HPLC-Chip/MS system with a 43-mm carbon graphite column (Agilent Technologies) coupled to an HCTplus ion trap (Bruker Daltonics). The voltage applied to the capillary cap was optimized to  $-1750$  V. The mass spectrometer ion transmission parameters were optimized for high  $m/z$  (1000  $m/z$ ). For tandem MS experiments, the system was operated with automatic switching between MS and MS/MS modes. The three most abundant peptides and preferentially doubly charged ions were selected on each MS spectrum for further isolation and fragmentation. The MS/MS scanning was performed in the ultrascan resolution mode at a scan rate of 26,000  $m/z$  per second. A total of six scans were averaged to obtain an MS/MS spectrum. The complete system was fully controlled by ChemStation (Agilent Technologies) and EsquireControl (Bruker Daltonics) softwares. For each protein identified, potential *N*-glycosylation sites were predicted using the NetNGlyc 1.0 Server, and *in silico* tryptic digestion was performed to generate the tryptic peptides containing a potential *N*-glycosylation site in their amino acid sequence. For each analysis, typical glycopeptide diagnostic ions were extracted from MS/MS ion current, and MS/MS spectra containing diagnostic ions were manually interpreted. Results were validated taking into account the following three types of fragment ions classically generated for glycopeptides: sugar neutral losses, non-glycosylated peptide fragments, and glycan backbone internal fragments.

**Western Blotting**—Glycoproteins, affinity-purified using a ConA column, resuspended in  $2\times$  SDS sample buffer and freshly isolated tachyzoites lysed with the same sample buffer were separated on SDS-polyacrylamide gels under non-reducing conditions and blotted to nitrocellulose by an electrophoretic transfer system (Bio-Rad system). Blots were blocked in TNT buffer (15 mM Tris-HCl (pH 8), 140 mM NaCl, 0.05% Tween 20) with 5% dry milk and incubated with primary antibodies (monoclonal or rabbit polyclonal antibodies) followed by goat anti-rabbit IgG or goat anti-mouse IgG conjugated to peroxidase (Sigma), and the signal was detected by chemiluminescence using the ECL kit (Amersham Biosciences). *T. gondii*-specific antibodies were generously provided by Drs. M. Lebrun (anti-rhoptry neck protein 2 (RON2) and anti-apical membrane antigen 1 (AMA1), D. Boothroyd (anti-AMA1), P. Bradley (anti-RON1 and anti-apicoplast), D. Soldati (anti-myosin A (MyoA), anti-MLC1, and anti-gliding-associated protein 45 (GAP45) and J.-F. Dubremetz (anti-ROP2/3/4, anti-MIC, and anti-GRA).

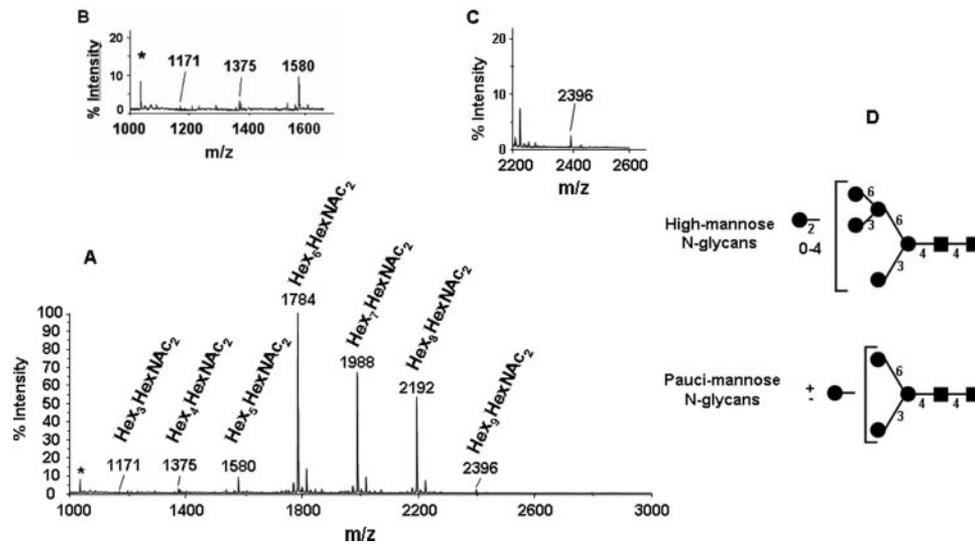
**Lectin and Immunofluorescence Localizations**—Freshly isolated and purified tachyzoites were fixed with 4% paraformaldehyde (PFA) in PBS for 30 min on ice. For the localization of glycoproteins, fixed slides were permeabilized with 0.1% Triton X-100 prepared in buffer B containing 10 mM Tris-HCl (pH 7.5), 0.15 M NaCl, 1 mM  $\text{CaCl}_2$ , 1 mM  $\text{MnCl}_2$  at room temperature. Free PFA was blocked by 0.1 M glycine in buffer C (similar to buffer B except that the Triton X-100 was at 0.01%) for 10 min at room temperature. Then parasites were incubated either with 10  $\mu\text{g/ml}$  biotinylated concanavalin A (Type IV, Sigma) or 100  $\mu\text{g/ml}$  FITC-coupled *Pisum sativum* agglutinin (PSA; Sigma) diluted in buffer C for 30 min at 37 °C. After three washes with buffer C, the ConA-treated slides were incubated with FITC-coupled streptavidin (Sigma) for 30 min at 37 °C and washed three times. For co-localization, slides containing intracellular parasites were incubated with lectins followed by incubation with monoclonal or polyclonal antibodies in PBS with 0.01% Triton X-100. Fluorescently conjugated secondary antibodies were added and incubated for 30 min at 37 °C. After three washes, slides were mounted in ImmunO (immunofluor mounting medium, MP Biochemicals Inc.), examined, and photographed on a Zeiss confocal microscope.

**In Vitro Gliding Assays**—Intracellular tachyzoites of the RH strain were treated with 5  $\mu\text{g/ml}$  tunicamycin (Sigma) prepared at an initial concentration of 1 mg/ml in DMSO for 40 h at 37 °C. The untreated intracellular parasites incubated in DMSO only or tunicamycin-treated parasites were filtered as described above; chased with DMEM containing 10 mM HEPES, 1 mM EGTA; and centrifuged at 1,000 rpm for 10 min at room temperature. The parasite pellet was resuspended in medium as above, and  $5 \times 10^8$  tunicamycin-treated or untreated tachyzoites were inoculated into a chamber containing a slide that had previously been coated overnight with 50% fetal calf serum in PBS. The slides were then incubated with the parasites at 37 °C for 15 min, rinsed in PBS, and fixed in 4% PFA in PBS. Trails left by gliding parasites were visualized by staining with the monoclonal antibody anti-surface antigen (SAG) 1 and photographed using a Zeiss Axiophot microscope.

**Parasite Attachment, Host Cell Invasion, and Growth**—To test parasite attachment and invasion, two assays were used to distinguish parasites that had attached and entered host cells from those that had entered and replicated in host cells. First, freshly purified tachyzoites of RH strain treated with tunicamycin or untreated parasites as described above were resuspended in DMEM containing 10% FCS, loaded onto slides containing confluent monolayer HFF cells, incubated at 37 °C for 10 min, then fixed with 4% PFA, and processed for immunofluorescence analysis. Second, tunicamycin-treated or untreated tachyzoites of RH strain were also loaded onto confluent monolayer HFF cells for 1 h, and intracellular growth was monitored for 24 h at 37 °C. Slides were rinsed with PBS, fixed with 4% PFA, washed with PBS, and stained with mAb anti-SAG1 followed by FITC-goat anti-mouse IgG. Attached/invaded parasites for 10 min-pulse infection or 1-h infection followed by growth for 24 h were quantified by single tachyzoites or replicated tachyzoites, and counting was performed for at least 500 host cells. Experiments were repeated three times, and data are expressed as the mean  $\pm$  S.D. Third, parasite growth was alternatively quantified by using freshly isolated  $\beta$ -galactosidase-expressing 76K strain parasites to infect slides in 24-well plates containing HFF cells for 4 h at 37 °C. Four wells were individually treated with 5  $\mu\text{g/ml}$  tunicamycin or DMSO alone. After 2 days of growth, one-fifth of each well was used to infect a new well containing monolayer HFF cells, and incubation was continued for 2 days under tunicamycin or DMSO treatment. The remaining parasites of the first cycle experiment were recovered, washed with PBS, and frozen. After the second cycle of treatment, one-fifth of each well was used to infect new monolayer HFF cells for the third cycle experiment. The parasite pellets were washed with PBS and lysed with buffer containing 100 mM HEPES (pH 8.0), 1 mM  $\text{MgSO}_4$ , 5 mM DTT, 1% Triton X-100. The  $\beta$ -galactosidase activity was quantified using a chlorophenol red- $\beta$ -D-galactopyranoside (Roche Applied Science) colorimetric assay (23). Three independent experiments were performed, and data are expressed as mean  $\pm$  S.D.

## RESULTS

**MALDI-TOF-MS of N-Glycans Released from Detergent Extracts**—To gain more insight into protein *N*-glycosylation by the parasite, we embarked on the determination of the major structures of *N*-glycans synthesized by *T. gondii*. Experimental strategies based on derivatization, chemical hydrolysis, exoglycosidase digestions, MALDI-TOF-MS, and GC-MS were used to characterize the glycans of *Toxoplasma* total protein extracts. Reduced and carboxamidomethylated *Toxoplasma* products were first digested with trypsin, and glycans were released from the resulting peptide/glycopeptide mixture by digestion with PNGase F. This enzyme is capable of releasing all known *N*-linked oligosaccharides except those



**FIG. 1. Positive MALDI-TOF mass spectrum of permethylated *N*-glycans from *T. gondii* glycoproteins.** A, the *N*-glycans were released from tryptic glycopeptides by digestion with PNGase F, separated from peptides by Sep-Pak purification, and permethylated. The three major molecular ions identified are Hex<sub>6</sub>(HexNAc)<sub>2</sub> + Na<sup>+</sup> (*m/z* 1784), Hex<sub>7</sub>(HexNAc)<sub>2</sub> + Na<sup>+</sup> (*m/z* 1988), and Hex<sub>8</sub>(HexNAc)<sub>2</sub> + Na<sup>+</sup> (*m/z* 2192), which correspond to Man<sub>6</sub>(GlcNAc)<sub>2</sub>, Man<sub>7</sub>(GlcNAc)<sub>2</sub>, and Man<sub>8</sub>(GlcNAc)<sub>2</sub>, respectively. \*, contaminants from the MALDI-TOF DHB solution used to resuspend the solution. B and C, four minor molecular ions, Hex<sub>3</sub>(HexNAc)<sub>2</sub> + Na<sup>+</sup> (*m/z* 1171), Hex<sub>4</sub>(HexNAc)<sub>2</sub> + Na<sup>+</sup> (*m/z* 1375), Hex<sub>5</sub>(HexNAc)<sub>2</sub> + Na<sup>+</sup> (*m/z* 1580), and Hex<sub>9</sub>(HexNAc)<sub>2</sub> + Na<sup>+</sup> (*m/z* 2396) corresponding, respectively, to Man<sub>3</sub>(GlcNAc)<sub>2</sub>, Man<sub>4</sub>(GlcNAc)<sub>2</sub>, Man<sub>5</sub>(GlcNAc)<sub>2</sub>, and Man<sub>9</sub>(GlcNAc)<sub>2</sub> were also detected and are depicted in the insets B and C to view these minor peaks relative to the three major peaks. Ions that appear at *m/z* values slightly higher than those assigned to Hex<sub>5-8</sub>(HexNAc)<sub>2</sub> in the spectrum correspond to species 30 Da larger than the fully methylated carbohydrate molecules (51). D, our proposed structures of *N*-glycans isolated from total extract glycoproteins of *T. gondii*. These fine *N*-glycan structures were determined twice in two independent productions of purified *T. gondii* materials by combining several enzymatic digestions, MALDI-TOF-MS, and GC-MS analyses. ■, *N*-acetylglucosamine; ●, mannose.

with fucose attached to the 3-position of the Asn-linked GlcNAc residue. Such PNGase F-resistant oligosaccharides have been found to be sensitive to PNGase A. PNGase F-released glycans were separated from peptides and were analyzed by MALDI-TOF-MS after permethylation and purification on a Sep-Pak C<sub>18</sub> (Fig. 1). The permethylation derivatization of glycans increases the sensitivity of the detection of molecular ions. The peptide fraction was further digested with PNGase A, and the putative glycans released were analyzed by MALDI-TOF-MS after permethylation and purification on a Sep-Pak C<sub>18</sub>. Fig. 1A shows the MALDI-TOF mass spectrum profile of the permethylated PNGase F-released glycans from *Toxoplasma* detergent extracts. Permethylated glycans give [M + Na]<sup>+</sup> species as the major ion. Three major molecular ions that correspond to Hex<sub>6</sub>(HexNAc)<sub>2</sub> + Na<sup>+</sup> (*m/z* 1784), Hex<sub>7</sub>(HexNAc)<sub>2</sub> + Na<sup>+</sup> (*m/z* 1988), and Hex<sub>8</sub>(HexNAc)<sub>2</sub> + Na<sup>+</sup> (*m/z* 2192) were observed. Four minor molecular ions corresponding to Hex<sub>3</sub>(HexNAc)<sub>2</sub> + Na<sup>+</sup> (*m/z* 1171), Hex<sub>4</sub>(HexNAc)<sub>2</sub> + Na<sup>+</sup> (*m/z* 1375), Hex<sub>5</sub>(HexNAc)<sub>2</sub> + Na<sup>+</sup> (*m/z* 1580), and Hex<sub>9</sub>(HexNAc)<sub>2</sub> + Na<sup>+</sup> (*m/z* 2396) were also identified. Fig. 1, insets B and C, show enlargements of a spectrum that revealed these four minor species. Based on the MALDI-TOF-MS data and currently accepted models of *N*-glycan biosynthesis (10, 11), we conclude that the major *N*-glycans of *T. gondii* have compositions consistent with high mannose type structures (Hex<sub>5-8</sub>(HexNAc)<sub>2</sub>). Very minor *N*-glycans have compositions consistent with paucimannose type struc-

tures (Hex<sub>3-4</sub>(HexNAc)<sub>2</sub>). After PNGase F digestion, the peptide fraction was further digested with PNGase A. Putative PNGase A-released *N*-glycans were permethylated, purified on a Sep-Pak C<sub>18</sub> cartridge, and analyzed by MALDI-TOF-MS. No signals corresponding to *N*-glycans were observed (data not shown). Thus, the PNGase F digestion was complete, and we could find no evidence that *N*-linked glycans with fucose attached to the 3-position of the Asn-linked GlcNAc were an important component of *Toxoplasma*. In addition, also no modification of terminal galactose and sialic acid was identified, suggesting that *T. gondii* may lack a part of the endoplasmic reticulum and Golgi trimming and maturation pathways that are highly conserved in other eukaryotes (10, 11).

**Structural Analyses of *N*-Glycans Released by PNGase F**—Linkage analyses on the PNGase F-released glycans by GC-MS shown in Table I are fully consistent with high mannose and paucimannose type structures being the constituents of the *N*-glycan population as terminal mannose and 2-linked mannose are the major residues. To define the anomeric configurations and also to confirm the presence of high mannose type structures, *N*-glycans released by PNGase F were subjected to treatment with  $\alpha$ -mannosidase. An aliquot was taken after the digestion, permethylated, and examined by MALDI-TOF-MS after Sep-Pak C<sub>18</sub> purification. After  $\alpha$ -mannosidase treatment, the molecular ions at *m/z* 1171 (Hex<sub>3</sub>(HexNAc)<sub>2</sub>), 1375 (Hex<sub>4</sub>(HexNAc)<sub>2</sub>), 1580 (Hex<sub>5</sub>(HexNAc)<sub>2</sub>), 1784 (Hex<sub>6</sub>(HexNAc)<sub>2</sub>), 1988 (Hex<sub>7</sub>(HexNAc)<sub>2</sub>), 2192 (Hex<sub>8</sub>-

TABLE I  
GC-MS analysis of partially methylated alditol acetates obtained from the PNGase F-released N-glycans

Retention time <i>min</i>	Characteristic fragment ions	Assignment	Relative abundance
18.40	102, 118, 129, 145, 161, 162, 205	Terminal mannose	0.81
22.34	129, 130, 161, 190	2-Linked mannose	1.00
28.36	118, 129, 189, 234	3,6-Linked mannose	0.58
33.30	117, 159, 233	4-Linked GlcNAc	0.38

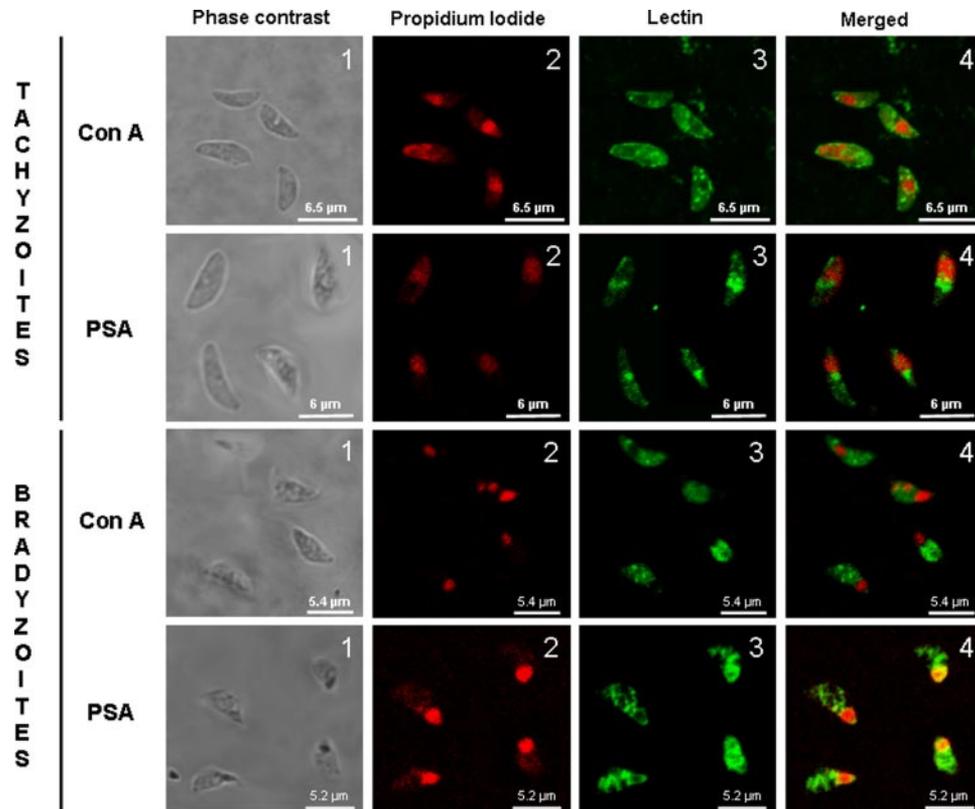
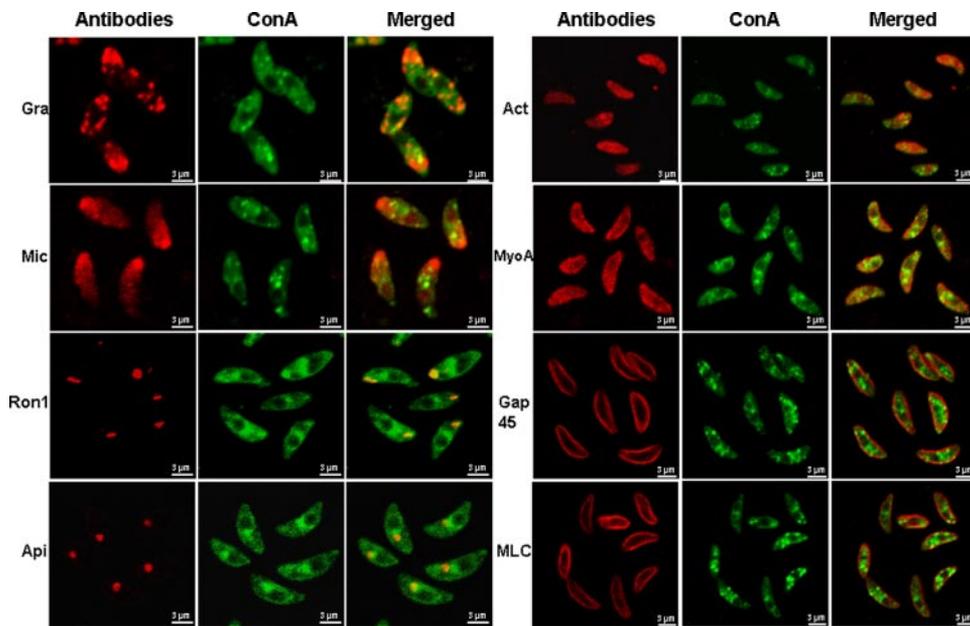


FIG. 2. Confocal microscopy analysis of lectin-staining formaldehyde-fixed *T. gondii*. Extracellular tachyzoites were purified from infected human foreskin fibroblasts, and bradyzoites were isolated from brains of mice chronically infected by *T. gondii*. ConA that is specific to  $\alpha$ -D-mannose and  $\alpha$ -D-glucose and PSA that exclusively recognizes  $\alpha$ -D-mannose were used. The ConA signal was amplified by the FITC-conjugated streptavidin, whereas PSA was directly conjugated to FITC. Panels 1, phase-contrast. Panels 2, propidium iodide staining of nucleus (red). Note that mRNAs are also labeled by this dye, and the yellow signals around the nucleus (panel 4, PSA) indicate co-localization between the lectin signal and the endoplasmic reticulum-containing mannose-rich glycoproteins and lipid precursors and the propidium iodide staining. Panels 3, lectins. Panels 4, merged signal between propidium iodide staining and lectin labels.

(HexNAc)<sub>2</sub>), and 2396 (Hex<sub>9</sub>(HexNAc)<sub>2</sub>) were abolished with a concurrent increase in signals consistent with the core of digested N-glycans Hex<sub>1</sub>(HexNAc)<sub>2</sub> (*m/z* 763) and Hex<sub>2</sub>- (HexNAc)<sub>2</sub> (*m/z* 967). Taken together, these detailed structural analyses indicated that the major *Toxoplasma* N-linked glycans are composed of high mannose (Man<sub>5-8</sub>(GlcNAc)<sub>2</sub>) and paucimannose (Man<sub>3-4</sub>(GlcNAc)<sub>2</sub>) type structures (Fig. 1D).

**Subcellular Localization of Toxoplasma N-Glycoproteins**—Indirect fluorescence staining of *T. gondii* was carried out to localize novel glycoproteins within extracellular parasites. Based on the structural data suggesting the apparent lack of complex type of N-glycans with sialic acid, fucose, and ga-

lactose residues in *T. gondii*, the two lectins ConA and PSA that recognize oligomannosidic glycans were tested. Although ConA is specific to both terminal  $\alpha$ -D-mannose and  $\alpha$ -D-glucose residues, PSA defines an exclusive terminal  $\alpha$ -D-mannose-binding lectin. We decided to test both lectins to discriminate the signal of a polymer glucose named amylopectin that is abundantly present in the cytoplasm of bradyzoites and to a lesser extent in tachyzoites of *T. gondii* (28). When stained with ConA, the rapidly replicating tachyzoite forms that cause the acute phase of the disease (toxoplasmosis) revealed a fluorescence signal covering the entire body of the tachyzoites with a few dots of cytoplasmic punctate

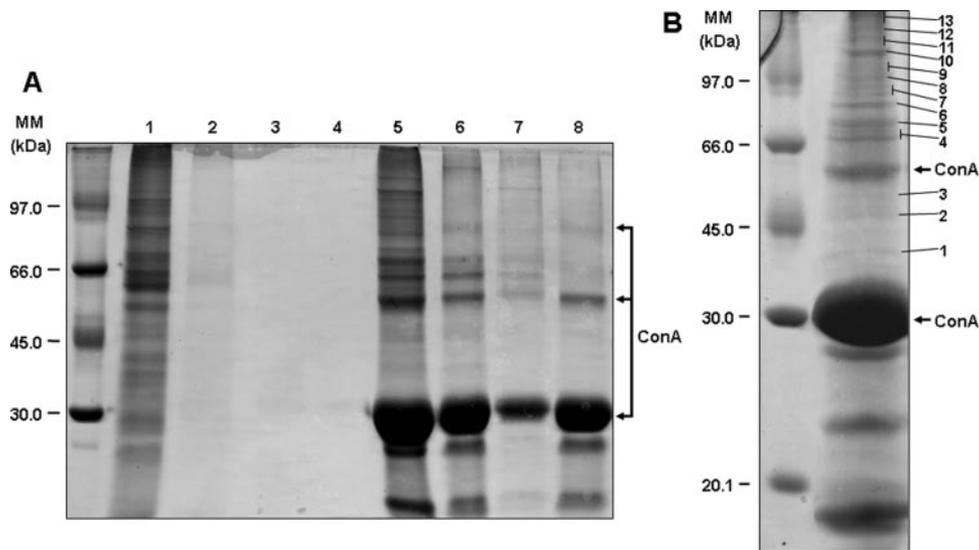


**FIG. 3. Colocalization of ConA fluorescence with different *T. gondii* organelle-specific markers.** Double fluorescence assays were performed using ConA revealed with FITC-conjugated streptavidin followed by incubation with monoclonal or polyclonal antibodies specific for protein markers of rhoptries (*Ron1*), microneme (*Mic1*), dense granules (*Gra1*), apicoplast (*Api*), actin (*Act*), MyoA, gliding-associated protein (*Gap45*), and myosin light chain 1 (*MLC*). The signals corresponding to monoclonal and polyclonal antibodies were revealed by a secondary goat Alexa Fluor 594-conjugated antibody (red). *Panel 1*, monoclonal or polyclonal antibody; *panel 2*, ConA; *panel 3*, merged signals between antibodies and ConA. The images were examined and photographed on a Zeiss confocal microscope.

fluorescence (Fig. 2, *panel 3*). No specific or obvious labeling of late secretory organelles could be distinctly observed (Fig. 2, *panel 3*). The lectin PSA also gave identical cytoplasmic patterns in tachyzoites (Fig. 2, *PSA, panel 3*) with the presence of pronounced staining of vesicles found close to the nucleus stained in red by propidium iodide (*panel 2*). We also investigated the presence of lectin-stained glycoproteins in the dormant encysted bradyzoite forms that are responsible for the chronic phase of the disease. The released bradyzoites obtained from 2-month-old cysts isolated from chronically infected mice also positively stained with both ConA and PSA, suggesting the presence of N-linked glycoproteins in the dormant form of *T. gondii* (Fig. 2, *Bradyzoites, Con A, panel 3*). The patterns of ConA and PSA showed clusters of scattered staining throughout the cytoplasm of bradyzoites (Fig. 2, *Bradyzoites, panels 3*), suggesting that these dormant parasite stages are still expressing N-linked glycans *de novo*. We cannot rule out that N-glycoproteins produced in tachyzoites or at the beginning of bradyzoite differentiation (in younger bradyzoites) are stably conserved throughout the life cycle of encysted bradyzoites. With the exception of lectin signals scattered throughout the whole body of the parasite and surrounding the nucleus (Fig. 2, *PSA, panel 4, yellow staining*) that suggest endoplasmic reticulum recognition, our experimental observations using a confocal microscope did not allow us to make a conclusion about the presence of N-glycoproteins in *T. gondii*-specific secretory organelles named rhoptries, micronemes, and dense granules that are known to be involved

in the host cell attachment, invasion, and intracellular growth of *T. gondii*.

**Colocalization of ConA Signal with Specific Markers of Subcellular Compartments of *T. gondii***—Indirect immunofluorescence and ConA staining of *T. gondii* tachyzoites were carried out to co-localize the lectin signal with that of several antibodies that are markers of different subcellular compartments. Fig. 3 shows diverse degrees of co-localization between ConA and markers of *T. gondii* late secretory organelles such as dense granules (mAb anti-*GRA1*, *Gra panel*) and micronemes (mAb anti-micronemal protein 1 (*MIC1*), *Mic panel*). Most importantly, a strong co-localization with *T. gondii* rhoptry neck protein 1 was revealed by a specific polyclonal antibody (*Ron1 panel*). Staining myosin A and actin with specific antibodies showed a pattern displaying weaker co-localization with ConA fluorescence (*Act and MyoA panels*) than that seen for *RON1* (*Ron1 panel*). In addition, no convincing co-localization was observed when polyclonal antibodies specific to gliding-associated protein 45 (*Gap45 panel*) and myosin light chain (*MLC panel*), two markers of the glideosome located between the inner complex membrane and the parasite's plasma membrane, were tested. Because it has been described that one major component of the glideosome, TgGAP50 is N-glycosylated (19), the lack of convincing co-localization of lectin fluorescence with glideosome proteins might be explained by the weakness of the lectin signal. Alternatively the N-glycans carried by these proteins may be involved in protein-protein and/or glycan-protein interactions



**FIG. 4. Purification of *T. gondii* N-linked glycoproteins using a ConA column.** *A*, the pilot experiment performed with detergent extract from  $10^9$  tachyzoites. *Lane 1*, total detergent extract used; *lanes 2–4*, efficacy of purity of glycoprotein-binding ConA column checked by taking three aliquots corresponding, respectively, to the first, third, and fifth/last washes; *lane 5*, eluate of ConA-binding glycoproteins using  $2\times$  SDS sample buffer; *lane 6*, competition assay performed by incubating the detergent protein extract in the presence of  $0.5\text{ M}$   $\alpha$ -methylmannoside inhibitor before elution as described in *lane 5*; *lane 7*, elution of ConA-binding glycoproteins with  $0.5\text{ M}$   $\alpha$ -methylmannoside; *lane 8*, elution of ConA beads alone by  $2\times$  SDS sample buffer. *B*, large scale purification of ConA-binding glycoproteins using detergent extract from  $10^{10}$  tachyzoites. After elution with  $2\times$  SDS sample buffer and SDS-PAGE, 13 bands were excised and processed for proteomics analyses. The contaminated proteins corresponding to different ConA species eluted from the beads are indicated by arrows. *MM*, molecular mass in kilodaltons.

that may render these structures inaccessible to lectin recognition. Interestingly an apparent co-localization was observed with the apicoplast, a plastid-like organelle found in apicomplexan parasites (depicted in Fig. 9C) using a monoclonal antibody specific to this organelle (Fig. 3, *Api panel*). As it has been described that proteins translocated into apicoplast are transported across the endoplasmic reticulum (29), we hypothesize that some apicoplast proteins containing potential glycosylation sequons may in fact be N-glycosylated. This issue will be further discussed.

**Nature of N-linked Glycoproteins Binding to ConA**—It was difficult to firmly establish which of the parasite's specific organelles and N-glycoproteins were labeled by fluorescence using these lectins. Affinity purification therefore was carried out for the isolation of N-linked proteins that bind to ConA. A pilot experiment was performed using a total detergent extract of  $10^9$  tachyzoites, and the ConA-binding glycoproteins eluted were checked by SDS-PAGE followed by Coomassie Blue staining (Fig. 4A). After five successive washes were tested for efficiency (Fig. 4A, *lanes 2–4*), N-linked glycosylated proteins bound to ConA-beads were eluted by  $2\times$  SDS sample buffer. Interestingly only glycoproteins ranging from 45 kDa to higher than 100 kDa were identified (Fig. 4, *lane 5*). The specificity of N-glycosylated proteins binding to ConA was demonstrated by a competition assay performed by incubating detergent protein extracts in the presence of  $0.5\text{ M}$   $\alpha$ -methylmannoside inhibitor (*lane 6*). Alternatively the elution was directly performed with the same inhibitor (*lane 7*) instead

of SDS sample buffer (*lane 5*). The disappearance of numerous glycoproteins when the competition assay was performed or their appearance during the specific elution with  $0.5\text{ M}$   $\alpha$ -methylmannoside suggested the presence of several N-glycosylated proteins ranging between 45 and  $>100$  kDa. Then large scale preparation was performed using total detergent extract from  $10^{10}$  tachyzoites. Thirteen gel slices (bands) were excised from the gel (as indicated in Fig. 4B) and digested with trypsin as described under "Experimental Procedures." The resulting peptides were analyzed by nano-LC-MS/MS, and the peak lists generated were submitted to classical protein database searches via the Mascot search engine that identified 26 proteins (proteins not footnoted in Tables II and III and supplemental data). Proteins were identified using a minimum of two matching peptides, and only unambiguous identifications with at least two matching peptides presenting high quality MS/MS spectra (MS/MS ion score higher than 40) were retained. As the predicted proteome of *T. gondii* was not fully represented and annotated in the available protein databases we decided to use a six-frame translation of the complete genome sequence in a second step for the protein identification. This approach allowed the identification of 11 additional proteins that were not present in the NCBI protein database of *T. gondii* (proteins footnoted in Table III and supplemental data). In many cases, the same protein was detected in multiple gel slices as mentioned above presumably due to carryover of high abundance proteins from one band to another during SDS-PAGE because of partial degra-

TABLE II  
Proteins identified by mass spectrometry using nano LC-MS/MS analysis and the NCBI/nr databases

Protein name	Organism	Accession number	Molecular mass (theoretical)	Start-end	Sequences	Gel bands	Potential N-glycosylation sites in sequence																
MyoA (TgM-A)	<i>T. gondii</i>	gi 50400802	93,334	5-12 35-56 126-143 144-155 156-164 165-174 176-185 222-240 253-260 266-276 288-297 307-316 317-342 374-386 394-404 413-422 425-434 440-455 465-479 480-497 498-509 520-531 537-563 577-602 609-634 676-684 685-701 709-719 756-761 774-793 803-808 815-831	TTSEELKTATALK GFQIWTDLAPSVKEEPDLM-FAK NQIYTTADPLVVAINPFR DLGNTTLDWIVR YRDTFDLSK LAPHVFYTAR ALDNLHAVNK IQNAIMAAANPVLEAFGNAK FMQLDVGR FGSVVAFLLLEK SYHIFYQMCK FHILPLSEYK YINPLCLDAPGIDDVAEFHE-VCESFR DGGIDDAAAIEGK ACGLLFLDAER VSYAGNQEIR WKQEDGDMLK AMYDKLFMWIIAVLNR IFMGMLDIFGFVFK NNSLEQFFINITNEMLQK NFVDIVFDRESK ELIFTSNAEVIK NNSVLALEDQCLAPGGSD-EKFLSTCK VSPNINFLISHTVGDIIQYNAE-GFLFK AEIMEIVQQSKNPVVAQLFAG-IVMEK KPLDWVPSK MLIQLHALSVLEALQLR RPFKEFLFQFK TMVFLK ECLSSWEPLVSVLEAYYAGR TPFIIR RHLVDNNSPATVQPAF	2, 3, 6, 7, 8, 9, 10, 11, 12, 13	9 sites (67, 147, 184, 245, 246, 345, 489, 659, 672)																
								MyoB/C (TgM-B)	<i>T. gondii</i>	gi 50400803	132,672	132-149 215-224 225-245 574-584 823-836 1069-1080	SIIYTTAEPLLVAINPFK FFASASSEVR TTIQDTIMAGNPVLEAFGNAK LEPSGFFLESR HLEPDSINISPEER DVSYLIGMLFQR	13	B: 8 sites (83, 94, 180, 181, 304, 427, 619, 648); C: 9 sites (42, 153, 164, 250, 251, 374, 497, 689, 718)								
																Membrane skeleton protein IMC2A	<i>T. gondii</i>	gi 133990372	168,268	65-78 117-124 152-159 244-253 261-273 288-303 337-343 347-369 475-483 555-566	VLNIEVYTYLAGAK GLPYNVLR TLNDLLLK NAVETITVEK LLLIGNTGIGEYK FLWTNEFDQTVSALAK WYDIFVK LDIPWMLTLGEEEEALVNPS-ASVR LQYVDDLYK GALYPILCWSSGR	11, 12, 13	5 sites (49, 363, 400, 409, 449)

TABLE II—continued

Protein name	Organism	Accession number	Molecular mass (theoretical)	Start-end	Sequences	Gel bands	Potential N-glycosylation sites in sequence
			Da				
				590–599	FFAPVNSQYK		
				620–630	DYLVAEFIDSR		
				649–660	DIQFMDPVAEGR		
				700–711	FYQTEINDLVSK		
				744–760	IAQAVSEMNVLAGDYDK		
				767–782	ELIDLKDTMPAADDPR		
				783–791	YAQLFQLEK		
				826–834	MILNMQVIR		
				836–845	EIVAVENDIK		
				933–944	GTQEELAAMQQR		
				957–963	VWQLLQR		
				976–989	ILVNQVEKLPEETR		
				1041–1054	AELVLAQESLLSEK		
				1158–1168	VSPLLSPEER		
				1174–1185	LVGPVQDLPQKK		
Actin	<i>T. gondii</i>	gi 1703160	41,908	30–40	AVFPSIVGKPK	2, 3, 6, 8, 9,	1 site (13)
				97–114	VAPEEHPVLLTEAPLNPK	11, 12	
				149–178	TTGIVLDSDGVSHTVPIY-EGYALPHAIMR		
				185–192	DLTEYMMK		
				198–211	GYGFTTSAEKEIVR		
				240–255	SYELPDGNIITVGNER		
				256–270	FRCPEALFQPSFLGK		
				317–327	ELTSLAPSTMK		
				361–373	EEYDESGPSIVHR		
Tubulin $\beta$ chain ( $\beta$ -tubulin)	<i>T. gondii</i>	gi 135499	49,814	63–77	AILMDLEPGTMDSVR	13	2 sites (184, 370)
				78–103	AGPFGQLFRPDNFVFGQT-GAGNNWAK		
				163–174	IMETFSVFPSPK		
				253–262	LAVNLIPFPR		
				283–297	ALSVPQLTQQMFDK		
Membrane anchor for myosin XIV precursor	<i>T. gondii</i>	gi 46948064	46,607	54–68	FVGLGNWGSYSGQK	1, 3	3 sites (101, 136, 228)
				105–121	WQSEFENVYSDANGALK		
				122–135	MPFFTVLGVDDWSR		
				144–156	TELTYAVTSEIKDGK		
				160–178	LAPADATEAAAAENHGYPK		
				240–246	TLELAPK		
				247–257	ILDYIIVADR		
				267–280	GDSMLQYYLQPLLK		
				322–347	HSGSLYYAGETGFCLFELTA-EGLVTR		

duction during sample preparation or because of the presence of multiple protein isoforms due to heterogeneity of N-linked glycans carried by these proteins. Where a protein was detected in multiple gel slices (column designated Gel bands) and multiple MS data sets were obtained, all the matching peptides identified are shown (column designated Sequences). The nature of *T. gondii* proteins identified during this proteomics study is in good agreement with the molecular masses in kilodaltons of proteins that specifically bind to the ConA as determined by SDS-PAGE (Fig. 4). Further experimental verification of proteomics data using Western blot analyses also validated the presence of different proteins such as myosin A, GAP45, and rhoptry proteins in the *T.*

*gondii* material eluted by affinity purification using ConA beads (data not shown). The three surface proteins including the minor glycosylphosphatidylinositol-anchored surface glycoprotein gp23, which has been described previously as an N-linked glycosylated protein (17), were not detected in the ConA-binding glycoproteins (data not shown). This suggests that either lower amounts of gp23 or its particular N-glycan modifications might explain the absence of gp23 identification. These results also confirm the absence of co-localization between the ConA signal and the major surface antigen SAG1 using a specific monoclonal antibody (data not shown). Several proteins of dense granules and micronemes tested by Western blots using specific monoclonal antibodies were also



TABLE III  
Identification of additional ConA binding proteins by mass spectrometry

Protein name	Organism	Accession number	Molecular mass (theoretical)	Start - end	Sequences	Gel bands	Potential N-glycosylation sites in sequence
			<i>Da</i>				
Rhoptry neck protein 2	<i>T. gondii</i>	gi 71559160	155,451	248–259 507–520 616–629 995–1010	VAMIDAVPSGIR VVADPTAYGEIFER TFCAQPTSFLLSSFR VPGFDTISAANEQLR	11	5 sites (275, 354, 393, 844, 1340)
Rhoptry neck protein 1	<i>T. gondii</i>	gi 71559158	127,355	900–914 1096–1107	DACVNQGYQSVSLTK GTLSSQQQAFMK	7	10 sites (69, 114, 632, 720, 738, 756, 774, 792, 810, 875)
Rhoptry protein 18	<i>T. gondii</i>	gi 84618297	62,342	241–250 423–432	FVSVTTGETR LFLGDFGTYR	3	1 site (377)
Rhoptry protein 7	<i>T. gondii</i>	gi 84618158	63,414	175–197 303–313 326–343 370–389 441–455 459–466 523–531 532–544	ERPQPVFTEGDPDLETNSL- YR- HQALAIGLFQK FLAPFDLVTIPGKPLVQK FVEELYELPTEDRPLADAAR VVAENAQGFSPPEVR GGLLFGPR LINPSVEAR LLALQATETPEYR	2, 3, 4, 5, 6, 7, 11, 12	1 site (525)
Apical membrane antigen 1 homolog	<i>T. gondii</i>	gi 2293476	59,978	84–94 108–121 122–142 143–153 166–172 185–194 197–211 218–231 232–257 276–293 342–355 412–425	EPAGLCPIWVGK NNFLEDVPTEKEYK QSGNPLPGGFNLNFVTPSG- QR ISPFPMELLEK CAEFAFK YRYPFVYDSK LCHILYVSMQLMEGK GPPDLTWYCFKPR KSVTENHHLIYGSAYGENP- DAFISK CLDYTELTDTVIERVESK NYGFYYVDTTGEGK FPDSFGACDVQACK	6	2 sites (58, 393)
Perforin-like protein 1	<i>T. gondii</i>	gi 118500931	124,646	423–435 448–458 797–813 1074–1088	QAVPQESVADLNR AAAPLSAVYTK AVGLTPQDLSALTGVTR TVNEPAMHVATDVGK	8	7 sites (246, 253, 333, 539, 607, 744, 1103)
Sortilin, putative <sup>a</sup>	<i>P. falciparum</i> (isolate 3D7)	gi 124809739	102,270		SWTEITDLLK SSEAATGTVAVDSIIVSPVDKR LVADYVVQFSWGDKK LGNTDHIFFQHR NVDLMYTPDFGATITR QTVSLLVSTDGCK LPVEIEER YLSLPPNNIR TSTGECEFDKVLSEGVYLAN- FK GGVWSYLK DGGVSWIEAHK LAPPRFDEDNVELL FLLSNGYFFVAK	10	8 sites (56, 112, 116, 123, 375, 869, 876, 881)
DNAJ domain protein <sup>a</sup>	<i>Plasmodium berghei</i>	gi 68072203	44,189		QNLYSVLGVK FKEISFAYEILNNAEK QVYDEYGEEGLER LNVSLEQLYK	2	1 site (44)



TABLE III—continued

Protein name	Organism	Accession number	Molecular mass (theoretical)	Start - end	Sequences	Gel bands	Potential N-glycosylation sites in sequence
			Da		VITQQMGPGFIVQNQIQDDT-CVDQGK LGHEPGDLVLVIQELPHKR IGDDLEMSIR ISLLEALVGFER SFIHLDNTPVR FMVSYPAALDEK SADPFYWWR GQQGSYPIK		
PfSec61 <sup>a</sup>	<i>P. falciparum</i>	gi 3057044	52,228			2	2 sites (139, 241)

<sup>a</sup> These proteins were identified by our in-house constructed genome database generated using complete genome sequence, downloaded as a single text file from the TIGR and segmented into regular segments of 7500 bps with a determined overlap length of 2500 bps. These nucleic acid databases were imported into the local Mascot server, and the segments were translated in all possible six reading frames. The other proteins were identified as described in Table II.

not detected in the ConA-binding material (data not shown). Taken together, these data are consistent with the presence of a limited number of surface proteins, dense granules, and micronemes in the ConA-binding material analyzed by proteomics approaches. Nevertheless many novel potentially N-glycosylated proteins in *T. gondii* were identified.

**Proteomics Analysis Reveals Many Constituents for Parasite Motility and Host-Parasite Interactions**—The proteomics approaches resulted in the unexpected identification of proteins involved in gliding motility such as myosin A (30), membrane anchor myosin XIV (also named GAP50) (19), actin, and the membrane skeleton protein IMC2A. These components play key roles in the particular gliding motility of parasites that allows *T. gondii* to enter into host cells (31–33). Strikingly bioinformatics analyses indicated that these components of the glideosome contain between one and nine Asn-X-(Ser/Thr) sequons required for N-glycosylation (Table II). The second class of N-linked glycosylated proteins that we identified are components of the late apical secretory organelles, named rhoptries (Table III). These Asn-X-(Ser/Thr)-containing glycoproteins include RON2 and AMA1 involved in moving junction formation that is required for host cell invasion (34–36). RON1 contains 10 Asn-X-(Ser/Thr) sequons, the highest number of consensus sites identified during this study, whereas RON2 has only five consensus sites (Table III). Another category of secreted proteins present in the rhoptry was also identified (Table III and supplemental data). These are known to be released from the rhoptry during host cell invasion and participate in the formation of the vacuole in which the parasites multiply (37–39). These ROPs that bound to ConA include the protein kinase-related family members ROP2, ROP5, ROP7, and ROP18 (Table III and supplemental data). However, only ROP7 and ROP18 appeared to contain one genuine N-glycosylation sequon (Table III), suggesting that the other ROPs (ROP2, ROP4, and ROP5) are unlikely to be glycosylated but were pulled down together with other *T. gondii* ConA-binding N-linked glycoproteins. In addition,

other proteins identified by proteomics also include myosin B/C known to be involved in parasite division (40) and the skeleton component tubulin (41) (Table II). Additional bioinformatics searches using our in-house genome database (see details under “Experimental Procedures”) identified all potentially N-linked glycoproteins, and in addition to the putative partners described above, 11 novel proteins were detected (proteins marked by asterisk in Table III and supplemental data). Some of these novel putative N-glycoproteins are highly conserved in the Apicomplexa, implying that they fulfill parallel roles in the biology of these parasites. For example, we identified the *Plasmodium* homologue of perforin-like protein 1 (Ref. 42 and Table III) and a putative sortilin (Ref. 43 and Table III). Surprisingly HSP70, HSP90, elongation factor 1a, and eukaryotic translation initiation factor 4A were also detected (supplemental data). Finally some proteins that do not contain the Asn-X-(Ser/Thr) sequons were also found, confirming that unglycosylated partners can be co-precipitated with *T. gondii* N-glycoproteins (supplemental data). However, HSP70 and HSP90 may be involved in the folding and stabilization of genuine N-linked glycoproteins of the glideosome, moving junction, or other subcellular compartments of the parasites.

**Deciphering the Precise Nature of N-Glycopeptides by Structural Analysis of Both N-Glycans and Peptides**—We established the structures of N-linked glycopeptides of a major *T. gondii* glycoprotein, GAP50, identified by proteomics analysis by determining the nature of both N-linked glycans and trypsin-generated peptides. Two MS/MS spectra show the presence of typical diagnostic ions, MH<sup>+</sup> *m/z* 366.14 and 528.19 corresponding to glycan backbone internal fragments. Both are found in bands 1 and 3 (MS/MS spectra of doubly protonated parent ions at *m/z* 1248.10 and 1329.50). The MS/MS spectrum of the doubly protonated parent ion at *m/z* 1248.10 shows fragment ions, which correspond to sugar neutral losses on the glycopeptide. Therefore, one ion of mass MH<sup>+</sup> *m/z* 953.47 was observed, corresponding to the peptide

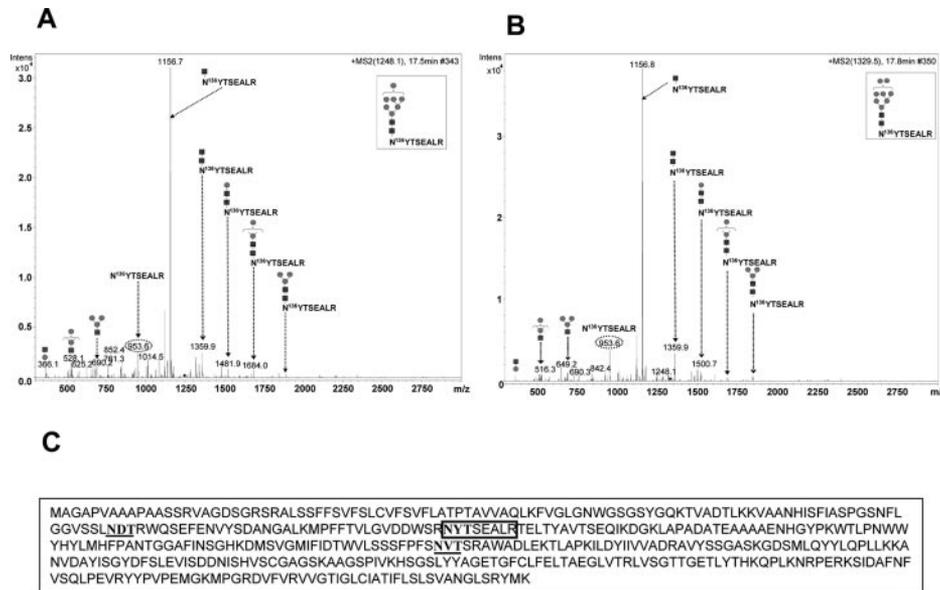
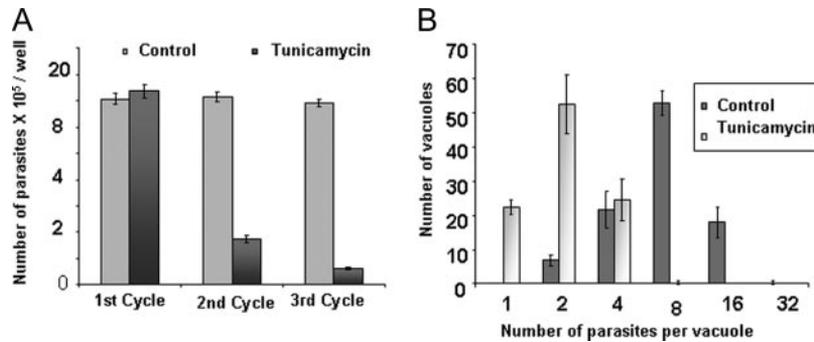


FIG. 5. Identification of peptide and N-glycans from the membrane anchor myosin XIV also named GAP50. A, the MS/MS spectrum from parent ion at  $m/z = 1248.10$  corresponding to the peptide NYTSEALR with the N-glycosylation  $\text{Man}_7(\text{GlcNAc})_2$  in position Asn<sup>136</sup>. B, the MS/MS spectrum from parent ion at  $m/z = 1329.50$  corresponding to the peptide NYTSEALR with the glycosylation  $\text{Man}_8(\text{GlcNAc})_2$  in position Asn<sup>136</sup>. C, the position of the peptide NYTSEALR with the consensus NYT alternatively modified by  $\text{Man}_7(\text{GlcNAc})_2$  and  $\text{Man}_8(\text{GlcNAc})_2$  is indicated (boxed) in the GAP50 protein sequence. The two additional N-glycosylation sequons are underlined and bold.

NYTSEALR that is a potential glycosylated tryptic peptide from the membrane anchor myosin XIV (GAP50), and was only identified in these two bands (Fig. 5). Combining the two data above allowed us to determine the glycan structure  $\text{Man}_7(\text{GlcNAc})_2$  at position Asn<sup>136</sup> of GAP50 (Fig. 5A). In addition, the MS/MS spectrum of the doubly protonated parent ion at  $m/z 1329.50$  is clearly similar to the MS/MS spectrum of parent ion  $m/z 1248.10$  (same fragment ions and presence of the ion of mass  $\text{MH}^+ m/z 953.47$ ). Only the doubly protonated parent ion masses show a mass difference of 162 Da corresponding to the mass of a mannose. These two data allowed us to define the glycan structure  $\text{Man}_8(\text{GlcNAc})_2$  at the same position, Asn<sup>136</sup>, of the membrane anchor myosin XIV (GAP50) (Fig. 5B). Taken together, the data confirmed the presence of the following two high mannose structures,  $\text{Man}_7(\text{GlcNAc})_2$  or  $\text{Man}_8(\text{GlcNAc})_2$ , on the peptide NYTSEALR with NYT as a consensus site in the membrane anchor myosin XIV. As mentioned, TgGAP50 is an integral membrane protein that anchors *T. gondii* myosin A (TgMyoA), its associated light chain (TgMLC1), actin, and TgGAP45 to the inner membrane complex, and these constitute the glideosome that is involved in motility required for host cell invasion (19). TgGAP50, which contains three Asn-X-(Ser/Thr) consensus sites (Fig. 5C, underlined and bold), has been reported previously as a genuine N-linked glycoprotein that can be cleaved off by PNGase F, and its mobility shift compared with PNGase F-untreated control suggested that the three consensus sites are likely occupied by N-glycans (19). We demonstrate here that the NYT sequon within the NYTSEALR peptide (boxed in TgGAP50 protein sequence in Fig. 5C) bears alternatively

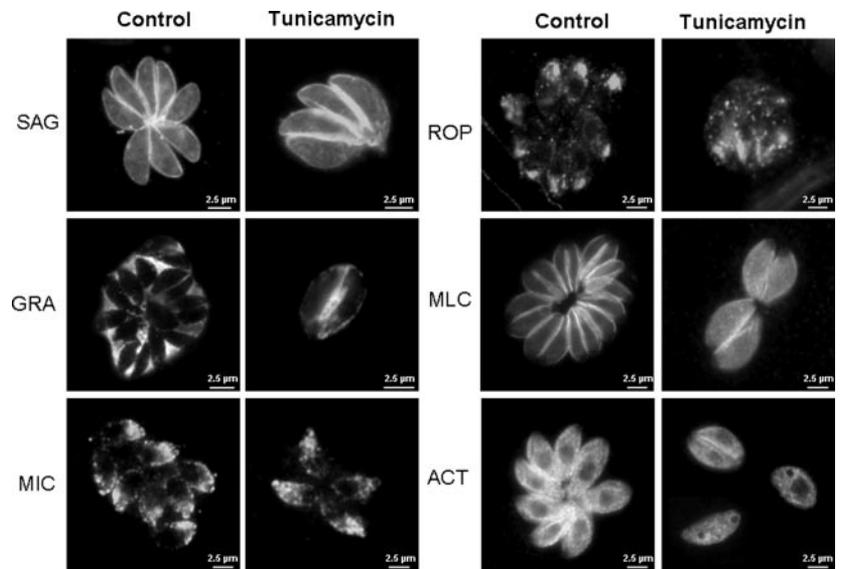
$\text{Man}_8(\text{GlcNAc})_2$  and  $\text{Man}_7(\text{GlcNAc})_2$ . The glycoprotein TgGAP50 represents a genuine pellicle component involved in the parasite's motile apparatus. However, the implication of N-glycosylation, not only in gliding motility but also in the biogenesis of the parasite's late secretory organelles, was further investigated in greater detail using both cell biology and pharmacological approaches.

*The N-Glycosylation Inhibitor Tunicamycin Blocks T. gondii Growth within Mammalian Cells*—The N-glycosylation inhibitor tunicamycin displayed an unexpected and intriguing mode of action on the intracellular tachyzoites of *T. gondii*. When parasite numbers were determined by a quantitative assay using the exogenous reporter enzyme  $\beta$ -galactosidase expressed by stable tachyzoites harboring the *lacZ* gene, we were surprised that 5  $\mu\text{g/ml}$  or higher concentration of tunicamycin showed no significant inhibitory effect on the intracellular growth of *T. gondii* tachyzoites during the first round of drug treatment (Fig. 6A, 1st Cycle). Instead the tunicamycin-treated tachyzoites displayed a normal replication rate until infected host cells were lysed at 48 h postinfection like the control tachyzoites treated with DMSO alone (Fig. 6A, 1st Cycle). In sharp contrast, when the first round of treated tachyzoites were used to infect new host cells in the presence or absence of tunicamycin, only a 25–30% growth rate relative to untreated tachyzoites was observed, suggesting that parasites were now significantly altered in division rate during the second cycle of infection (Fig. 6A, 2nd Cycle). A more drastic effect, up to 90% growth rate inhibition, was observed during the third cycle of host infection (Fig. 6A, 3rd Cycle). As mentioned, the presence of tunicamycin was not required to



**FIG. 6. Inhibitory effect of tunicamycin on *T. gondii* growth.** A, the delayed effect of tunicamycin on *T. gondii* tachyzoites growth quantified using stable  $\beta$ -galactosidase-expressing 76K strain parasites. Four wells of 24-well plates containing HFF cells infected by tachyzoites were treated with 5  $\mu$ g/ml tunicamycin or with DMSO alone. After 2 days of infection, one-fifth of each well of the first cycle treated or untreated parasites was used to infect new wells under tunicamycin or DMSO treatment. The remaining parasites of this first cycle experiment were recovered, washed with PBS, and frozen. After the second cycle of treatment, one-fifth of each well was used again to infect new monolayer HFF cells for the third cycle experiment. All parasite pellets were lysed and tested for  $\beta$ -galactosidase activity. Three independent experiments were performed, and data are expressed as mean  $\pm$  S.D. B, comparison of the growth rate of tunicamycin-treated and untreated tachyzoites. Tachyzoites treated with tunicamycin or with DMSO alone during the first cycle of infection as described were loaded onto new monolayer of HFF cells and grown for 24 h in the absence of drug. The percent distribution of vacuole size (number of parasites per vacuole) was counted for 100 vacuoles in three independent experiments, and data are expressed as mean  $\pm$  S.D.

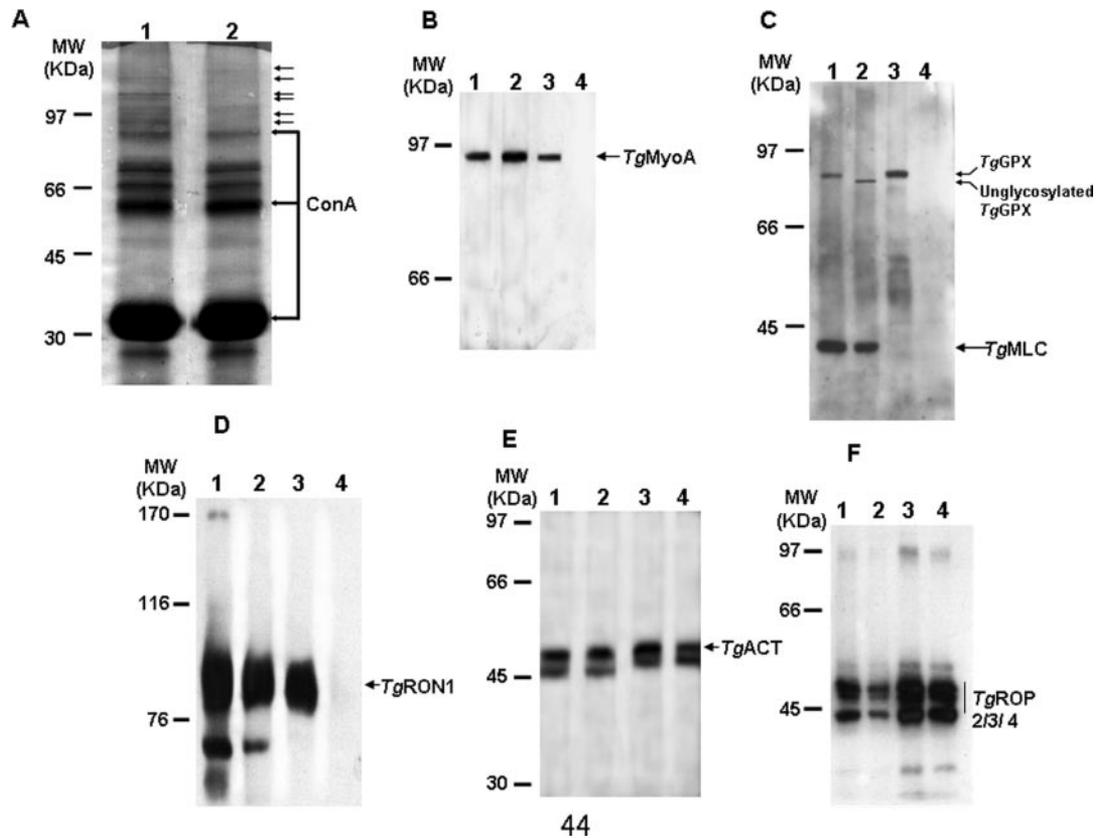
**FIG. 7. Tunicamycin alters the morphogenesis of *T. gondii* organelles.** Tachyzoites treated with tunicamycin or with DMSO alone for 48 h during the first cycle of infection were released and loaded on new monolayer HFF cells and grown for 24 h without drug. The morphology of different organelles verified by markers such as the monoclonal antibody anti-surface protein SAG1, anti-dense granule protein GRA1, anti-MIC1, anti-rhoptries ROP2/3/4, anti-MLC, and anti-actin (*ACT*) were compared between tunicamycin-treated and the control DMSO-treated *T. gondii* tachyzoites. Note the pronounced effect of tunicamycin on the morphogenesis of inner membrane complex.



obtain the significant effect on intracellular growth of the parasites during the second and third cycles. Because *T. gondii* divides by a unique binary division and because the number of parasites per vacuole reflects division rates, we examined the intracellular growth of tachyzoites after the first round of tunicamycin treatment. As can be seen in Fig. 6B, tachyzoites treated once with tunicamycin displayed a slower multiplication rate in their newly infected host cells. The control tachyzoites treated during the first cycle of infection with DMSO multiplied at normal rates, and 4–16 parasites/vacuole were obtained at 24 h postinfection (Fig. 6B). By comparison, first round tunicamycin-treated tachyzoites that received no further treatment during the second round of new host cell infection still grew more slowly. Values of one and four para-

sites/vacuole at 24 h post-infection were scored (Fig. 6B). This reduced multiplication rate that led to 1–4 parasites/vacuole in tunicamycin-treated tachyzoites versus untreated controls (8–16 parasites/vacuole) was also confirmed by direct visualization using immunofluorescence assays and different monoclonal antibodies (Fig. 7).

**Tunicamycin Alters Biogenesis of Late Secretory Organelles and the Inner Membrane Complex**—To determine whether tunicamycin could alter the morphogenesis of different secretory organelles and other subcellular compartments of *T. gondii*, tachyzoites, drug-treated for 48 h, were used to infect new monolayer HFF cells and grown for 24 h without drug. The morphology of different organelles was checked and compared with the control tachyzoites treated with DMSO and

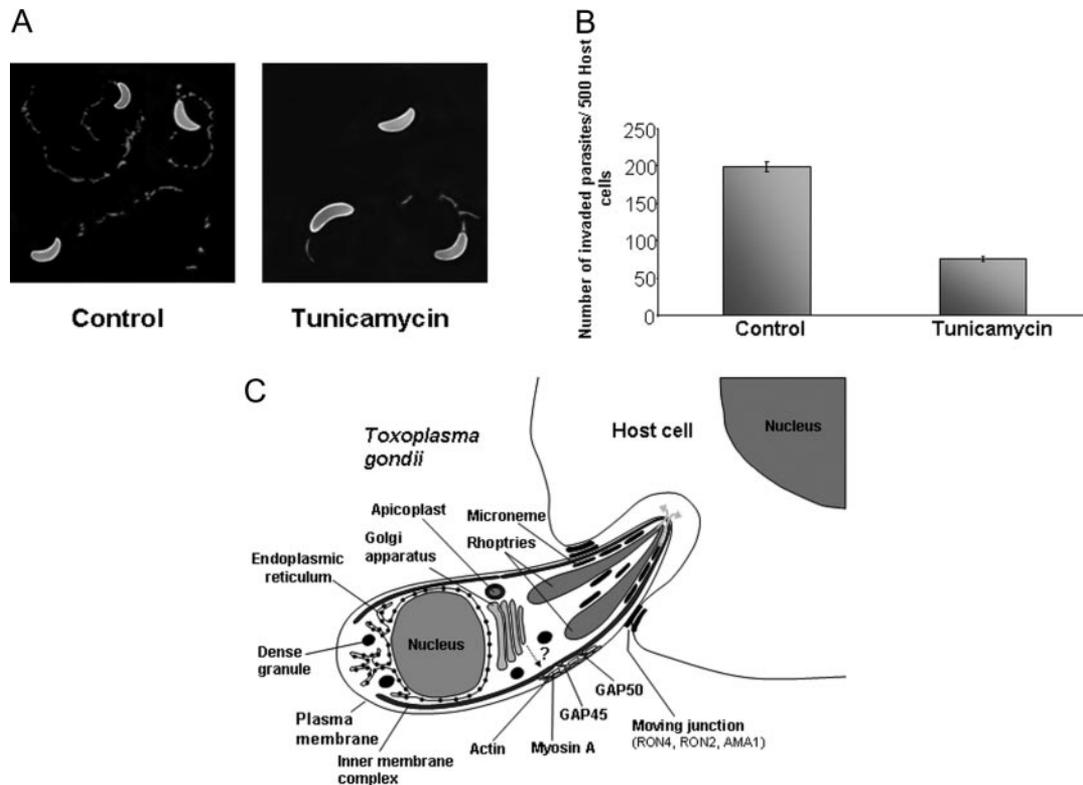


**FIG. 8. Effect of tunicamycin on ConA binding and identification of novel N-glycoproteins of *T. gondii*.** A, total detergent extract from the same number of tunicamycin-treated or untreated control (DMSO only) tachyzoites was incubated with ConA beads. After elution, the samples were analyzed by SDS-PAGE (7.5% polyacrylamide gel) followed by Coomassie Blue staining. The absence of glycoprotein binding to ConA is indicated by arrows. B–F, further verification of the tunicamycin effect on ConA binding by TgMyoA (B), TgMLC (C), TgRON1 (D), actin (TgACT; E), and TgROP2/3/4 (F) using lectin purification followed by Western blots. Lane 1, total detergent extract from the control tachyzoites; lane 2, total detergent extract from tunicamycin-treated tachyzoites; lane 3, presence of the parasite protein from the control tachyzoites detected in the ConA-binding material; lane 4, absence or presence of parasite protein from the material eluted from ConA beads when a lysate of tunicamycin-treated tachyzoites was used. Note the identification of a novel N-glycoprotein (TgGPX) that cross-reacted to anti-TgMLC antibodies. This TgGPX displayed an electrophoretic shift when the total lysate was prepared from tunicamycin-treated tachyzoites (C, lane 3) and could not bind to ConA (C, lane 4) confirming that TgGPX is N-glycosylated. The nature of this novel N-glycoprotein remains to be determined.

processed under the same experimental conditions. The plasma membrane appeared not to be affected in tunicamycin-treated tachyzoites (Fig. 7, SAG panel). Neither the constitutive secretion of dense granules within the vacuole nor the morphology of micronemes was affected in tunicamycin-treated tachyzoites (Fig. 7, GRA and MIC panels). In addition, the fluorescence pattern of rhoptries was not significantly changed after tunicamycin treatment (Fig. 7, ROP panel). In contrast, the classical fluorescence pattern of the inner membrane complex could not be visualized, suggesting that the morphogenesis of this compartment had been altered (Fig. 7, MLC panel). A more diffuse cytoplasmic staining was detected with anti-myosin, and the inner membrane complex that is known to be labeled by this myosin marker could not be distinguished (Fig. 7, MLC panel). Some treated tachyzoites also showed a few unknown structures when anti-actin was tested (Fig. 7, ACT panel). Collectively these data indicate that

tunicamycin induces important alterations in the biogenesis of *T. gondii* inner membrane complex, suggesting that the function of the parasite's motile apparatus or glideosome that is located in this compartment may be impaired.

**Inhibition of *T. gondii* N-Glycosylation by Tunicamycin**—At the molecular level, we also demonstrated that binding of *T. gondii* glycoproteins to ConA was affected when protein extracts were derived from tunicamycin-treated tachyzoites prior to lectin affinity purification. As shown in Fig. 8A, several N-glycosylated proteins ranging from 45 to >100 kDa that specifically bound to ConA, as reported in Fig. 4A, were no longer detected (see Fig. 8A, small arrows). Western blots in Fig. 8B further validated the absence of TgMyoA from the tunicamycin-treated material eluted from ConA (lane 4). The polyclonal antibodies specific to myosin light chain (MLC) cross-reacted with a novel *T. gondii* glycoprotein of 85 kDa that displayed a difference in electrophoretic mobility be-



**FIG. 9. Effect of tunicamycin on *T. gondii* gliding motility and host cell invasion.** *A*, effect of tunicamycin on gliding motility of *T. gondii*. Tachyzoites treated with tunicamycin or with DMSO alone (*Control*) for 48 h during the first cycle of infection were released and allowed to glide on serum-coated slides. Trails were visualized by staining with anti-SAG1 antibodies followed by a fluorescent secondary antibody. *B*, effect of tunicamycin on host cell entry. The tunicamycin-treated and untreated tachyzoites during the first cycle of infection described above were also used to infect new monolayer HFF cells and grown for 24 h. The intracellular tachyzoites were paraformaldehyde-fixed and stained with anti-SAG1 or anti-GRA1 antibodies. 500 vacuoles were counted, and three different experiments were performed. The data are expressed as mean  $\pm$  S.D. *C*, schematic drawing illustrating the host cell entry by *T. gondii*, the basic elements (ER, Golgi apparatus, and plasma membrane), and parasite-specific organelles (dense granules, micronemes, rhoptries, and apicoplast). The current accepted model of host cell entry or invasion proposes contributions from both the glideosome and the moving junction. The glideosome is composed of TgMyoA, actin filaments, and membrane anchor myosin XIV or GAP50 and GAP45, which are located between the plasma membrane and the inner membrane complex, whereas the moving junction contains rhoptry proteins that are secreted during host cell entry. These moving junction proteins include RON2, RON4, and the microneme AMA1. In light of the observations related in this study, we describe that several components of both glideosome and moving junction are potentially *N*-glycosylated or contain a number of consensus sites suggesting them to be *N*-glycosylated. We believe that this post-translational modification may be required for their proper intracellular transport to the pellicle and for protein-protein interactions.

tween tunicamycin-treated and untreated tachyzoites, respectively (Fig. 8C, lanes 1 and 2). As expected, when deglycosylated the 85-kDa protein was unable to bind ConA (lane 4). The precise nature of this novel 85-kDa glycoprotein remains to be determined. In addition, we demonstrated that the rhoptry neck protein RON1 was not capable of binding ConA when tunicamycin-treated (lane 4) compared with untreated lysate (lane 3). Because of the lower amount of rhoptry neck protein 2 detected by Western blots combined with our identification of only one peptide isolated by proteomics approaches, it was not surprising that we were unable to validate the *N*-glycosylated status of RON2 using total extract of tunicamycin-treated extracts (data not shown). In contrast, we were able to show that actin (Fig. 8E), ROP2/3/4 (Fig. 8F), and probably other cytoplasmic proteins can bind to ConA despite tunicamycin treatment, suggesting that these abundant pro-

teins interact nonspecifically with the lectin. Nevertheless our proteomics data together with the direct evidence of the *N*-glycan structures  $\text{Man}_7(\text{GlcNAc})_2$  and  $\text{Man}_8(\text{GlcNAc})_2$  at position Asn<sup>136</sup> of the peptide NYTSEALR of GAP50 and the lack of ConA binding by TgMyoA after tunicamycin treatment also suggest that several crucial proteins involved in the parasite's motility and host-parasite interactions are likely *N*-glycosylated.

**Tunicamycin Blocks *T. gondii* Motility and Host Cell Invasion**—To investigate the physiological role of *N*-glycosylation in the parasite's motility, we monitored the gliding of drug-treated and untreated tachyzoites using immunofluorescence staining to detect the trails left on slides by motile parasites (Fig. 9A). When intracellular tachyzoites were treated with 5  $\mu\text{g/ml}$  tunicamycin for 40 h and purified, the freshly released and tunicamycin-treated tachyzoites were unable to display parasite-specific gliding as no characteristic trail motility was

visualized for individual parasites stuck to serum-coated glass (Fig. 9A, *Tunicamycin panel*). Only a few tachyzoites treated by tunicamycin showed smaller and aborted trails, whereas the untreated tachyzoites (incubated with DMSO only) that were analyzed under the same experimental conditions were entirely motile as demonstrated by the prominent trail length for each individual tachyzoite (Fig. 9A, *Control panel*). These data on the inhibition of N-glycosylation of components involved in parasite's gliding motility are consistent with tunicamycin treatment affecting the host cell invasion by *T. gondii*.

To establish whether tunicamycin was blocking the attachment or invasion of host cells, we monitored host cell reinfection after the first cycle of intracellular parasite treatment. To do this, we examined and quantified the number of parasites attached to or invading host cells. Invasion was scored during a 1-h pulse infection followed by 24 h of intracellular growth. Fig. 9B shows an approximately 70% decrease in tachyzoite numbers within host cells, suggesting that tunicamycin treatment impaired host cell invasion.

#### DISCUSSION

In this study, we have provided new evidence that *T. gondii* contains numerous N-glycosylated proteins that unexpectedly are components of the glideosome, moving junction, and other glycoproteins involved in host cell-parasite interactions. The cell wall or pellicle of apicomplexan parasites like *T. gondii* consists of the plasma membrane and the closely associated, flattened cisternae of the inner membrane complex. Both actin and myosin A homologues have been localized to the space between the plasma membrane and the inner membrane, and this glideosome is a key player in motility and host cell entry by *T. gondii* (see schematic model in Fig. 9C). Another essential element represents the moving junction, a structure built at the interface between the host cell and the parasite during its active entry into any kind of mammalian cell. The moving junction is a circumferential zone that forms at the apical tip of the parasite, moves backward, and pinches the vacuole from the host cell membrane (34, 36). Several components of the moving junction are secreted from late secretory organelles such as rhoptries (RON2) and micronemes (AMA1) (depicted in Fig. 9C). It is clear that the results presented here identified N-linked glycosylation on several key constituents involved in host-parasite interactions. The major *Toxoplasma* N-linked glycans released from total detergent extract proteins by PNGase F are oligomannosidic ( $\text{Man}_{5-8}(\text{GlcNAc})_2$ ) and paucimannosidic structures ( $\text{Man}_{3-4}(\text{GlcNAc})_2$ ), sugars that are rarely present on mature N-glycoproteins of higher eukaryotes. The three major N-glycans linked to proteins correspond to  $\text{Man}_6(\text{GlcNAc})_2$ ,  $\text{Man}_7(\text{GlcNAc})_2$ , and  $\text{Man}_8(\text{GlcNAc})_2$ , respectively. In addition, four other minor oligosaccharides,  $\text{Man}_3(\text{GlcNAc})_2$ ,  $\text{Man}_4(\text{GlcNAc})_2$ ,  $\text{Man}_5(\text{GlcNAc})_2$ , and an extremely limited amount of  $\text{Man}_9(\text{GlcNAc})_2$ , were also found. It has been

shown that other eukaryotes can transfer structures other than the largest lipid-linked oligosaccharide precursor, Dol-PP-Glc<sub>3</sub>Man<sub>9</sub>(GlcNAc)<sub>2</sub> (12–14, 20). Thus, we postulate that the paucimannosidic N-linked glycan structures found on *T. gondii* glycoproteins could point to the transfer of  $\text{Man}_5(\text{GlcNAc})_2$  and its subsequent glucosylation as reported previously in *Trypanosoma brucei* (14). The genes encoding numerous endoplasmic reticulum-located glycosyltransferases, named ALG (asparagine-linked glycosylation), required for the biosynthesis of the lipid (dolichylpyrophosphate)-linked glycosyl donors and oligosaccharyltransferases needed for the N-glycan transfer to proteins, can be readily identified in the *T. gondii* genome (ToxoDB) using detailed bioinformatics and gene predictions.<sup>3</sup> All of these genes involved in both synthesis of lipid precursor and transfer of N-glycan structures are transcribed in *T. gondii* tachyzoites and bradyzoites.<sup>4</sup> In contrast, our biochemical analyses failed to detect sialylated, galactosyl, or fucosylated structures consistent with either the likely absence or the limited presence of complex glycan types containing sialic acid, galactose, and fucose residues in *T. gondii*. This also indicates that a part of the endoplasmic reticulum and Golgi trimming and maturation pathways that are highly conserved in other eukaryotes is absent in the parasite. However, only structures of major N-glycans were determined when the total detergent glycoprotein extracts were analyzed. We therefore cannot rule out the presence of other minor modifications and branching monosaccharides that escaped detection in this study.

We identified a number of known proteins that have not been suspected until now to be N-glycosylated. Conversely N-glycosylation is considered as a rare post-translational modification in apicomplexan parasites because most investigations have focused attention on surface and other antigens that are important for host cell-parasite interactions and future vaccine development. Our data confirmed that the key glideosome membrane anchor myosin XIV (also named GAP50), known to be involved in both parasite motility and host cell entry, contains N-glycosylated structures. From the mixed glycopeptide populations, we were fortunate to identify the trypsin-generated peptide NYT<sub>5</sub>SRALR of the membrane anchor myosin XIV (GAP50) that is truly modified by two N-glycan structures at the same position, Asn<sup>136</sup>. We report for the first time in *T. gondii* the fine N-glycan structures that are present on the GAP50 sequon NYT that is alternatively modified by  $\text{Man}_8(\text{GlcNAc})_2$  or  $\text{Man}_7(\text{GlcNAc})_2$ . However, it was somewhat unexpected to find that TgMyoA, another component of the glideosome, was also capable of specifically binding to ConA. We have shown that TgMyoA failed to bind to ConA when the parasites were treated with the N-glycosylation poison tunicamycin. Neither the N-terminal signal peptide nor a transmembrane domain is present in TgMyoA.

<sup>3</sup> F. Dzierszinski, D. Ross, and S. Tomavo, unpublished data.

<sup>4</sup> S. Fauquenoy and S. Tomavo, unpublished data.

These data therefore suggest that an *N*-glycosylated binding partner might bring TgMyoA into interaction with ConA. As stated, no obvious domains of TgMyoA presently explain how it might be translocated across the ER membrane to the lumen where the addition of *N*-glycans is expected to occur co-translationally (10). However, almost nothing is known about ER translocation machinery in *T. gondii*, and previous work on *Saccharomyces cerevisiae* has described that carboxypeptidase Y and an acid phosphatase can be translocated to the ER, enter the secretory pathway, and be *N*-glycosylated without having their N-terminal signal sequences (44, 45). In addition, it has been shown that cytoplasmic and nuclear proteins are *N*-glycosylated, suggesting the existence of post-translational *N*-glycosylation mechanisms in higher eukaryotes (46, 47). It remains to be determined whether such a non-conventional *N*-glycosylated biosynthetic pathway operates in *T. gondii*. To our knowledge, almost nothing has been reported on *N*-glycosylation of myosins and components involved in the motile apparatus in other eukaryotic cells. Gliding is a unique form of apicomplexan parasite motility that manifests as either circular spirals or a series of helical turns revolving around the long axis of the body of the parasite (30–32). It has been suggested previously that the soluble protoglideosome and membrane-associated Tg-GAP50 are transported separately and are assembled into the glideosome in the inner membrane complex of mature parasites (19). The *N*-linked glycosylation of glideosome components such as GAP50 suggests that this post-translational modification may be involved in their intracellular trafficking mechanisms or in their transport to the parasite's space where the glideosome are assembled. It is also interesting to note that two key components of the moving junction (RON2 and AMA1) required for host cell invasion (Fig. 9C) are also present in material purified by ConA. These moving junction proteins also contain putative *N*-glycosylation sites. In addition, we showed that another rhoptry neck protein, RON1, with presently unknown function may be *N*-glycosylated. It should be mentioned that RON1 displayed the strongest fluorescence that co-localized with the ConA signal, and its deglycosylation by tunicamycin prevented lectin binding. However, we cannot rule out that RON1 may indirectly bind to ConA via other *N*-glycan-bearing RONS. We were not able to confirm that RON2 and AMA1 are *N*-glycosylated using Western blots because of the limited amount of protein detected after lectin purification. Their glycopeptides modified by *N*-glycan structures were also not identified using MALDI-MS/MS analyses because only a few peptides (only one peptide from RON2) from these proteins were obtained. Further biochemical analyses including purification of higher amounts of RON2 and AMA1 by affinity purification with their respective specific antibodies followed by direct determination of *N*-glycan structures using mass spectrometry are required. Many other proteins have been isolated by ConA affinity purification and identified by proteomics analyses. Among these

are also proteins that do not have any predicted *N*-glycosylation sites (supplemental data), suggesting that these proteins may interact with other genuine *N*-glycosylated proteins. This may be the case for MIC1 (supplemental data) that has been reported previously as a parasite lectin (48) that might be retained on ConA by interacting with *N*-glycan structures carried by other *N*-glycoproteins that specifically bind to ConA. Even if some proteins without *N*-glycosylated sites (supplemental data) may behave like MIC1, we cannot exclude that these components may also interact with ConA nonspecifically as further demonstrated during this study for actin and other rhoptry proteins such as ROP2/3/4.

It is intriguing that unlike all other eukaryotic cells, *T. gondii* tachyzoites treated with tunicamycin at conventional inhibitory concentrations did not display any obvious growth defect within its host cells during its first cycle of intracellular development that lasts 48 h. The growth defect appeared only during the second cycle of reinvasion of host cells with a very prominent effect during the third cycle. The effect of tunicamycin took place in the absence of drug during the second and third cycle of host cell infection, suggesting that this delayed effect, or unusual kinetics of tunicamycin on *T. gondii*, may reflect drastic differences in host cell re-entry by tunicamycin-treated relative to untreated tachyzoites. This conclusion is supported by the results that demonstrate the ability of tunicamycin to affect gliding motility of tachyzoites released after intracellular drug treatment. We cannot exclude that host cell attachment may also be affected by tunicamycin treatment. However, the lack of *N*-glycosylation of several SAGs and most MICs reported by lectin purification and proteomics analyses indicates that any interference of host cell attachment by tunicamycin may be marginal. Therefore, we suspect that it is the parasite's motility and new host cell invasion that are likely impaired in tunicamycin-treated tachyzoites. It remains to be determined whether moving junction formation is also inhibited by tunicamycin and absence of *N*-glycosylation. However, the delayed effect on intracellular replication of tunicamycin-treated parasites that have entered new host cells is highly reminiscent of the delayed death phenomenon seen with apicoplast inhibitors (49). It appears that ConA fluorescence co-localizes with the signal of apicoplast-specific monoclonal antibodies. Because it has been described that some apicoplast proteins transit through the ER (29, 50), it is likely that this plastid-like organelle may also contain *N*-glycosylated proteins and that the delayed effects seen with tunicamycin might be partially attributable to inhibition of apicoplast functions.

In summary, the current study illuminates a diverse repertoire of *N*-glycosylated proteins that contribute to parasite survival and pathogenesis during mammalian host cell infection. Novel parasite-specific *N*-glycoproteins identified here can significantly expand the number of potential targets for therapeutic intervention. We anticipate that ongoing in-depth exploration of biological functions of the identified *N*-linked



glycans carried by components involved in gliding motility, moving junction, protein secretion, and protein-protein interaction would further unravel the invasive and pathogenic mechanisms developed by *T. gondii* and other apicomplexan parasites during infection.

**Acknowledgments**—We thank Drs. Dominique Soldati, Maryse Lebrun, Jean-François Dubremetz, Peter Bradley, and John Boothroyd for generous gifts of antibodies and Drs. Steven Ball and Gordon Langsley for a critical review of the manuscript. We acknowledge the *Toxoplasma* Genome Sequencing Consortium for making available the genome database. Preliminary genomic and/or cDNA sequence data were accessible via ToxoDB. The mass spectrometry facility used at the University of Lille was funded by the European Community (FEDER), the Région Nord-Pas de Calais (France), the CNRS, and the Université des Sciences et Technologies de Lille.

\* This work was supported by the CNRS and INSERM. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

☐ The on-line version of this article (available at <http://www.mcponline.org>) contains supplemental material.

|| To whom correspondence should be addressed: Equipe de Parasitologie Moléculaire, Unité de Glycobiologie Structurale et Fonctionnelle, CNRS UMR 8576, Bâtiment C9, Université des Sciences et Technologies de Lille, 59655 Villeneuve d'Ascq, France. Tel.: 33-3-20-43-69-41; Fax: 33-3-20-33-65-55; E-mail: Stan.Tomavo@univ-lille1.fr.

## REFERENCES

- Kim, K., and Weiss, L. M. (2004) *Toxoplasma gondii*: the model apicomplexan. *Int. J. Parasitol.* **34**, 423–432
- Dubey, J. P. (1994) Toxoplasmosis. *J. Am. Vet. Med. Assoc.* **205**, 1593–1598
- Dubey, J. P. (1998) Advances in the life cycle of *Toxoplasma gondii*. *Int. J. Parasitol.* **28**, 1019–1024
- Luft, B. J., Hafner, R., Korzun, A. H., Lepout, C., Antoniskis, D., Bosler, E. M., Bourland, D. D., Uttamchandani, R., Fuhrer, J., Jacobson, J., Morlat, P., Vilde, J.-L., and Remington, J. S. (1993) Toxoplasmic encephalitis in patients with the acquired immunodeficiency syndrome. Members of the ACTG 077p/ANRS 009 study team. *N. Engl. J. Med.* **329**, 995–1000
- Dobrowolski, J. M., Carruthers, V. B., and Sibley, L. D. (1997) Participation of myosin in gliding motility and host cell invasion by *Toxoplasma gondii*. *Mol. Microbiol.* **26**, 163–173
- Carruthers, V., and Boothroyd, J. C. (2006) Pulling together: integrated model of *Toxoplasma* cell invasion. *Curr. Opin. Microbiol.* **10**, 83–89
- Carruthers, V., and Sibley, L. D. (1997) Sequential protein secretion from three distinct organelles of *Toxoplasma gondii* accompanies invasion of human fibroblasts. *Eur. J. Cell Biol.* **73**, 114–123
- Jewett, T. J., and Sibley, L. D. (2003) Aldolase forms a bridge between cell surface adhesins and the actin cytoskeleton in apicomplexan parasites. *Mol. Cell* **11**, 885–894
- Soldati, D., and Meissner, M. (2004) *Toxoplasma* as a novel system for motility. *Curr. Opin. Cell Biol.* **16**, 32–40
- Kornfeld, R., and Kornfeld, S. (1985) Assembly of asparagine-linked oligosaccharides. *Annu. Rev. Biochem.* **54**, 631–664
- Aebi, M., and Hennet, T. (2001) Congenital disorders of glycosylation: genetic model systems lead the way. *Trends Cell Biol.* **11**, 136–141
- Turco, S. J., and Pickard, J. L. (1982) Altered G-protein glycosylation in vesicular stomatitis virus-infected glucose-deprived baby hamster kidney cells. *J. Biol. Chem.* **257**, 8674–8679
- Parodi, A. J. (1993) N-Glycosylation in trypanosomatid protozoa. *Glycobiology* **3**, 193–199
- Jones, D. C., Mehlert, A., Güther, L. S., and Ferguson, M. A. J. (2005) Deletion of the glucosidase II gene in *Trypanosoma brucei* reveals novel N-glycosylation mechanisms in the biosynthesis of variant surface glycoprotein. *J. Biol. Chem.* **280**, 35929–35942
- Handman, E., Goding, J. W., and Remington, J. S. (1980) Detection and characterization of membrane of *Toxoplasma gondii*. *J. Immunol.* **124**, 2578–2583
- de Carvalho, L., Souto-Padron, T., and de Souza, W. (1991) Localization of lectin-binding sites and sugar-binding proteins in tachyzoites of *Toxoplasma gondii*. *J. Parasitol.* **77**, 156–161
- 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**, 713–721
- Shans-Eldin, H., Blaschke, T., Anhlan, D., Niehus, S., Müller, J., Azzouz, N., and Schwarz, R. T. (2005) High-level expression of the *Toxoplasma gondii* STT3 gene is required for suppression of the yeast STT3 gene mutation. *Mol. Biochem. Parasitol.* **143**, 6–11
- 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
- Samuelson, J., Banerjee, S., Magnelli, P., Cui, J., Kelleher, D. J., Gilmore, R., and Robbins, P. W. (2005) The diversity of dolichol-linked precursors to Asn-linked glycans likely results from secondary loss of sets of glycosyltransferases. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 1548–1553
- Dzierszynski, F., Popescu, O., Torsel, C., Slomianny, C., Yahiaoui, B., and Tomavo, S. (1999) The protozoan parasite *Toxoplasma gondii* expresses two functional plant-like glycolytic enzymes. Implications for evolutionary origin of apicomplexans. *J. Biol. Chem.* **274**, 24888–24895
- Dzierszynski, F., Mortuaire, M., Dendouga, N., Popescu, O., and Tomavo, S. (2001) Differential expression of two plant-like enolases with distinct enzymatic and antigenic properties during stage conversion of the protozoan parasite *Toxoplasma gondii*. *J. Mol. Biol.* **309**, 1017–1027
- Seeber, F., and Boothroyd, J. C. (1996) *Escherichia coli*  $\beta$ -galactosidase as an *in vitro* and *in vivo* reporter enzyme and stable transfection marker in the intracellular protozoan parasite *Toxoplasma gondii*. *Gene (Amst.)* **169**, 39–45
- Ciucanu, I., and Kerek, F. (1984) A simple and rapid method for the permethylation of carbohydrates. *Carbohydr. Res.* **131**, 209–217
- Carr, S., Aebersold, R., Baldwin, M., Burlingame, A., Clauser, K., and Nesvizhskii, A. (2004) The need for guidelines in publication of peptide and protein identification data. Working Group on Publication Guidelines for Peptide and Protein Identification Data. *Mol. Cell. Proteomics* **3**, 531–533
- Bradshaw, R. A. (2005) Revised draft guidelines for proteomic data publication. *Mol. Cell. Proteomics* **4**, 1223–1225
- 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
- Coppin, A., Varré, J.-S., Lienard, L., Dauvillée, D., Guérardel, Y., Soyergobillard, M.-O., Buléon, A., Ball S., and Tomavo, S. (2005) Evolution of plant-like crystalline storage polysaccharide in the protozoan parasite *Toxoplasma gondii* argues for a red alga ancestry. *J. Mol. Evol.* **60**, 257–267
- Waller, R. F., Keeling, P. J., Donald, R. G., Striepen, B., Handman, E., Lang-Unnasch, N., Cowman, A. F., Besra, G. S., Roos, D. S., and McFadden, G. I. (1998) Nuclear-encoded proteins target to the plastid in *Toxoplasma gondii* and *Plasmodium falciparum*. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 12352–12357
- 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
- Dobrowolski, J. M., and Sibley, L. D. (1996) *Toxoplasma* invasion of mammalian cells is powered by the actin cytoskeleton of the parasite. *Cell* **84**, 933–939
- 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
- Sibley, L. D. (2004) Intracellular parasite invasion strategies. *Science* **304**, 248–253
- Alexander, D. L., Mital, J., Ward, G. E., Bradley, P., and Boothroyd, J. C. (2005) Identification of the moving junction complex of *Toxoplasma*

- gondii*: a collaboration between distinct secretory organelles. *PLoS Pathog.* **1**, e17
35. Bradley, P. J., Ward, C., Cheng, S. J., Alexander, D. L., Collier, S., Coombs, G. H., Dunn, J. D., Ferguson, D. J., Sanderson, S. J., Wastling, J. M., and Boothroyd, J. C. (2005) Proteomic analysis of rhoptry organelles reveals many novel constituents for host-parasite interactions in *Toxoplasma gondii*. *J. Biol. Chem.* **280**, 34245–34258
  36. Lebrun, M., Michelin, A., El Hajj, H., Poncet, J., Bradley, P. J., Vial, H., and Dubremetz, J.-F. (2005) The rhoptry neck protein RON4 re-localizes at the moving junction during *Toxoplasma gondii* invasion. *Cell. Microbiol.* **7**, 1823–1833
  37. El Hajj, H., Demey, E., Poncet, J., Lebrun, M., Wu B., Galeotti, N., Fourmeaux, M. N., Mercereau-Puijalon, O., Vial, H., Labesse, G., Dubremetz, J.-F. (2006) The ROP2 family of *Toxoplasma gondii* rhoptry proteins: proteomic and genomic characterization and molecular modeling. *Proteomics* **6**, 5773–5784
  38. El Hajj, H., Lebrun, M., Fourmaux, M. N., Vial, H., and Dubremetz, J.-F. (2007) Inverted topology of the *Toxoplasma gondii* ROP5 rhoptry protein provides new insights into the association of the ROP2 protein family with the parasitophorous vacuole membrane. *Cell. Microbiol.* **9**, 54–64
  39. Saeij, J. P., Boyle, J. P., Collier, S., Taylor, S., Sibley, L. D., Brooke-Powell, E. T., Ajioka, J. W., and Boothroyd, J. C. (2006) Polymorphic secreted kinases are key virulence factors in toxoplasmosis. *Science* **314**, 1780–1783
  40. Delbac, F., Sanger, A., Neuhaus, E., Stratmann, R., Ajioka, W. J., Toursel, C., Herm-Gotz, A., Tomavo, S., Soldati, T., and Soldati, D. (2001) *Toxoplasma gondii* myosins B/C: one gene, two tails, two localizations, and a role in parasite division. *J. Cell Biol.* **155**, 613–623
  41. Morrissette, N. S., Murray, J. M., and Roos, D. S. (1997) Subpellicular microtubules associate with an intramembranous particle lattice in the protozoan parasite *Toxoplasma gondii*. *J. Cell Sci.* **110**, 35–42
  42. Kadota, K., Ishino, T., Matsuyama, T., Chinzei, Y., and Yuda, M. (2004) Essential role of membrane-attack protein in malarial transmission to mosquito host. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 16310–16315
  43. Nielsen, M. S., Madsen, P., Christensen, E. I., Nykjaer, A., Gliemann, J., Kasper, D., Pohlmann, R., and Peterson, C. M. (2001) The sortilin cytoplasmic tail conveys Golgi-endosome transport and binds the VHS domain of the GGA2 sorting protein. *EMBO J.* **20**, 2180–2190
  44. Blachy-Dyson, E., and Stevens, T. H. (1987) Yeast carboxypeptidase Y can be translocated and glycosylated without its amino-terminal signal sequence. *J. Cell Biol.* **104**, 1183–1191
  45. Silve, S., Monod, M., Hinnen, A., and Haguenaer-Tsapis, R. (1987) The yeast acid phosphatase can enter the secretory pathway without its N-terminal signal sequence. *Mol. Cell. Biol.* **7**, 3306–3314
  46. Tsai, B., Ye, Y., and Rapoport, T. A. (2002) Retro-translocation of proteins from the endoplasmic reticulum into the cytosol. *Nat. Rev. Mol. Cell Biol.* **3**, 246–255
  47. Kane, R., Murtagh, J., Finlay, D., Marti, A., Jaggi, R., Blatchford, D., Wilde, C., and Martin, F. (2002) Transcription factor NFIC undergoes N-glycosylation during early mammary gland involution. *J. Biol. Chem.* **277**, 25893–25903
  48. Lourenço, E. V., Pereira, S. R., Faça, V. M., Coelho-Castelo, A. A., Mineo, J. R., Roque-Barreira, M. C., Greene, L. J., and Panunto-Castelo, A. (2001) *Toxoplasma gondii*: micronemal protein MIC1 is a lactose-binding lectin. *Glycobiology* **11**, 541–547
  49. Fichera, M. E., and Roos, D. S. (1997) A plastid organelle as a drug target in apicomplexan parasites. *Nature* **390**, 407–409
  50. Parsons, M., Karnataki, A., Feagin, J. E., and DeRocher, A. (2007) Protein trafficking to the apicoplast: deciphering the apicomplexan solution to secondary endosymbiosis. *Eukaryot. Cell* **6**, 1081–1088
  51. Robinson, S., Routledge, A., and Thomas-Oates, J. (2005) Characterization and proposed origin of mass spectrometric ions observed 30 Th above the ionised molecules of per-O-methylated carbohydrates. *Rapid Commun. Mass Spectrom.* **19**, 3681–3688

### 3. Caractérisation des glycosylations de la protéine GAP50

#### 3.1. Contexte de l'étude

La protéine GAP50 fait partie des constituants majeurs du glideosome. Cette protéine est entièrement membranaire et constitue le point d'ancrage de l'ensemble du mécanisme d'invasion des cellules hôtes par *T. gondii* [Gaskins et al. 2004]. L'étude protéomique globale réalisée précédemment (§ 2) a permis de montrer que cette protéine est glycosylée. Elle possède trois séquences consensus (Asn<sup>101</sup>, Asn<sup>136</sup> et Asn<sup>228</sup>) mais seules deux glycoformes du site Asn<sup>136</sup> ont été caractérisées jusqu'à présent : Man<sub>7</sub>GlcNAc<sub>2</sub> et Man<sub>8</sub>GlcNAc<sub>2</sub>.

Pour mieux comprendre son rôle, nous avons choisi de caractériser plus finement les glycosylations et plus particulièrement la micro-hétérogénéité de cette protéine en appliquant la méthodologie développée dans le Chapitre 2 de ce manuscrit. L'objectif de ce travail était de déterminer le rôle de chacune des glycosylations de GAP50 dans le mécanisme du glideosome. Cette étude fait l'objet d'une publication en cours de rédaction.

#### 3.2. Méthodologie

##### 3.2.1. Préparation d'échantillon

L'extrait protéique purifié contenant GAP50 et ses partenaires a été séparé sur gel d'électrophorèse monodimensionnel SDS-PAGE. Les bandes ont été révélées par une coloration au bleu de Coomassie. Les protéines ont ensuite été réduites et alkylées puis digérées avec une endoprotéase.

##### 3.2.2. Caractérisation des glycosylations de GAP 50

La bande de gel contenant la protéine GAP50 a été divisée en deux. La 1<sup>ère</sup> moitié a été digérée avec de la trypsine et la 2<sup>nde</sup> avec de la chymotrypsine. Cette deuxième enzyme permet d'obtenir des glycopeptides de taille plus homogène et ainsi d'optimiser le paramétrage du spectromètre de masse pour maximiser leur intensité. Après extraction, les peptides ont été analysés en utilisant la stratégie développée au cours de ce travail de thèse (Chapitre 2). Cette stratégie comprend une analyse nanoLC-MS en mode scan alterné pour générer des ions diagnostiques et une analyse nanoLC-MS/MS. Les expériences ont été réalisées avec un couplage nanoAcquity (Waters) / maXis (Bruker Daltonics).

L'analyse nanoLC-MS a permis de déterminer les zones d'élution des glycopeptides en extrayant les ions diagnostiques (m/z 204,09 et 366,14) et ainsi d'identifier facilement les spectres de

fragmentations correspondant à des glycopeptides dans l'analyse nanoLC-MS/MS. Ces spectres ont été interprétés manuellement en tenant compte des règles de fragmentation propres aux glycopeptides.

Le digestat chymotrypsique a ensuite été analysé par nanoLC-MS avec un couplage nanoAcquity/Synapt G1 (Waters). Les courants des ions des glycopeptides ont été extraits du chromatogramme. La somme des spectres pour chacun des pics chromatographique obtenus a permis d'obtenir la distribution des glycoformes pour chacun des sites de glycosylation.

### 3.3. Résultats

#### 3.3.1. Digestion enzymatique

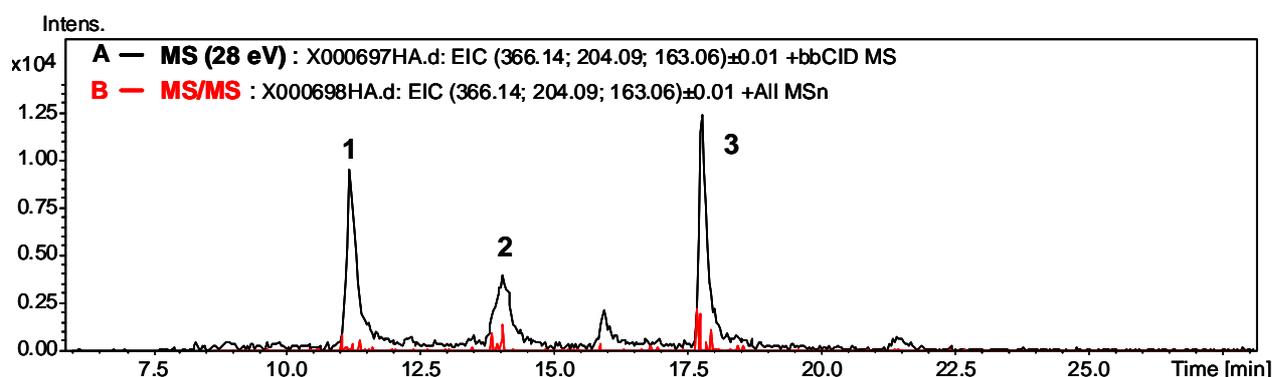
La protéine GAP50 possède trois sites de glycosylation potentiels (Asn<sup>101</sup>, Asn<sup>136</sup> et Asn<sup>228</sup>). Lors de l'étude précédemment publiée (annexe 1) la trypsine a été utilisée comme endoprotéase pour la digestion de la glycoprotéine. Dans ce cas, une digestion trypsique ne permet de générer qu'un seul glycopeptide remplissant les critères adaptés pour sa caractérisation par spectrométrie de masse. Il s'agit du peptide <sup>136</sup>NYTSEALR<sup>143</sup> glycosylé en Asn<sup>136</sup> qui a été identifié lors de l'étude protéomique précédente. Les deux autres peptides trypsiques qui contiennent les deux autres sites de glycosylation, sont de trop grande taille pour être séquencés (2859 et 2922 Da).

C'est pourquoi, après cette étude assez générale, nous avons caractérisé spécifiquement GAP50 en utilisant la méthodologie développée dans le Chapitre 2. Nous avons donc choisi une enzyme permettant d'obtenir des peptides de taille moyenne et relativement homogène afin d'être facilement analysés par spectrométrie de masse : la chymotrypsine. Le clivage de GAP50 par cette enzyme génère des peptides portant les sites de glycosylation de 900 à 1300 Da environ. Il est ainsi possible d'optimiser le paramétrage du spectromètre de masse pour obtenir une intensité maximale pour les trois glycopeptides de GAP50 en même temps.

#### 3.3.2. Analyse nanoLC-MS : détection des ions diagnostiques

Dans l'étude précédente, le digestat trypsique de GAP50 a été analysé avec une trappe ionique. Nous avons vu dans le Chapitre 2 que ce type d'instrument n'est pas le plus adapté pour détecter les ions diagnostiques. Le digestat chymotrypsique de GAP 50 a donc été analysé par nanoLC-MS sur une Q-TOF (maXis, Bruker Daltonics) en utilisant une méthode alternant un scan à basse (10 eV) et un scan à haute (28eV) énergie dans la cellule de collision (mode scan alterné) pour générer des ions diagnostiques et déterminer les zones d'éluion des glycopeptides. Cette analyse a été suivie d'une analyse nanoLC-MS/MS pour caractériser les glycopeptides. La Figure 4 montre que les ions diagnostiques générés en appliquant une énergie élevée dans la cellule de collision lors de l'analyse en mode scan alterné sont nettement plus intenses que lorsqu'ils sont issus d'une

fragmentation par MS/MS. L'extraction des courants des ions diagnostiques permet de détecter trois zones d'élution de glycopeptides (Figure 4).



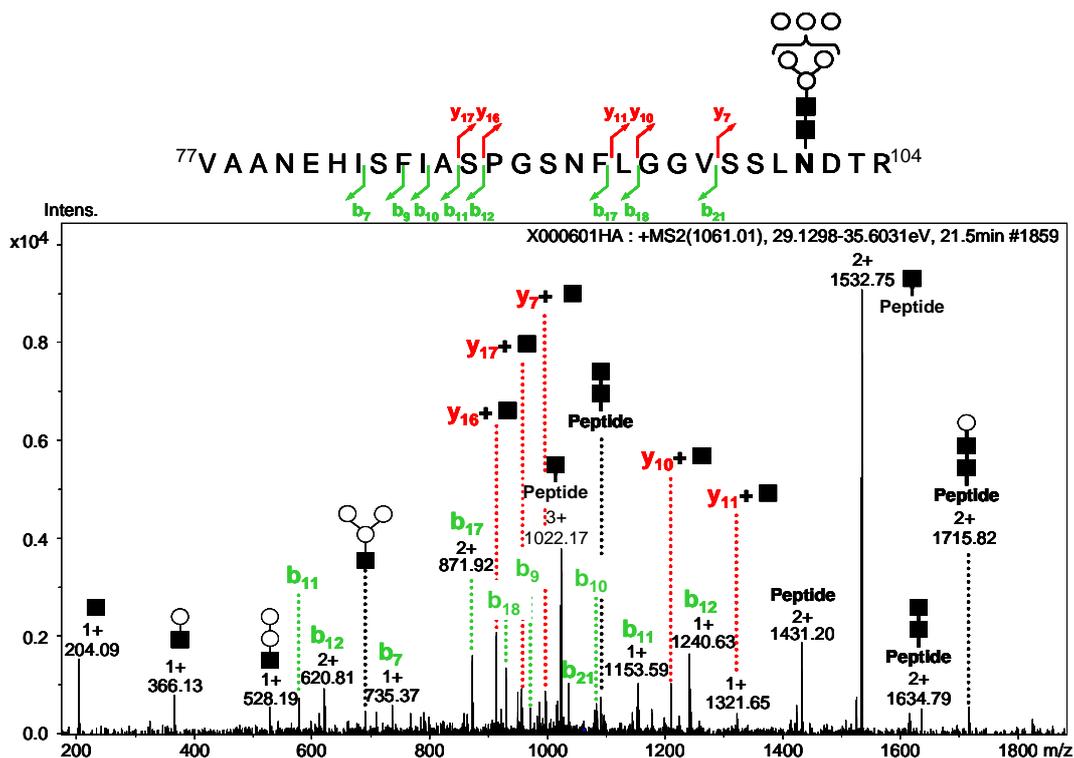
**Figure 4.** (A) Chromatogramme correspondant aux ions diagnostiques extrait du signal correspondant à la haute énergie (28 eV) dans la cellule de collision lors de l'analyse nanoLC-MS en mode scan alterné du digestat chymotrypsique de la glycoprotéine GAP50. (B) Chromatogramme des ions diagnostiques extrait du signal MS/MS de l'analyse nanoLC-MS/MS du digestat chymotrypsique de la glycoprotéine GAP50

### 3.3.3. Analyse nanoLC-MS/MS : caractérisation des glycopeptides

Lors de l'étude précédente, une trappe ionique avait été utilisée pour effectuer les analyses nanoLC-MS/MS. Les trappes ioniques sont des instruments sensibles et permettant d'obtenir des spectres de fragmentation de très bonne qualité. Cependant, nous avons vu dans le Chapitre 2, que ce type de spectromètre de masse n'est pas très adapté pour l'analyse des glycopeptides. Les deux digestats (trypsique et chymotrypsique) ont donc été analysés par nanoLC-MS/MS sur un Q-TOF (maXis, Bruker Daltonics) afin de caractériser les glycopeptides de GAP50.

L'utilisation d'un Q-TOF maXis permet d'obtenir une meilleure résolution qu'une trappe ionique (30000 à 35000 sur les glycopeptides) et une meilleure précision de mesure de masse sur les ions parents (5 ppm) mais aussi sur les ions fragments ( $\pm 0,02$  Da). Ces performances permettent de déterminer précisément la masse des glycopeptides et des fragments et ainsi de limiter le champ des possibilités pour l'interprétation des spectres MS/MS.

Les spectres correspondant à la fragmentation de glycopeptides ont été interprétés manuellement en assignant les ions correspondants aux fragments les plus courants : les fragments internes du glycane et les pertes de neutre de sucres. Des ions correspondant à des fragments internes du peptide ont également été observés pour certains glycopeptides (Figure 5).



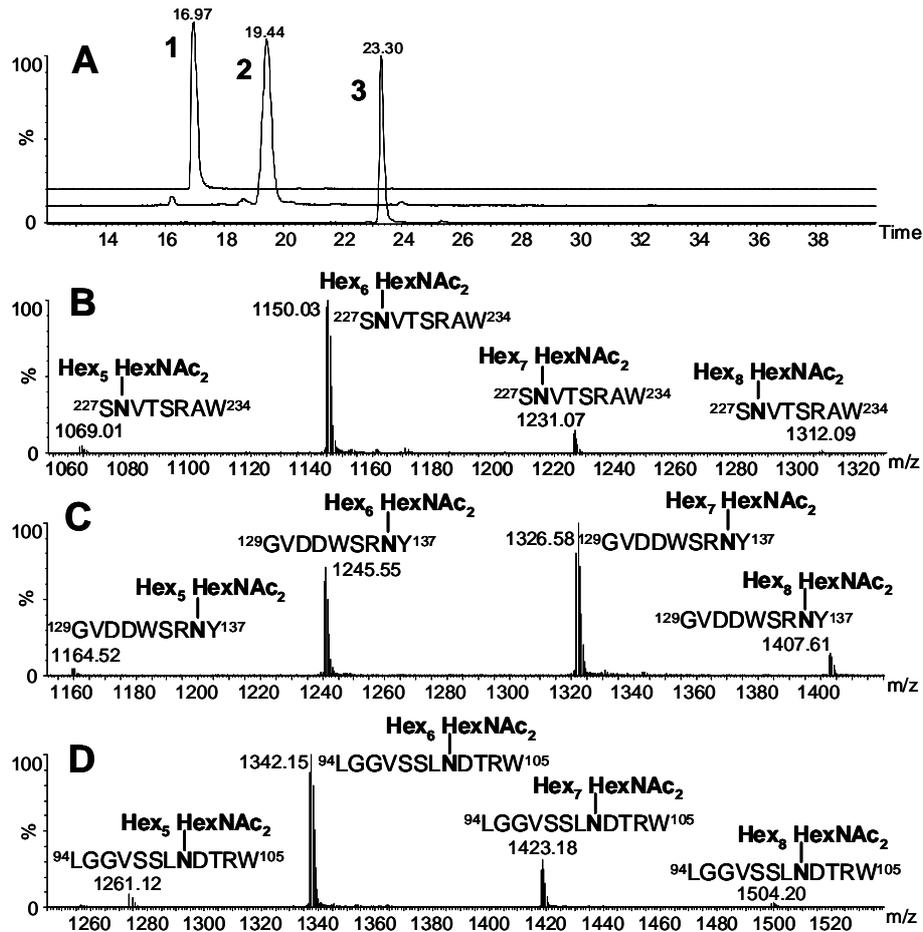
**Figure 5.** Spectre MS/MS (MaXis, Bruker Daltonics) de l'ion  $(M+4H)^{4+}$  1061,01 Da d'un glycopeptide tryptique de GAP50 glycosylé en Asn<sup>101</sup> (Hex<sub>6</sub> HexNAc<sub>2</sub>). Les ions b et y correspondent à des fragments internes du peptide

### 3.3.4. Résultats

L'ensemble de ces expériences a permis de caractériser les trois sites de glycosylation de GAP50 et d'identifier plusieurs glycoformes au niveau de chaque site :

- Asn<sup>101</sup> : Hex<sub>5-6-7</sub>HexNAc<sub>2</sub>
- Asn<sup>136</sup> : Hex<sub>6-7-8</sub>HexNAc<sub>2</sub>
- Asn<sup>228</sup> : Hex<sub>6-7</sub>HexNAc<sub>2</sub>

Une analyse LC-MS complémentaire a permis de montrer qu'il existait également des glycoformes mineures Hex<sub>5</sub>HexNAc<sub>2</sub> pour les sites Asn<sup>136</sup> et Asn<sup>228</sup>, mais aussi Hex<sub>8</sub>HexNAc<sub>2</sub> au niveau des sites Asn<sup>101</sup> et Asn<sup>228</sup> (Figure 6). Les ions de ces glycoformes ne sont pas assez intenses pour être sélectionnés pour la MS/MS.



**Figure 6.** Analyse du digestat chymotrypsique de la bande 1 par LC-MS; (A) Les chromatogrammes extraits pour les ions  $m/z$  1150,0, 1231,0, 1245,5, 1326,5, 1407,5, 1261,0, 1342,1, 1423,1,  $\pm 0.2$  Da, indiquent les zones d'élution des glycopeptides. (B, C et D) Somme des spectres de masses obtenus dans les zones d'élution des glycopeptides. (B) Site Asn<sup>228</sup>, élués à 17 min (pic 1). (C) Site Asn<sup>136</sup>, élués à 19 min (pic 2). (D) Site Asn<sup>101</sup>, élués à 23 min (pic 3)

En combinant les spectres de masses de chaque pic chromatographique, on obtient un profil global des glycopeptides pour chacun des sites de glycosylation. La distribution des glycopeptides montre une prédominance du glycane Hex<sub>6</sub>HexNAc<sub>2</sub>, à l'exception d'Asn<sup>136</sup> pour qui Hex<sub>7</sub>HexNAc<sub>2</sub> est majoritaire. Les tendances obtenues sont similaires à celle observée par nos collaborateurs lors de l'étude sur les glycanes libérés après déglycosylation de GAP50. Cependant, l'étude des glycanes libérés ne permet pas d'observer la distribution des glycosylations au niveau de chaque site, contrairement à notre approche qui consiste à analyser les glycopeptides intacts.

Au total, 4 glycoformes ont été identifiées pour chaque site de glycosylation : Hex<sub>5</sub>HexNAc<sub>2</sub>, Hex<sub>6</sub>HexNAc<sub>2</sub>, Hex<sub>7</sub>HexNAc<sub>2</sub> et Hex<sub>8</sub>HexNAc<sub>2</sub>.

### 3.4. Conclusion et perspectives

Alors que l'étude précédente (§ 2) n'avait permis d'identifier qu'un seul site de glycosylation (Asn<sup>136</sup>) de la glycoprotéine GAP50, la méthodologie que j'ai développée et qui a été appliquée dans cette étude a permis d'identifier les trois sites de glycosylations (Asn<sup>101</sup>, Asn<sup>136</sup> et Asn<sup>228</sup>) et la présence de 4 chaînes glycaniques par site Hex<sub>5-8</sub>HexNAc<sub>2</sub>.

La digestion de GAP50 par la chymotrypsine a généré des glycopeptides ayant des tailles plus homogènes ce qui permet d'optimiser une méthode de masse adaptée pour l'ensemble des glycopeptides. L'utilisation d'un analyseur de type Q-TOF qui possède une excellente résolution spectrale et précision de mesure de masse, a été déterminante pour l'interprétation et la validation des spectres de fragmentation des glycopeptides. De plus, ce type d'instrument est beaucoup moins limité en gamme de masse que les trappes ioniques. Ainsi, un glycopeptide tryptique de haut poids moléculaire a pu être caractérisé alors qu'il n'avait pas pu être observé lors de l'étude précédente.

L'identification de ces glycoformes va aider nos collaborateurs à déterminer le rôle biologique de chacune de ces glycosylations et ainsi de mieux comprendre les mécanismes moléculaires du gliedosome. Dans cet objectif, la caractérisation des autres glycoprotéines identifiées au cours de l'analyse protéomique globale réalisée précédemment doit être poursuivie à l'aide de la méthodologie que nous avons mis en place. Ainsi, la compréhension du rôle de ces protéines et de leurs glycosylations pourrait conduire au développement de nouvelles cibles thérapeutiques.

## **4. Etude des facteurs nucléaires impliqués dans la régulation de l'expression des gènes de *T. gondii***

### **4.1. Contexte de l'étude**

La capacité de *T. gondii* à passer de la forme tachyzoïte qui est infectieuse à la forme bradyzoïte qui est dormante rend l'infection par ce parasite incurable. En effet, la forme infectieuse tachyzoïte se multiplie très rapidement mais est vulnérable. En revanche, une fois convertie en bradyzoïte et enkysté dans les tissus de l'hôte, le parasite devient résistant aux attaques du système immunitaire mais aussi aux médicaments [Kim et Weiss 2004]. Ainsi, la forme avirulente enkystée persiste toute la vie de l'individu sans aucun symptôme. Cependant, en cas de défaut du système immunitaire, le parasite peut reprendre sa forme virulente tachyzoïte et peut être responsable d'encéphalites mortelles comme chez les malades du SIDA par exemple.

De précédents travaux ont montré que les gènes exprimés pour les deux stades (virulent et dormant) sont différents [Dzierszinski et al. 2001, Li et al. 2003, Kibe et al. 2005]. Bien qu'il soit établi que le contrôle de la régulation de la transcription soit un facteur clef dans les étapes du développement et de la virulence du parasite [Gissot et al. 2009], les protéines qui contrôlent ce mécanisme n'ont toujours pas été identifiées.

Le but de cette étude a été d'identifier les facteurs de transcription impliqués dans la régulation de l'expression des gènes de *T. gondii* pour tenter de mieux comprendre les mécanismes d'interconversion de la forme virulente à la forme dormante et ainsi pouvoir développer des solutions thérapeutiques. La préparation et la purification de ces facteurs de transcription est délicate à réaliser. Ainsi peu de protéines sont disponibles, ce qui nécessite l'amélioration de la sensibilité des méthodes protéomiques utilisées. Ce travail fait l'objet d'une publication soumise présentée en annexe 2.

### **4.2. Méthodologie**

#### **4.2.1. Préparation d'échantillon**

Les facteurs nucléaires provenant de la forme virulente tachyzoïte ont été purifiés par chromatographie d'affinité avant d'être séparés par un gel d'électrophorèse monodimensionnel SDS-PAGE. Les protéines ont ensuite été révélées à l'aide d'une coloration à l'argent. L'ensemble des méthodes de préparation et des techniques de biologie moléculaires utilisé par nos collaborateurs est détaillé dans la partie matériel et méthode de la publication en annexe 2.

#### 4.2.2. Analyse protéomique

Après découpage des bandes, les protéines ont été digérées dans le gel avec de la trypsine et analysées par nanoLC-MS/MS. Deux méthodes différentes ont été utilisées pour l'interprétation des données. Au début de l'étude, le génome de *T. gondii* était séquencé mais en cours d'annotation. Par conséquent, il existait peu de séquences protéiques correspondant à cet organisme dans les banques. C'est pourquoi nous avons tout d'abord adopté une stratégie de recherche directe dans le génome (§ 2.2.2 et annexe 1). Par la suite, le génome du parasite étant complètement annoté, les analyses ont été réinterprétées en utilisant une banque protéique target-decoy restreinte aux Apicomplexa extraite de NCBIInr, et en utilisant deux moteurs de recherche, Mascot et OMSSA. Le but de cette seconde interprétation étant de confirmer les résultats obtenus par recherche dans le génome suivi de MS-Blast, les critères de validation des interprétations ont été très strictes : 2 peptides minimum, ayant un score Mascot supérieur à "l'identity score" ou une "E-value" dans OMSSA inférieure à 2. L'ensemble des paramètres chromatographiques, du spectromètre de masse et d'interprétation des données est détaillé dans la partie matériel et méthode de la publication en annexe 2.

#### 4.3. Résultats

Il est très difficile d'obtenir des protéines nucléaires à partir des formes dormantes bradyzoïtes qui se répliquent peu, c'est pourquoi les analyses protéomiques ont été effectuées sur des protéines provenant de *T. gondii* sous sa forme virulente tachyzoïte. Cependant, les facteurs nucléaires restent difficiles à purifier. Les analyses ont donc été effectuées sur des échantillons en très petites quantités, ce qui a nécessité l'application d'une méthodologie protéomique adaptée.

De plus, lorsque le projet a débuté, très peu de séquences protéiques de *T. gondii* étaient disponibles dans les banques de données. C'est pourquoi, les premières interprétations ont été effectuées par recherche dans le génome complet non annoté du parasite. Par la suite le génome de *T. gondii* a fini d'être annoté et les recherches ont été refaites dans une banque de données restreinte target-decoy en utilisant deux moteurs de recherche afin de confirmer les résultats. Parmi les protéines identifiées on trouve des protéines non spécifiques (ribosomales, rhopty) issues de la purification, et 39 facteurs nucléaires qui peuvent être classés en trois catégories :

- 11 protéines présentant une homologie avec des facteurs nucléaires connus;
- 7 protéines correspondant à des kinases, des phosphatases acides et des protéines heat shock;
- 21 protéines étant des "hypothetical protein" dont plus de la moitié n'a aucune homologie de séquence avec des facteurs nucléaires connus.

De cette liste de protéines identifiées par analyse protéomique, sept ont été sélectionnées pour des études biologiques plus approfondies et sont nommées TgNFs pour "*T. gondii* nuclear factor". Des expériences de marquage fluorimétrique ont permis de montrer que six protéines sur sept sont

présentes dans le noyau du parasite. TgNF1 et TgNF4 ne possèdent pas d'homologie avec des facteurs nucléaires connus et TgNF2, TgNF5 et TgNF6 sont probablement impliqués dans le métabolisme de l'ARN. Les expériences de biologie moléculaire ont alors été focalisées sur la protéine TgNF3 qui possède une forte homologie avec un facteur de transcription connu chez *Saccharomyces pombe* et qui est impliqué dans la répression de l'ADN ribosomique [Kuzuhara et Horikoshi 2004].

Les expériences réalisées par nos collaborateurs ont permis de montrer que pour la forme virulente tachyzoïte, TgNF3 est présent majoritairement dans le nucléole, alors que pour la forme dormante bradyzoïte, ce facteur de transcription est localisé dans le cytoplasme. Au cours de la phase infectieuse et de la multiplication du parasite, le taux de TgNF3 varie en fonction du stade de la division cellulaire. Nos collaborateurs ont également généré un mutant de *T. gondii* dans lequel, TgNF3 est surexprimé. Des expériences ont montré que lorsqu'une souris est infectée par ce mutant, la virulence du parasite et la formation de kystes dans le cerveau sont fortement atténuées. Ainsi, 100% des souris infectées par les mutants survivent et ne présentent aucun symptôme lorsqu'elles sont réinfectées par une souche sauvage.

#### 4.4. Discussion et conclusion

L'application de la stratégie protéomique développée dans la première partie de ce manuscrit a permis d'identifier plusieurs facteurs nucléaires présents en très faibles quantités.

L'interprétation des données en utilisant une banque restreinte target-decoy et deux moteurs de recherche a permis non seulement d'identifier plus de protéines (103 au lieu de 76) mais aussi de confirmer les résultats obtenus par recherche dans le génome complet non annoté de *T. gondii*. En effet, la méthode qui permet d'identifier les protéines par recherche dans le génome n'est pas automatisée et utilise la recherche par homologie de séquence (MS-Blast). Ainsi, bien que la validation des identifications soit très stricte, elle reste manuelle et donc subjective.

Actuellement, la validation des interprétations tient un rôle central dans les analyses protéomiques [Nesvizhskii et al. 2007, Deutsch et al. 2008], et l'utilisation combinée de banques restreintes target-decoy [Balgley et al. 2007, Elias et Gygi 2007] mais aussi de plusieurs moteurs de recherches [Kapp et al. 2005, Searle et al. 2008, Jones et al. 2009] tend à se généraliser [Carrascal et al. 2010]. En effet, cette technique permet d'améliorer la confiance dans la validation des résultats et de calculer un taux de faux positifs. Ainsi, grâce à cette méthodologie, nous avons pu fournir à nos collaborateurs une liste de protéines candidates pour des études de biologie moléculaire.

Les travaux de protéomique et de biologie moléculaire ont permis de proposer un mécanisme moléculaire pour TgNF3. Ce facteur de transcription pourrait avoir un rôle dans le contrôle de la transcription des gènes qui régulent, *in vivo*, la virulence du parasite mais aussi dans sa différenciation et dans la formation des kystes. Ces résultats, très encourageants, constituent une piste pour le développement de vaccins contre *T. gondii*.



## **ANNEXE 2: Publication**

**A novel nuclear factor TgNF3 of *Toxoplasma gondii* is a dynamic modulator of nucleolar architecture and parasite virulence**

Alejandro Olguin-Lama, Edwige Madec, Agnès Hovasse, Elisabeth Werkmeister, Isabelle Callebaut, Christian Slomianny, Stéphane Delhay, Thomas Mouveaux, Christine Schaeffer-Reiss, Alain Van Dorsselaer and Stanislas Tomavo

Soumis au journal Plos Pathogens



# **A novel *Toxoplasma gondii* nuclear factor TgNF3 is a dynamic chromatin-associated component, modulator of nucleolar architecture and parasite virulence**

Alejandro Olguin-Lama<sup>1,2</sup>, Edwige Madec<sup>2</sup>, Agnès Hovasse<sup>3</sup>, Elisabeth Werkmeister<sup>1</sup>, Isabelle Callebaut<sup>5</sup>, Christian Slomianny<sup>4</sup>, Stephane Delhayé<sup>1,2</sup>, Thomas Mouveaux<sup>1,2</sup>, Christine Schaeffer-Reiss<sup>3</sup>, Alain Van Dorsseleer<sup>3</sup> and Stanislas Tomavo<sup>1,2,6</sup>

<sup>1</sup>Center for Infection and Immunity of Lille, CNRS UMR 8204, INSERM U 1019, Institut Pasteur de Lille, Université Lille Nord de France, 59000 Lille, France

<sup>2</sup>Centre National de la Recherche Scientifique, CNRS UMR 8576, UGSF, Université de Lille 1, 59655 Villeneuve d'Ascq, France

<sup>3</sup>Laboratoire de Spectrométrie de Masse Bioorganique, IPHC, CNRS UMR 7178, Université de Strasbourg, 67087 Strasbourg, France

<sup>4</sup>Laboratoire de Physiologie Cellulaire, INSERM U1003, Université de Lille 1, 59655 Villeneuve d'Ascq, France

<sup>5</sup>Centre National de la Recherche Scientifique, Universités Pierre et Marie Curie-Paris 6 et Denis Diderot-Paris 7, UMR7590, 75005 Paris, France

<sup>6</sup>Corresponding author: Stan.Tomavo@pasteur-lille.fr

## ABSTRACT

In *Toxoplasma gondii*, *cis*-acting elements present in promoter sequences of genes that are stage-specifically regulated have been described. However, the nuclear factors that bind to these *cis*-acting elements and regulate promoter activities have not been identified. In the present study, we performed affinity purification, followed by proteomic analysis to identify nuclear factors that bind to a stage-specific promoter in *T. gondii*. This led to the identification of several putative nuclear factors in *T. gondii* including a novel factor, designated herein as TgNF3. The N-terminal domain of TgNF3 shares similarities with the N-terminus of yeast nuclear FK506-binding protein (FKBP), known as a histone chaperone regulating rDNA silencing. Using anti-TgNF3 antibodies and yellow fluorescence protein-tagged TgNF3, we show that TgNF3 is predominantly a parasite nucleolar, chromatin-associated protein that binds specifically to *T. gondii* promoters *in vivo*. The over-expression of TgNF3 in the tachyzoites revealed dynamic changes in the size and position of the nucleolus, leading to a severe attenuation of virulence *in vivo*. We demonstrate that TgNF3 physically interacts with H3, H4 and H2A/H2B assembled into bona fide core and nucleosome-associated histones. Furthermore, TgNF3 is present in parasite nucleosomes and interacts specifically to histones only in the context of stage-specific gene silencing of a promoter that lacks active epigenetic acetylated histone marks. In contrast to virulent tachyzoites, which express the majority of TgNF3 in the nucleolus, the protein is exclusively located in the cytoplasm of the avirulent and dormant encysted bradyzoites. We propose a model where TgNF3 acts essentially to coordinate nucleolus and nuclear functions by modulating nucleosome activities during the intracellular proliferation of the virulent tachyzoites of *T. gondii*.

**Running title: A novel *Toxoplasma* chromatin-associated factor**

## Author Summary

Apicomplexa including *Toxoplasma gondii* are responsible for a variety of deadly infections. These protozoan intracellular parasites have complex life cycles within different hosts and their infectivity relies on their capacity to regulate gene expression in response to different environments. However, to date almost nothing is known about nuclear factors that regulate gene expression in these pathogens. Initial studies suggest that transcriptional control is a primary means to regulate genes involved in developmental transitions of *T. gondii*. We have combined proteomic analyses and reverse genetics to characterize parasite nuclear factors that bind to a stage-specific promoter. This led to the identification of several putative nuclear factors in *T. gondii* including a novel factor, designated herein as TgNF3. The N-terminal domain of TgNF3 shares similarities with the N-terminus of yeast nuclear FK506-binding protein (FKBP), known as a histone chaperone regulating rDNA silencing. We show that TgNF3 is predominantly a nucleolar, chromatin-associated protein that specifically binds to *T. gondii* nucleosome-associated histones and promoters. Over-expression of TgNF3 induces dynamic changes in size and position of the nucleolus, and a severe attenuation of parasite virulence *in vivo*. In avirulent encysted bradyzoites, TgNF3 is found exclusively in the cytoplasm, suggesting a potential role for TgNF3 in regulating nucleolar and nuclear functions during the intracellular proliferation of the virulent tachyzoites of *T. gondii*.

## INTRODUCTION

*Toxoplasma gondii* has long been a major medical and veterinary problem capable of causing abortion, or congenital birth defects in both humans and livestock. The advent of AIDS has drawn even more attention to *T. gondii* as a serious opportunistic pathogen. *T. gondii* is distinct from nearly all of the other members of the phylum Apicomplexa, owing to the exceptional range of all warm-blooded animals and humans that serve as hosts. The infection is incurable because of its ability to differentiate from the rapidly replicating tachyzoite stages into latent cysts containing the bradyzoite stages that are impervious to immunity and current drugs. *T. gondii* cysts and dormant bradyzoites persist in the brain of the infected host and also play key roles in pathogenesis because they can convert to virulent tachyzoites in immune compromised individuals with AIDS and in transplant patients. This stage conversion is triggered by the host immune response and impairment of the immune system in HIV infected individuals can lead to lethal toxoplasmic encephalitis.

Although the basal core transcriptional machinery, the protein-coding genes involved in nucleosome assembly and chromatin remodelling machinery were found to be conserved in *T. gondii* genome (<http://www.toxodb.org>), a surprising finding was the identification of a relatively low number of genes encoding transcription factors in the parasite [1-6]. This has led to the proposal that gene regulation in *T. gondii* and other apicomplexan parasites is controlled mainly by epigenetic mechanisms [7-9]. However, bioinformatics searches for DNA-binding domains identified, in *Plasmodium spp* and in all apicomplexan parasite genomes sequenced to date, a family of proteins homologous to the plant transcription factor Apetala2, named ApiAP2 for apicomplexan AP2-like factors [10]. De Silva *et al.* have demonstrated the DNA-binding specificities of two ApiAP2 proteins that have a high specificity for unique DNA sequence motifs found in the upstream regions of distinct sets of genes co-regulated during asexual development [11]. One *Plasmodium* ApiAP2 factor has an important role in stage-specific gene regulation by activating a set of genes, including genes reported to be required for midgut invasion. It has also been described that this ApiAP2 factor binds to specific six-base sequences in the proximal promoters [12]. In *T. gondii*, our current transcriptome knowledge indicates that mRNA pools are dynamic and transcriptional control is also a primary means to regulate the developmental transitions of the parasites, suggesting that gene regulation occurs mostly at the transcriptional level [13, 14]. Microarray studies have demonstrated that transcriptional regulation required timed expression of clusters of genes during the bradyzoite development and that for most genes changes in transcription are tied to modulations in protein expression [15-18]. Further confirmatory data is provided by the Serial Analysis Gene Expression (SAGE), which supports the notion that transcriptional regulation plays a key role in the developmental program of *T. gondii* [19]. We and others have previously established that *T. gondii* stage conversion is accompanied by the expression of

a variety of genes that displayed diverse functions, suggesting that parasite differentiation is clearly regulated in part at the transcriptional level [20-28]. In addition, several promoter sequences have been characterized in *T. gondii* [29, 30]. We have shown that the promoter regions of two stage-specifically expressed genes displayed promoter autonomy that can be exploited to achieve developmental expression of reporter genes [31]. Yet almost nothing is known about the nature of nuclear factors that can specifically bind to *T. gondii* promoters and regulate transcriptional activity. Here, we report the isolation and characterization of a novel *T. gondii* promoter-specific binding factor designated herein as TgNF3, which shares similarities with yeast nuclear FK506-binding protein (FKBP), known to be a histone chaperone regulating rDNA silencing. We demonstrate that TgNF3 protein is predominantly a nucleolar, chromatin-associated protein that binds specifically to *T. gondii* gene promoters *in vivo*. Importantly, TgNF3 interacts directly to core and nucleosome-associated histones in the context of gene silencing. Furthermore, we show that TgNF3 is a dynamic chromatin-associated factor, a modulator of nucleolus biogenesis, parasite replication and virulence. Taken together, our findings suggest an important role of TgNF3 in nucleosome activity that may regulate nucleolar and nuclear functions during the intracellular proliferation of *T. gondii*.

## RESULTS

**Isolation of nuclear factors that bind to *T. gondii* promoter.** To test the suitability of using a stage-specific promoter to purify and determine the identity of nuclear factors that interact and control the transcriptional repression of a *T. gondii* stage-specific promoter, we performed affinity chromatography using biotinylated DNA sequence (Figure S1) from the previously reported bradyzoite-specific *ENO1* promoter [31] used as bait (experimental strategy outlined in Figure 1A). After biotinylation, we checked whether the probe still binds to parasite nuclear factors. As shown in Figure 1B, the biotinylated probe strongly interacts with nuclear factors in gel retardation. The specificity of the DNA-protein complexes visualized (lane 2) was demonstrated by a competition assay using unlabelled probe (lane 3), confirming the presence of bound parasite nuclear factors. To determine the nature of these nuclear factors, large scale affinity purification was carried out using the biotinylated bait incubated with a nuclear extract containing about 17 mg of total nuclear proteins obtained from  $4 \times 10^{10}$  tachyzoites. Thirty five nuclear proteins were identified and isolated after SDS-PAGE and silver staining (Figure 1C). These proteins were excised as gel slices and subjected to proteomic analysis. Using mass spectrometry and database searches, we identified thirty nine putative nuclear factors, which are presented as three groups in Supplementary Table 1 (Table S1): 1) A class of 11 proteins, which displayed significant similarities to known nuclear factors. These include a protein possessing a RNA-specific DEAD/DEAH box helicase domain (genbank identifier (gi) number

211966692) and a protein having a pinin domain (genbank identifier (gi) number 211966969). Members of the pinin family have various localisations (including nuclear location) within eukaryotic cells and are thought to regulate protein-protein interactions. A protein (gi number 211962881), which shares similarities with a nuclear FK506-binding protein (*Spodoptera frugiperda* FKBP46), is also found in this group. The FKBP46 homologue in *Saccharomyces pombe*, SpFKBP39, was reported to be involved in transcription repression of ribosomal DNA [32]. Other nucleolar factors and nuclear proteins containing DNA-binding motifs were also present. 2) A second class of 7 proteins corresponds to kinases, acid phosphatase and heat shock proteins. 3) Finally, 21 hypothetical proteins, more than half of the total number of factors discovered, displayed no obvious similarity to known factors. However, our further bioinformatic analyses identify two proteins in this group (genbank identification (gi) numbers 211967631 and 211968320 in Table S1) as homologues of Alba, ancient archaeal chromatin-associated factors [our unpublished data]. The Alba factors are known to be involved in gene silencing operating through chromatin regulation in Archaea [33, 34]. Several proteins identified during the proteomic analysis could clearly be expected to be present in the nucleus of the parasite, for example those with strong similarities to known nucleolar factors. However, the majority of enzymes and hypothetical proteins identified could not be obviously considered as genuine parasite nuclear factors. Therefore, we decided to verify whether some of these factors are truly localized in the nucleus of *T. gondii*.

**Nuclear localization of candidate factors identified by proteomic analysis.** We have randomly chosen 7 candidate proteins and used two distinct tagged constructs to examine the presence of these factors in the parasite's nucleus. The full-length cDNAs of these candidate proteins were fused to YFP or HAFLAG tags and their expression was driven by tubulin (Tub 1) and dense granule 1 (GRA1) promoters, respectively. We compared the location of YFP tagged proteins relative to ENO2, a glycolytic enzyme known to be predominantly detected in the nucleoplasm but not in the nucleolus of active replicated intracellular tachyzoites [35]. Figure 2A illustrates four YFP-tagged candidate proteins with convincing nuclear location assessed by direct fluorescence microscopy. These candidate proteins were therefore named TgNFs for *T. gondii* nuclear factors. TgNF1 (gi number 211966692) and TgNF2 (gi number 211966969), containing a DEAD/DEAH box helicase domain and a pinin domain, respectively, showed perfect overlapping fluorescence signals with anti-ENO2 staining, suggesting that these two factors localized in the parasite nucleoplasm (Figure 2A). In contrast, two distinct patterns of fluorescence were observed for TgNF3 (gi number 211962881) and TgNF4 (gi number 211965453) with strong fluorescence in the parasite nucleolus in addition to a faint signal, which co-localized with ENO2 signal in the nucleoplasm (Figure 2A). In a similar approach, we confirmed the location of TgNF1 and TgNF2 in the parasite nucleoplasm (Figure 2B). Both diffuse

nucleoplasm and strong nucleolar patterns were confirmed for TgNF3 and TgNF4 in transgenic HAFLAG-tagged protein, whose expression was driven by another promoter (Figure 2B). Furthermore, the two homologues (gi numbers 211967631 and 211968320 in Supplementary Table S1) of ancient archaeal chromatin-associated factors named Alba1 and Alba2 [33, 34], and here designated TgNF5 and TgNF6, displayed fluorescence signals on the nucleus periphery. In addition, a signal was observed in the parasite cytoplasm, suggesting that these Alba homologues may have regulatory functions in both nuclear and cytoplasmic compartments as previously reported [36]. One candidate protein (TgNF7) with no known similarity (gi number 211964891, Table S1) showed a predominantly cytoplasmic signal even though the corresponding recombinant protein produced in *E. coli*, specifically binds to the ENO1 promoter by gel retardation. Furthermore, chromatin immunoprecipitation (ChIP) using antibodies specific to TgNF7 also identified the ENO1 promoter suggesting that a tiny amount of this candidate factor may be present in the parasite nucleus *in vivo*. We conclude that six out of the seven candidate proteins have been shown to enter the nucleus of *T. gondii*, which represent 85% of proteins experimentally tested. Because one candidate factor TgNF4 has no obvious similarity to known nuclear factors and four other candidates TgNF1, TgNF2, TgNF5 and TgNF6 are also suspected to have functions in RNA metabolism [36, 37], we decided to further investigate in detail the biological functions of TgNF3, which shares similarities with nuclear FK506 binding proteins (FKBP), known in yeast as a histone chaperone that regulates rDNA silencing [32].

**TgNF3 belongs to a large superfamily of nuclear chaperones, including insect and fungi FKBP, plant HD2 histone deacetylases and members of the nucleoplasm family.** Because TgNF3 was selected for detailed molecular and functional characterization, we wanted to determine definitively whether this putative homologue of fungi, nuclear FK506-binding protein is a genuine member of this family. This is important to further elucidate potential nucleolar/nuclear functions of TgNF3 in the parasite. A first analysis of the TgNF3 protein sequence using Hydrophobic Cluster Analysis (HCA) [38] indicated that it contains two globular domains (boxes, Figure 3A), separated by a linker sequence, rich in acidic residues (unboxed area, Figure 3A). A PSI-BLAST search using TgNF3 first domain as query (aa 1-100) indicated a significant similarity with the histone deacetylase 2 (HD2)/nuclear FK506-binding protein (FKBP) family, which was first reported by Aravind & Koonin [39]. The N-terminal domain of nuclear members of the FKBP family found in yeasts and insects has nucleosome assembly activity, which is independent from the activity of their C-terminal FKBP domain, having peptidyl-prolyl isomerase (PPIase) activity [32]. This N-terminal domain shares significant similarities with the N-terminal domain of HD2 proteins, which are described as plant-specific histone deacetylases (HDACs) [39]. The HD2/nuclear FKBP family, thus combining the N-terminal domains of plant HD2 and nuclear FKBP found in fungi and insects, also includes the N-

terminal domains of parasitic apicomplexan proteins, among which is TgNF3. Interestingly, no nuclear factors of vertebrates are members of this family [39]. However, after further PSI-BLAST iterations, we found significant similarities of the HD2/nuclear FKBP family with the N-terminal (Np) core domain of the nucleoplasmin/nucleophosmin (NPM) family, which are nuclear chaperones from vertebrates involved in chromatin remodeling ([40] and Figure 3B). Reciprocal searches using sequences from the NPM family also highlighted the similarity with the whole HD2/nuclear FKBP family. The N-terminal domains of the HD2/nuclears FKBP and nucleoplasmin/nucleophosmin (NPM) families form a unique large structural superfamily, sharing a common ability to bind histones. For all members of the family, the N-terminal Np core is followed by acidic stretches, thought to play an important role in histone binding (Figure 3B), but only in few cases, a C-terminal globular domain follows this acidic stretches as for nuclear FKBP in which this domain has peptidyl-prolyl isomerase (PPIase) activity. For TgNF3, which lacks the peptidyl-prolyl isomerase domain, a novel small domain consisting of  $\alpha$ -helices is also present (Figure 3A), but this domain shares no obvious similarity with any other known functional domain.

The relationship between the HD2 and NPM families has been previously mentioned [41] but to our knowledge, without a detailed description of the sequence similarities. We thus presented in this study a refined alignment between these two families, which was performed using sensitive tools (Figure 3B). Although sequence identities are low (below 20 %), the hydrophobic amino acids are well conserved, supporting a common fold for the two families. In addition, we provided a comprehensive analysis of this alignment, using the structural information coming from crystal structures of NPM family members. These have revealed that the Np core adopts an eight-stranded beta-barrel structure and organizes itself into pentameric or decameric structures [42-44]. These decameric structures (dimers of pentamers) appeared to have direct relevance to histone binding and models have been proposed in which histone octamers dock around the NPM decamer periphery, the binding especially involving an acidic stretch (NPM A1 tract) and a signature  $\beta$ -hairpin of the Np core (pink and orange in Figure 3B and 4A). The examination of our alignment showed that the NPM-specific acidic A1 tract is not present in the HD2/nuclear FKBP. Instead, acidic stretches of variable length within the  $\beta$ -hairpin linking strands  $\beta$ 4 and  $\beta$ 5 are present in the nuclear FKBP (FKBP acidic tract, Figure 3B). This insertion would be located in close proximity to the loop integrating the NPM A1 tract (Figure 4A). Following the HD2 N-terminal domain alignment, it has been suggested that the two conserved polar residues, namely an invariant aspartic acid and a histidine (arrows on Figure 3B and 4B), may play a key role in lysine deacetylation [39], a prediction which was partly supported by experimental investigations [45]. We show for the first time and based on the NPM template, that these two residues

are located in close proximity of the NPM Np core (blue and red in Figure 4A), at the end of the funnel shaped cavity formed by the different subunits, and on the subunit distal face (thus opposite to the pentamer-pentamer interface). The two residues are also close to the positions of the acidic stretches and of the  $\beta$ -hairpin. We hypothesize from this 3D mapping that the binding of histones on the NPM/HD2 scaffold may not involve the decamer periphery as previously suggested [42-45]. Rather the distal ends with the pentameric organization may be important for histone binding. However, our analyses also revealed that even if the SpFKBP39 and TgNF3 proteins have a common "NPM/HD2" N-terminal domain, they differ in their C-terminal extremities, which is highlighted in Figure 4B. The C-terminal globular domain of *T. gondii* TgNF3, which is specific of apicomplexans, has no similarity with the C-terminal propyl peptidyl isomerase of nuclear FKBP and to other known proteins. Therefore, we embarked on the functional characterization of TgNF3 in the parasite.

**TgNF3 gene is transcribed and translated in both virulent tachyzoites and avirulent bradyzoites of *T. gondii*.** Toward the identification of the potential nuclear functions of TgNF3 during intracellular development of the parasite, we analyzed the pattern of expression of TgNF3 gene in the two invasive life stage forms of *T. gondii* present in intermediate hosts. Specific transcript coding TgNF3 were amplified by RT-PCR using total RNA isolated from the rapidly replicating tachyzoites and the dormant encysted bradyzoites (Figure 5A). A comparison of the amount of specific transcripts between virulent tachyzoites and dormant bradyzoites was assessed by RT-PCR using a normalization step with transcript coding the housekeeping actin. Figure 5A showed that TgNF3 mRNA seems to be more abundant in the persistent and dormant bradyzoites. Polyclonal antibodies were also raised against the purified recombinant TgNF3-GST fusion protein. Specific IgGs from the immune sera were affinity purified using recombinant non-fusion TgNF3 protein, which was coupled to Sepharose 4B-CNBr column. Figure 5B shows the specificity of the anti-TgNF3 purified sera tested by Western blots using the recombinant nonfusionTgNF3 (lane 2), GST-TgNF3 fusion expressed in *E. coli* (lane 1), total extract proteins from *T. gondii* tachyzoites (lane 3), from uninfected human fibroblast (lane 4) and from brain cells (lane 5). The sera specifically recognized a single band corresponding to a 43-kDa protein in tachyzoites (lane 3), which is in good agreement with the expected molecular mass of TgNF3. This observation is also supported by the co-migration between the bacterial recombinant non-fusion TgNF3 protein (lane 2) and the native parasite TgNF3 protein (lane 3). Neither anti-GST antibodies nor the pre-immune mice sera reacted with *T. gondii* proteins on Western blots (data not shown). To determine the pattern of expression of TgNF3 in the different life stages of the parasite, cell lysates of tachyzoites and bradyzoites were resolved by SDS-PAGE and probed by Western blots using purified antisera. Figure 5C illustrates the detection of similar levels of TgNF3 protein in tachyzoites and bradyzoites, as immunoblots in parallel with monoclonal antibody anti-*Toxoplasma*

actin provided loading controls. These data suggest that the difference in TgNF3 transcript level does not correlate with the amount of protein detected in the two invasive life stage forms of *T. gondii*. We also transfected parasites with a vector expressing TgNF3 tagged to YFP under the control of the constitutive TUB1 promoter (pTUB1-TgNF3-YFP). These and other expression vectors used in this study are depicted schematically in Figure 2 (on the top) and the stable parasites were selected by chloramphenicol and cloned. Western blot analysis of total protein extracts from transgenic parasites expressing TgNF3 (Figure 5D, arrowheads) confirmed that both native TgNF3 (43-kDa protein) and transgenic TgNF3-YFP fusion protein (~70-kDa protein) were detected with anti-TgNF3 antibodies (lane 7). As anticipated, only TgNF3 was recognized in the wild type parasites (lane 6). The monoclonal antibody specific to GFP recognized only the TgNF3-YFP fusion protein in the transgenic tachyzoites (lane 9) whereas no protein was detected in wild type tachyzoites (lane 8), as expected.

**TgNF3 is a prominent nucleolar factor of the virulent tachyzoites of *T. gondii*.** We next compared the localization of TgNF3 in both parental and over-expressing tachyzoites, when they freshly ruptured from host cells. Figure 6A shows confocal images of extracellular transgenic tachyzoites over-expressing TgNF3-YFP. In contrast to nucleoli with weaker fluorescence signal in the parental tachyzoites stained with the anti-TgNF3 antibodies (Figure 6A, upper panels), the TgNF3-YFP over-expressers contain nucleoli with strong fluorescent signals, which occupied an important proportion of the nuclear volume of the parasites (Figure 6A, lower panels). Z-stacks acquisitions followed by 3D-reconstruction and rotation around the y-axis enabled us to visualize in other sets of extracellular tachyzoites, the strong expression of TgNF3-YFP in nucleoli of over-expressers relative to wild type parasites (Figure 6B). In addition, we performed both direct fluorescence and indirect immunofluorescence assays using wild type and TgNF3-YFP over-expressing tachyzoites that are actively replicating within human fibroblasts *in vitro*. We confirmed that TgNF3 was most prominent in the nuclear areas, which defines *T. gondii* nucleoli (Figure 7A). The indirect immunofluorescence in the upper panel of Figure 7A shows four daughter tachyzoites, which are derived from two divisions of one parental parasite while the lower panel shows the direct fluorescence in the YFP-tagged TgNF3 over-expressers using a heterologous tubulin promoter. This strongly suggests that TgNF3 protein contains itself all the sequence information required for nucleolar retention after being targeted into the nucleus.

**Dynamics of nucleolar TgNF3-YFP expression in intracellular dividing tachyzoites.** Having observed the presence of TgNF3 in the parasite nucleoli, we determined the dynamics of its expression by live imaging using time lapse video-microscopy. The direct fluorescence collected for TgNF3-YFP in the intracellular tachyzoites revealed in some cases the presence of two smaller nucleoli in close vicinity of one larger nucleolus (Figure 7B, blue arrowheads), suggesting a dynamic biogenesis of the

nucleolus in *T. gondii*. To determine the precise location of smaller sized nucleoli in the nucleus, we simultaneously introduced hydroethidine (HET), a live nuclear dye staining during time-lapse imaging. Tracking three sets of independent vacuoles containing actively replicating TgNF3-YFP expressing tachyzoites confirmed two smaller sized and spherical nucleoli that positioned separately in the nucleus stained by HET (Figure 7C, red arrows) although a less intense staining was observed in the entire nuclear areas, which defines the nucleoplasm surrounding the nucleolus (red arrows). In some cases, only one very small and spherical nucleolus was also detected. These two complementary approaches revealed the dynamics of nucleolar biogenesis during the intracellular replication of *T. gondii* tachyzoites *in vitro*, a phenomenon that so far has not been reported. Collectively, these data demonstrate that TgNF3 is predominantly a resident factor even if it is also convincingly detected in the other areas of the nucleus, likely in the nucleoplasm.

**Overexpression of TgNF3 induces an increased nucleolus size and repositioning.** The nucleolar localization of TgNF3 was also demonstrated using electron microscopy and cryo-immunogold labeling performed using intracellular TgNF3-YFP over-expressing tachyzoites and a monoclonal antibody specific to GFP (Figure 8). Using immuno-gold staining and electron microscopy, we also discovered that the expression of the YFP-tagged version of TgNF3 induced profound changes in morphology of the parasite nucleus with a considerable increase of nucleolus size (Figure 8A and 8B). These observations were confirmed in all transgenic extracellular or intracellular tachyzoites, which over-expressed TgNF3-YFP. Using confocal acquisitions that allow 3D-reconstruction of the whole tachyzoite's body, we showed that TgNF3-YFP over-expressers contain nucleoli with a remarkable increase in size (Video S2) relative to that of the parental parasite (Video S1). We conclude that the most direct effect of TgNF3-YFP over-expression in all cases is the presence of huge nucleoli very close to nuclear membrane with a protuberance at one pole of the nucleus, thus deforming the nuclear shape externally (Figure 8B). In contrast, the nucleoli in the wild type parasites are more centrally positioned (Figure 8C). We estimated that over-expression of TgNF3 proteins can induce about 4-5 fold-increase in nucleolar size of transgenic parasites (Figure 8B) relative to that of the wild type (Figure 8C).

**Over-expression of TgNF3 enhances parasite replication *in vitro* but drastically attenuates parasite virulence *in vivo*.** Figure 9A shows that over-expression of TgNF3-YFP in tachyzoites of *T. gondii* induces a faster replication rate relative to the parental parasites. This increased replication rate is not correlated to the elevation of host cell invasion since no significant difference in fibroblast cells or macrophage cells entry was observed between transgenic over-expressers and wild type parasites. In addition, neither stable transgenic tachyzoites over-expressing another nuclear candidate factor TgNF2 nor transgenic tachyzoites over-expressing the nuclear enolases (ENO1 and ENO2) displayed changes

in nucleoli morphology and in replication rate relative to wild type parasites using the same type II 76K strain (our unpublished data). This suggests that the increased replication rate may be specific to TgNF3. Therefore, we conclude that TgNF3 may be involved in nucleolar dynamics and functions that are important for the replication rate of *T. gondii* within the host cell *in vitro*. To assess the influence of stable TgNF3 over-expression *in vivo*, TgNF3 over-expressers and parental wild type parasites were used to inoculate a group of 10 mice at doses up to  $10^4$  tachyzoites. After 4-5 days of infection, all mice from two genetically distinct groups (Balb/c and CBA/J), infected with the parental tachyzoites show the same characteristic symptoms of disease and succumbed approximately 12-14 days (Figure 9B and 9C). Surprisingly, all mice (Balb/c or CBA/J) infected with TgNF3 over-expressing tachyzoites displayed only mild symptoms of disease and recovered faster than those infected with the parental tachyzoites. In this case, 100% of mice survival was obtained with the two groups infected with TgNF3 over-expressers (Figure 9B and 9C). In order to ensure that the surviving mice had truly been infected, immune sera of each group of infected mice were collected 30 days post-inoculation and were tested in Western blots. All mice infected with the TgNF3 over-expressers were positive and they also survived a subsequent challenge with  $10^4$  RH (type I strain) and  $10^5$  76K (type II strain) wild type tachyzoites, doses that confer 100% mortality in the primo-infected mice used as controls (Figure 9B and 9C). These data were confirmed when a group of mice was respectively infected with  $10^4$ ,  $10^5$  and  $10^6$  TgNF3-YFP over-expressing tachyzoites, again no mice succumbed and all reacted positively against *T. gondii* total protein extracts as above. All sera from surviving mice recognized parasite antigens ranging mostly from 15-35 kDa whereas the sera from non-infected mice failed to detect any *T. gondii* antigens, as expected (Figure 9D). These results demonstrated that the mice were truly infected with TgNF3 over-expressing tachyzoites and consequently, positive immune responses have been developed. Therefore, we conclude that TgNF3 over-expression attenuates *T. gondii* virulence in mice and induces live vaccination in mice that can confer protection against *T. gondii* challenge. To examine if the attenuated TgNF3 over-expressers were able to establish a chronic infection in mice, we searched for the presence of cysts using specific staining with the *Dolichos biflorus* lectins, which specifically labeled the cyst wall [46]. The presence of few cysts *per* brain with fluorescently positive cyst-wall stained with the lectin and the presence of YFP-expressing encysted dormant bradyzoites (Figure 10A) was indicative of a chronic infection in all survived mice monitored and analyzed. Movies representing cysts at different depth and created with the Zen software, at a rate of 5 frames per second, confirmed the presence of encysted bradyzoites that expressed YFP-TgNF3 proteins (Video S3). These observations were also validated by direct confocal imaging of lectin-unlabeled cysts by Z-stacks acquisitions enabling the 3D localization of fluorescent signals (Video S4). Interestingly, all cysts analyzed in TgNF3-YFP infected mice, which were challenged with either RH

or 76K parental tachyzoites contained only YFP-positive encysted bradyzoites, suggesting that mice infection with TgNF3-YFP over-expressing tachyzoites raises an efficient sterile protection against *T. gondii*.

**TgNF3 is present exclusively in the cytoplasm of dormant bradyzoites and relocates in the nucleolus during bradyzoite to tachyzoite conversion *in vitro*.** The confocal observations described above confirmed the presence of TgNF3-YFP expression in the encysted bradyzoites during chronic infection of mice. We next determined whether TgNF3-YFP factor actually displayed the dual nucleolar/nuclear localization in the dormant encysted bradyzoites of *T. gondii*. Surprisingly, we found that bradyzoites isolated from brains of infected mice expressed TgNF3-YFP proteins exclusively in the cytoplasm, as demonstrated by ten Z-stack acquisitions, which allows the whole bradyzoite's body to be visualized at different depth (Figure 10B). Interestingly, when the bradyzoites were used to infect fibroblast cells and growth *in vitro* for 36 hours, TgNF3-YFP factor was again relocated in the nucleoli of all intracellular newly transformed tachyzoites (Figure 10C). The location of native TgNF3 in the cytoplasm of wild type bradyzoites was also validated using the anti-TgNF3 antibodies (Figure 11). We invariably detected native TgNF3 in the cytoplasm of all bradyzoites investigated and the signals varied from very intense fluorescent signal (panels 1-3) to weaker cytoplasmic signal (panel 4-7) and to very focused signal close to the nucleus (8-9). Despite the presence of some apparent overlapping signal between the cytoplasm and nucleus (panel 1-3), 3D constructions of panels 1 and 2 from Figure 11 showed that TgNF3 signal is exclusively localized to the cytoplasm of the bradyzoites (Video S5 and S6). This novel localization of TgNF3 becomes obviously apparent when TgNF3 signal decreased in some dormant encysted bradyzoites (panels 6-7). During the course of the confocal imaging, we discovered that the nucleus, which is more posterior in bradyzoites, displays a half reduction in size relative to nucleus of tachyzoites. In some cases, profound alterations (panels 8-10) and a total absence of nucleus were observed in few bradyzoites (panels 11 and 12). The alterations of nuclear morphology and the complete lack of nucleus always correlated with the strong decrease in TgNF3 signals. In the meantime, we also observed the striking absence of nucleoli in all bradyzoites analyzed (Figure 11). 3D-reconstructions of confocal imaging of the entire bodies of all bradyzoites investigated confirmed the exclusive presence of TgNF3 in the cytoplasm of these dormant encysted *T. gondii* forms, suggesting that the nucleolar and nuclear functions of TgNF3 are only operating in the rapidly replicating and virulent tachyzoites. Therefore, we decided to further investigate in more detail how TgNF3 protein functions biochemically *in vitro* and *in vivo*.

**TgNF3 physically interacts with both free core and nucleosome-associated histones.** Because we found a striking structural similarity between TgNF3, nuclear FKBP and nucleoplasmin-like proteins, which function as chaperones binding directly to histones and assemble histone octamers involved in

nucleosome activity [42-44, 47], we asked whether TgNF3 can associate with purified mammalian core histones. To this end, purified recombinant GST-tagged TgNF3 and GST alone produced in *E. coli* were purified and immobilized on glutathione-Sepharose, which was used to perform a series of pull-down experiments. To ensure that equal amount of TgNF3 and GST proteins were being used, aliquots of bound TgNF3-GST and GST alone to beads were eluted in SDS buffer and analyzed by SDS-PAGE (Figure 12A). Afterwards, the beads containing TgNF3-GST or GST alone were incubated with purified core histones from Hela cells, which were tested for quality before use (Figure 12B). Only TgNF3-GST was found to specifically pull down the core histones of Hela cells, as shown by SDS-PAGE and silver staining of Figure 12C (lane 2). No binding was observed when the GST alone was used (Figure 12C, lane 5). In addition, the level of core histones pulled down by TgNF3 was significantly reduced by direct competition assay in which the core histones were incubated with recombinant non-fusion TgNF3 protein prior to pull down experiments (Figure 12C, lane 3). No significant decrease of bound core histones was observed by competing with the GST alone (Figure 12C, lane 4). The specific binding of TgNF3 to histones was also confirmed when the same pulled down material analyzed by silver staining above was also subjected to Western blots using anti-H3 antibodies (Figure 12D). We found that approximately 25% of the original input bound specifically to TgNF3 (Figure 12D, lane 2). We confirmed the specificity of the competition with non-fusion recombinant TgNF3 (Figure 12D, lane 3). In addition, no binding of core histones to GST alone was observed (Figure 12D, lane 1). Next, we showed that *T. gondii* histone H3 (here named TgH3, lane 2) and TgNF3 (lane 1) are present in nucleosomes, which were purified from isolated nuclei of *T. gondii* (Figure 11E). In addition, we further demonstrated that histone TgH3 cannot be pulled down by GST-TgNF3 beads, suggesting that its association with native TgNF3 in the parasite nucleosome prevents binding site recognition and histone TgH3-TgNF3 complex formation (Figure 12F). This conclusion is supported by reciprocal immunoprecipitation using anti-TgNF3 antibodies, which validated TgNF3 as a genuine nucleosome-associated factor that interacts directly and specifically to *T. gondii* histone H3, present in parasite nucleosomes (Figure 12G). Altogether, these data support the notion that *T. gondii* TgNF3 has histone binding activity that is likely required for nucleosome functions in *T. gondii*.

***In vivo* binding of TgNF3 to ENO1 and 18S rDNA promoters is pertinent to stage-specific gene silencing in *T. gondii*.** We next investigated the direct and physical interactions of TgNF3 with promoter sequences *in vivo* using chromatin immunoprecipitation (ChIP). Towards this goal, intracellular and actively dividing tachyzoites were fixed by formaldehyde and released before chromosome fragmentation by sonication. Both TgNF3-YFP over-expressers and wild type tachyzoites of the *T. gondii* 76 K strain were used for ChIP assays using the monoclonal antibody anti-GFP and polyclonal antibodies anti-TgNF3, respectively. Figure 13 shows that all polyclonal

antibodies anti-TgNF3 individually collected from five immunized mice (designated immune sera IS1 to IS5) can positively pull down ENO1 promoter from the chromatin extract of wild type tachyzoites (Figure 13A). These results confirm our proteomic data, which identified TgNF3 among ENO1 binding nuclear factors (Supplementary Table S1). Moreover, the monoclonal antibody anti-GFP also immunoprecipitated the ENO1 promoter from the chromatin extract of TgNF3-YFP over-expressing tachyzoites (Figure 13A). We next showed that the strongest positive polyclonal antibody anti-TgNF3 (IS3) and the monoclonal antibody anti-GFP can also specifically pull down the ribosomal DNA 18S chromatin *in vivo* (Figure 13B). A pool of pre-immune sera (lanes labeled NS on the top) is not able to immunoprecipitate chromatin DNA of the ENO1 promoter and 18S rDNA in both independent experiments described in Figure 13A and 13B, and irrelevant DNA encompassing the coding region of ENO1 can not be precipitated using both anti-GFP and anti-TgNF3 antibodies (Figure 13C). Neither anti-GFP nor anti-TgNF3 can immunoprecipitate the ENO2 promoter that is active in the virulent tachyzoites using ChIP assays performed on the chromatin extracts from TgNF3-YFP over-expressing and parental tachyzoites, respectively (Figure 13D). This indicates that TgNF3 is exclusively capable of binding to ENO1 promoter that is silent in the tachyzoites [31]. Indeed, we confirmed the silent status of ENO1 promoter in tachyzoites by probing this promoter with three distinct epigenetic histone marks such as acetylated and methylated histones [7, 8]. The data showed that these three epigenetic histone marks are absent on ENO1 promoter, which is consistent with this promoter being silent in tachyzoites (Figure 13E). In contrast, the ENO2 promoter is readily modified by all three acetylated histone marks (Figure 13F), which confirmed data previously reported for active promoters [7, 8]. Taken together, we propose a functional model for TgNF3 (Figure 13G), which we define as a truly chromosome-associated factor that is probably involved in gene silencing through modulation of nucleosome assembly and/or disassembly.

## **DISCUSSION**

In this study, we sought new nuclear factors that bind to a parasite stage-specific promoter, because almost nothing is known about transcription factors or other nuclear components involved in *T. gondii* gene regulation. To this end, we used affinity purification and extensive proteomic analysis. Since it is very difficult to obtain enough nuclear protein from the persistent and dormant bradyzoites, we focused our efforts on the characterization of nuclear factors that bind to the *ENO1* promoter, which is transcriptionally silent in the rapidly replicating tachyzoites. Several novel factors isolated and identified from nuclear extracts of *T. gondii* tachyzoites have either putative DNA-binding motifs and/or putative nuclear localization sequences. Consistent with the *ENO1* gene being transcriptionally silent in tachyzoites, two candidate nuclear factors appear to be homologues of the chromatin associated protein called Alba, which is known to be involved in transcription repression that required

Sir2 and acetylation/deacetylation of lysine residue for DNA-binding in Archaea [48, 49]. Another candidate displays high sequence similarity in its N-terminal domain to yeast nuclear FK506-binding protein (FKBP), whose translocation into the nucleolus is associated with the transcriptional silencing of ribosomal RNA genes [32]. The identification in *T. gondii* of proteins sharing similarities with yeast and archaeal proteins known as transcriptional repressors is in good agreement with the *ENO1* promoter being silent in the rapidly replicating tachyzoites. We have validated that most of the candidate factors tested, including the homologues of the two archaeal Alba and the protein sharing similarity with yeast FKBP39, are capable of entering the nucleus of transiently transfected *T. gondii* tachyzoites. We showed that the homologue of the yeast nuclear FKBP39, which we termed TgNF3 for *Toxoplasma gondii* nuclear factor 3, is largely present in the nucleolus of the parasites even if the factor can also be weakly detected in the nucleoplasm. This dual nucleolar/nucleoplasm location of TgNF3 is consistent with the behavior of yeast FKBP39, which displays a similar dual sub-nuclear distribution [32], suggesting that these two factors may share similar nuclear functions. The *Saccharomyces pombe* SpFKBP39, which is a novel chromatin-modulating factor that acts as a histone chaperone regulating rDNA silencing, is a bi-partite protein with its N-terminal domain having the chaperone activity required in nucleosome assembly and disassembly that involves interaction with chromatin-remodelling factors and chromatin-modifying enzymes. This function is independent from the peptidyl-prolyl isomerase (PPIase) activity of its C-terminal FKBP domain, which is not present in TgNF3 [32]. Our bioinformatic analyses have provided evidence that the N-terminal domain of TgNF3 belongs to the HD2/nuclear FKBP family, including nuclear FKBP factors previously found only in yeasts and insects, plant-specific histone deacetylase 2 (HD2) and proteins from parasitic apicomplexans such as *Plasmodium falciparum*, the causative agent of malaria [39]. In the present study, we show that members of the HD2/nuclear FKBP family including TgNF3 are related to the nucleoplasmin/nucleophosmin (NPM) family, which was the first family of histone chaperones to be identified [40, 50]. The relationship to the NPM family, for which 3D structures are known [42-44], allowed the mapping on the Np core 3D structure of two residues, highly conserved in the HD2/nuclear FKBP family, an aspartic acid and a histidine, which may play a key role in lysine deacetylation [45]. Thus, our data provide an alternative hypothesis for the binding of histones on the NPM/HD2 scaffold to that previously proposed, which implicated the decamer periphery [42-45]. Rather, it is suggested here that the center of the pentamers, at the distal ends of the dodecameric organization, may be involved in the interaction with histones. This hypothesis would also fit well with implications of the acidic loops in histone binding, suggesting that TgNF3/HD2/nuclear FKBP would adopt a NPM-like quaternary association. Here, we also provide experimental evidence that TgNF3 binds specifically to both mammalian core histones containing H3, H4 and H2A/H2B complexes and

to parasite nucleosome-associated histones. Furthermore, we showed that TgNF3 is a chromatin-associated protein predominantly accumulating in the nucleolus of the parasite. TgNF3 can also specifically bind *in vivo* to the 18S ribosomal RNA gene promoter of *T. gondii*. High resolution confocal and electron microscopy imaging demonstrated that TgNF3 is a predominant resident factor of the nucleoli of the actively replicating intracellular virulent tachyzoites of *T. gondii*, the parasite forms that are responsible of the acute and life-threatening phase of the disease. Our attempts to knock-out the *TgNF3* gene in *T. gondii* failed even if a  $\Delta$ ku80 parasite strain lacking random DNA integration was used [51], suggesting that the *TgNF3* gene codes for an essential function, or the locus is inaccessible to double homologous recombination. Therefore, we decided to over-express the *TgNF3* gene in the parasite. The difference in *TgNF3* transcript levels between the two parasite forms, the rapidly replicating tachyzoites and persistent dormant bradyzoites is not only pertinent to this strategy, but also to investigate its biological relevance. Using a strong tubulin promoter, we have successfully generated transgenic tachyzoites, which over-express TgNF3 fused to YFP. Over-expression did not appear to be toxic, even though the size of parasite nucleoli was 4-5 fold increased in size. The over-expresser's nucleoli were often moved to one end of the nucleus with important modifications of the nuclear shape, whereas the nucleolar size and position in wild type tachyzoites remained centrally positioned in the nucleus. The TgNF3 over-expression allowed us to investigate the dynamics of the nucleolus in *T. gondii* during the parasite development inside the host cell, an issue that has not been investigated previously. Time-lapse images and movies consistently showed a dynamic nucleolar biogenesis with the appearance of two smaller sized nucleoli, or one small nucleolus close to a larger one during parasite replication, suggesting either fragmentation and/or *de novo* synthesis of the parasite nucleolus. The over-expression of TgNF3 in the tachyzoites induces two-fold rapid replication rate *in vitro* in transgenic over-expressers relative to wild type parasites. Because TgNF3 and the other nuclear factors are identified based on their binding to a transcriptionally silent gene in tachyzoites, it is tempting to conclude that silencing of bradyzoite-specific *ENO1* gene and other genes may activate expression of tachyzoite-specific genes and promote the rapid parasite replication and growth *in vitro*. However, the fact that transgenic tachyzoites over-expressing TgNF3 displayed a severe attenuation of virulence *in vivo* accompanied by cyst formation in the brain of infected mice suggests that the increased rapid replication rate of *T. gondii* tachyzoites *in vitro* is not necessary linked to virulence *in vivo*. This is in good agreement with the recent elucidation of the molecular basis of *T. gondii* virulence and the discovery of kinase-like proteins in rhoptries, the apical secretory organelles. Rhoptries release the kinase-like proteins into the nuclei of cells of infected hosts leading to death [52, 53]. Interestingly, we have also shown that 100% of mice survived when infected with TgNF3 over-expressers. When these survivors were challenged with high

doses of two distinct *T. gondii* strains including the highly virulent RH type I strain, they displayed no symptoms of disease, with the complete absence of cysts derived from parasites used to challenge the mice. These data also suggest that the mice displayed a sterile protection and the transgenic TgNF3 over-expressers may serve as a useful basis for the development of vaccine strategies, especially for livestock. It will also be very interesting to examine the ultra-structural morphology of these cysts. However, their limited numbers has hampered such analysis so far. Nonetheless, we have been able to release fewer bradyzoites from the cysts obtained *in vivo* during the primo-infection of mice with TgNF3-YFP transgenic parasites. We found that despite the presence of abundant level of both TgNF3 transcript and protein, the presence of TgNF3 protein is exclusively restricted to the cytoplasm of the avirulent and dormant bradyzoites. We noticed that the decrease of TgNF3 signal in the cytoplasm correlates with a profound size reduction and degeneration in some cases of bradyzoite nuclei. The exclusive location of TgNF3 in the cytoplasm of dormant bradyzoites together with its over-expression, which leads to an enlarged nucleoli with more rapid replication rate in the virulent tachyzoites, suggest that nucleolar functions may be key for the replication of *T. gondii* tachyzoites, at least *in vitro*. It is interesting to note that TgNF3 is also substantially present outside the nucleolus, namely in the nucleoplasm, where it can interact with other gene promoters. This has been demonstrated for the bradyzoite-specific *ENO1* promoter. Since the *ENO1* gene is transcriptionally silent in tachyzoites and its promoter has been used to isolate TgNF3 and the other novel nuclear factors, an attractive model, which is schematically presented in Figure 13G, is that TgNF3 may be a *T. gondii* nuclear chaperone involved in alterations in chromatin structure and functions. TgNF3 could also affect protein translation and ribosomal synthesis both known to be widely implicated in diverse cellular pathways during growth of metazoan and protozoan species. We propose that changes in chromatin structures and functions may link the dual nucleolar/nuclear TgNF3 functions to the control of genes that regulate parasite proliferation *in vitro* and virulence *in vivo* including differentiation and cyst formation.

## **MATERIALS AND METHODS**

***In vitro* growth of *T. gondii* and nuclear extract preparation.** Human foreskin fibroblasts (HFFs) were maintained in Dulbecco's modified Eagle's medium (DMEM, BioWhittaker, Verviers, Belgium) supplemented with 10% fetal calf serum, 2 mM glutamine (Sigma), and 0.05% gentamycin (Sigma). Tachyzoites from *T. gondii* 76K strain were grown in monolayers of HFF cells until they lysed the host cells spontaneously. Freed tachyzoites were harvested and purified using glass wool columns and 3- $\mu$ m pore filters. Encysted bradyzoites were purified from brains of mice chronically infected by *T. gondii* 76K strain, as previously described [23]. Nuclear extracts were obtained from about  $4 \times 10^{10}$

purified tachyzoites of *T. gondii* 76 K strain, as previously described [31]. The parasite nuclear extracts were aliquoted and stored at -80°C before use.

**Promoter DNA-biotinylation and affinity purification.** The nucleotide sequence corresponding to ENO1 promoter was amplified by PCR and cloned into BlueScript. The 200 µg DNA fragment was sequentially digested with EcoRI and HincII enzymes at 37°C for 2 hours. The released DNA fragment was purified by GeneClean after electrophoresis on agarose gel. The purified DNA fragment was biotinylated using Klenow enzyme, 1mM of dATP, 1mM dCTP, 1mM dGTP and 8 µl of dUTP 16-Biotin at 37°C for 2 h. The biotinylated DNA was purified by High Pack PCR kit (Roche). Total nuclear extract of about 17 mg of proteins was diluted with 7 ml of binding buffer (12% glycerol, 1.5mM MgCl<sub>2</sub>, 12 mM Hepes pH 7.9, 4 mM Tris.HCl pH 7.9, 60 mM KCl, 1 mM EDTA, 1 mM DTT) and incubated with 800 µl of Streptavidine-agarose at 4°C for 1h. After centrifugation, the supernatant was mixed with 350 µl of biotinylated DNA and the mixture was incubated at room temperature for 30 minutes. The DNA-protein complexes were isolated by affinity purification using 800 µl of Streptavidine-agarose for 4 h at 4°C. The beads were washed with binding buffer 10 times and eluted with 3ml of elution buffer (12% glycerol, 20 mM Tris.HCl pH 6.8, 1M KCl, 5mM MgCl<sub>2</sub>, 1 mM EDTA, 1 mM DTT) at 4°C for 40 min. The eluted proteins were dialyzed and concentrated by centrifugation using Centricon tubes (MWCO10, Millipore) and analyzed by SDS-PAGE followed by silver staining.

**Sample preparation for proteomic analysis.** Silver stained gels were systematically cut into slices, and in-gel digestion was performed with an automated protein digestion system, MassPREP Station (Waters, Milford, MA). The gel slices were washed three times in a mixture containing 25 mM NH<sub>4</sub>HCO<sub>3</sub>: CH<sub>3</sub>CN (1:1, v/v). The cysteine residues were reduced with 50 µL of 10 mM dithiothreitol at 57°C and alkylated with 50 µL of 55 mM iodoacetamide. After dehydration with acetonitrile, the proteins were cleaved in gel with 40 µL of 12.5 ng/µL of modified porcine trypsin (Promega, Madison, WI, USA) in 25 mM NH<sub>4</sub>HCO<sub>3</sub> at 37°C for 4 hours. The tryptic peptides were extracted with 60% acetonitrile in 0.1% formic acid.

**Mass spectrometry analysis.** NanoLC-MS/MS analyses were performed using an Agilent 1100 series nanoHPLC-Chip/MS system (Agilent Technologies, Palo Alto, USA) coupled to a HCT Plus ion trap (Bruker Daltonics, Bremen, Germany). The chip contained a Zorbax 300SB-C18 column (43 mm x 75 µm, 5 µm particle size) and a Zorbax 300SB-C18 enrichment column (40 nL, 5 µm particle size). The solvent system consisted of 2% acetonitrile, 0.1% formic acid in water (solvent A) and 2% water, 0.1% formic acid in acetonitrile (solvent B). 2 µL of each sample was loaded into the enrichment column at a flow rate set to 3.75 µL/min with solvent A. Elution was performed at a flow rate of 300 nl/min with

a 8-40% linear gradient (solvent B) in 7 minutes followed by a 3 min stage at 70% of solvent B before reconditioning the column at 92% of solvent A. The system was fully controlled by ChemStation Rev B.01.03.SRI [204] (Agilent Technologies). The MS instrument was operated with the following settings: source temperature was set to 320°C while cone gas flow was at 3 l /min. The capillary voltage was optimized to -1850 V. The MS spectra were acquired in the positive ion mode on the mass range 250 to 2500 m/z using the standard enhanced resolution mode at a scan rate of 8.100 m/z/s. The Ion Charge Control was fixed at 100000 with a maximum accumulation time of 200 ms and the number of averages was set to 4. For tandem MS experiments, the system was operated with automatic switching between MS and MS/MS modes. The 3 most abundant peptides were selected on each MS spectrum for further isolation and fragmentation with a preference for doubly charged ions (absolute threshold of 2000 and a relative of 5%). Ions were excluded after the acquisition of 2 MS/MS spectra and the exclusion was released after one minute. The Smart Parameters Setting option was used for the selected precursor ions. The MS/MS spectra were acquired on the mass range 50 to 2800 m/z using the ultrascan resolution mode at a scan rate of 26.000 m/z/s. The Ion Charge Control was fixed at 300000 and 6 scans were averaged to obtain a MS/MS spectrum. The MS/MS fragmentation amplitude was set to 1.5 V. The system was fully controlled by the EsquireControl 5.3 Build 11.0 software (Bruker Daltonics). Mass data collected during the nanoLC-MS/MS analyses were processed and converted into \*.mgf files using the DataAnalysis 3.3 Build 146 software (Bruker Daltonics).

**Protein identification.** The MS/MS data were analyzed using the MASCOT 2.2.0. algorithm (Matrix Science, London, UK) and Open Mass Spectrometry Search Algorithm (OMSSA) [55] for search against an in-house generated protein database [56], which is composed of protein sequences of *Apicomplexa* downloaded from <http://www.ncbi.nlm.nih.gov/sites/entrez> (on may 25, 2009) concatenated with reversed copies of all sequences (total 307386 entries). Searches were performed with a mass tolerance of 0.25 Da in both MS and MS/MS mode and with the following parameters: full trypsin specificity with one missed cleavage, methionine oxidation, protein amino-terminal acetylation and cystein carbamidomethylation. The Mascot and OMSSA results were loaded into the Scaffold software (Proteome Software, Portland, USA). To minimize false positive identifications, results were subjected to very stringent filtering criteria as follows. For the identification of proteins with two peptides or more, a Mascot ion score above the identity score or an OMSSA E-value below  $-\log(e^2)$  was required. The target-decoy database search allowed us to control and estimate the false positive identification rate of our study [57, 58]. Thus, the final catalogue of proteins presents an estimated false positive rate below 1%.

**Plasmid construction, transfection and transformation.** The full-length cDNAs corresponding to several proteins identified by proteomic analysis were fused to YFP using pTub1-YFP and pGRA1-

HAFLAG plasmids, which were generously provided by Dr Dzierszinski F & Roos DS (University of Pennsylvania, USA) and Dr Hakimi MA (University of Grenoble, France), respectively. The expression vectors were transiently transfected in the tachyzoites and processed for IFA after 24 h post-infection. For stable transformation of *T. gondii*, Tub1-TgNF3-YFP plasmid was transfected in  $10^7$  tachyzoites of *T. gondii* 76K strain and grown in the presence of 20 µg/ml of chloramphenicol until emerged resistant and stable parasites were cloned.

For recombinant protein expression, the full-length cDNA of TgNF3 was cloned in pGEX and used to transform *E. coli*. The recombinant TgNF3 was induced by IPTG and the soluble recombinant protein was purified using GST column according to the manufacturer's recommendations. A lot of five mice were immunized by 100 µg of recombinant protein per mouse in the presence of complete Freund adjuvant (Sigma). After two boosts with 50 µg of recombinant TgNF3 per mouse in the presence of incomplete adjuvant (Sigma), the sera were tested by Western blots before a final challenge.

**RT-PCR, GST pull down assays and Western blotting.** For RT-PCR, the RNA from  $10^8$  tachyzoites purified from infected HFF cells and  $10^6$  bradyzoites isolated from 2,000 cysts of brains of chronically infected mice were used. The total RNA were reverse transcribed for one hour at 42°C in a buffer containing 1 M oligo(dT)<sub>18</sub> primer, 2 mM dNTPs, 40 U of rRNasin (Promega) and 25 U of AMV reverse transcriptase (Roche). The primers used for RT-PCR were as follows: *T. gondii* actin gene, forward, 5'-atggcggatgaagaagtgcaa-3', and reverse, 5'-gatgcacttgcggtggacgat-3'; and *T. gondii* TgNF3, forward, 5'-atgttctacggtgtcgtgtgc-3', reverse, 5'-tttcagcgagcacacgcc-3'. PCR products were electrophoresed on agarose gels, stained with ethidium bromide, and photographed using UV-light scanner. For GST-pull down experiments, 100 µg of purified recombinant GST-TgNF3 and GST alone were resuspended in binding buffer (20 mM Hepes pH 7.9, 150 mM NaCl, 0.5 mM EDTA, 10% glycerol, 0.1% Tween 20 and a cocktail of protease inhibitor and PMSF) and incubated with 100 µl of glutathione-Sepharose 4B (Amersham Pharmacia Biotech) at 4°C overnight. The beads were washed three times with the binding buffer without protease inhibitors. Five µg of purified core histones from HeLa cells (Upstate, Millipore) or purified nucleosomes from isolated nuclei of *T. gondii* [7], kindly provided by Dr Hakimi MA, were resuspended with the binding buffer and added to 10 µl of GST-TgNF3, GST and to anti-TgNF3 antibodies coupled to Sepharose beads. For competition assays, 5 µg of core histones or purified *T. gondii* nucleosomes were incubated with recombinant non-fusion TgNF3 or GST for one hour at room temperature prior pull down assays. After incubation at room temperature and four washes, pulled down proteins were eluted by boiling by 25 µl of Laemmli sample buffer and analyzed by SDS-PAGE for silver staining or Western blots using rabbit polyclonal anti-Histone H3 antibody (Upstate, Amersham).

Total protein extracts from *T. gondii* tachyzoites ( $5 \times 10^6$  tachyzoites per lane) or purified *T. gondii* nucleosomes were also boiled in Laemmli's buffer, separated by SDS-PAGE and transferred to Hybond ECL nitrocellulose (Amersham). Immunoblots were carried out using murine immune sera, non-immune sera as control, monoclonal antibodies specific to GFP (Clontech) or with mouse polyclonal antibody raised against recombinant anti-TgNF3. The blots were incubated with peroxidase conjugated secondary antibodies followed by chemiluminescence detection.

**Electron microscopy and immunofluorescence assays.** For immuno-electron microscopy, the intracellular transgenic tachyzoites expressing TgNF3-YFP grown in HFF cells were fixed overnight at 4°C in 8% paraformaldehyde in PBS buffer, thoroughly washed in the same buffer and infused in sucrose 2.3 M containing 20 % polyvinyl pyrrolidone 10000 in phosphate buffer 0.1 M. The pellets were mounted on ultracryotome supports and rapidly frozen in melting nitrogen. Ultrathin sections of about 90-100 nm were obtained using Reichert UltraCut E ultramicrotome equipped with a FC4 device. Before mounting in methyl cellulose, sections were incubated in blocking medium (0.05 M glycine, 5% fish gelatine in 0.1 M PBS buffer) for 30 min. The grids were incubated with the monoclonal antibody anti-GFP for 1 hour at 37°C or overnight at 4°C. After washing, sections were incubated at room temperature for 30 min in the corresponding secondary gold conjugates (Jackson ImmunoResearch Laboratories Inc.) diluted in the same buffer. Following a final thorough wash in PBS alone, the grids were fixed in 4% paraformaldehyde for 10 min at room temperature and washed in water. After staining with 0.5 % uranyl acetate in 1.5 % methyl cellulose, sections were observed on a Hitachi H600 transmission electron microscope at 75kV accelerating voltage.

For immunofluorescence assays (IFA), intracellular tachyzoites were fixed with 4% paraformaldehyde in PBS for 15 minutes on ice, followed by two PBS washes and dried on Teflon slides. Intracellular parasites were permeabilized with 0.1% Triton X-100 in PBS containing 0.1% glycine for 10 minutes at room temperature. Samples were blocked with 3% BSA in the same buffer and mice immune sera diluted at 1:1000 were added on parasites in the same buffer for one hour at 37°C. Rabbit secondary antibody coupled to Alexa-488 (Molecular Probes) diluted at 1:1000 was added in addition to DAPI for nucleus staining. For co-localization assay, the rabbit anti-MPS1 serum and the goat secondary antibody coupled to Alexa-594 (Molecular Probes) were used at the same dilution. Fluorescence was visualized with a ZEISS Axiophot microscope.

**Imaging by confocal microscopy.** Confocal imaging was performed with a LSM710 microscope (Zeiss) and a Plan Apochromat objective (Plan-Apochromat 63x/1.40 Oil DIC M27, Zeiss). The associated software (Zen 2008) enabled the adjustment of acquisition parameters. YFP was excited at 514 nm and its fluorescence emission was collected from 520 to 620 nm. The Alexa-488 signal was excited at 488 nm and emission was collected from 500 to 580 nm. Nuclear DAPI signal were excited

at 405 nm and emission was collected from 410 to 500 nm. Fluorescent signals were collected sequentially, with a 4 lines average, a zoom factor (varying between 2 and 4) and resulting images are 512x512 pixels in size, and 8 bits in resolution (256 gray levels). By setting the photomultiplier tubes and the pinhole size (1 AU) correctly, there was no signal bleed-through. The images were treated with ImageJ (NIH). Z-stack acquisitions enabled to visualize the 3D localization of fluorescent signals. Rotation (360° around the y axis) movies were created with the Zen software, at a rate of 5 frames per second.

**Live video microscopy.** Images were obtained from an AxioObserver Z1 (Zeiss), equipped with a regulation chamber for temperature and CO<sub>2</sub>, through a Plan Apochromat objective (Plan-Apochromat 100x/1.46 Oil, Zeiss). Samples were excited with a Colibri source (Zeiss) and fluorescence signal was collected with an Axio mRm Camera (Zeiss), through different filters sets (Filter Set 38HE, 60HE, Zeiss). Experiments were performed by exposing the sample at a rate of one image every 30 minutes during 72 h. Z-stack acquisitions were performed, and resulting images, which correspond to the projections of these stacks, were shown. The images were treated using ImageJ (NIH).

**Chromatin immunoprecipitation.** ChIP was performed as described [7] with slight modifications. Briefly, chromatin from extracellular or intracellular parasites grown in HFF cells was cross-linked for 10 min with 1% formaldehyde at room temperature and purified. Chromatin extract was obtained after sonication yielding fragments of 500-1,000 bp. Immunoprecipitations were performed with the monoclonal antibody anti-GFP or the mouse serum anti-TgNF3 at 4°C overnight and washed as described [7]. DNA was then subjected to proteinase K digestion for 2 h and purified using the Qiagen PCR purification kit (<http://www.qiagen.com>). As a negative control, naive mouse serum was used. For PCR, specific primers of ENO1 gene promoter, forward 5'-atg tgetgctggtttttgttc-3', reverse, 5'-ttaaggcgtccgcaagactagtg-3' and ribosomal RNA 18S gene promoter, forward, 5'-ggggtggtggatggggacgggagc-3', reverse 5'-gccgttctctgaccccgtgcc3' were used. ChIP products amplified by PCR using these specific primers were electrophoresed on agarose gels, stained with ethidium bromide, and photographed using UV-light scanner.

**Bioinformatics.** Sequence similarity searches were performed using PSI-BLAST at the National Center for Biological Information (NCBI, nonredundant (nr) database, default parameters). Fold recognition was performed using Phyre [59]. Hydrophobic Cluster Analysis (HCA, [39]) was used to analyze protein domain architecture and refine sequence alignments.

**Statistical analysis.** Statistical differences between groups of mice used in this study were evaluated by the Student's t-test. The Mann-Whitney test was also used for counting intracellular growth of TgNF3-YFP and wild type tachyzoites after colorimetric staining and microscopic observation, and also for the test for mice survival curves.

## **ACKNOWLEDGMENTS**

We would like to thank Dewailly E and Mortuaire M for excellent technical assistance, especially in the earlier phase of this study. We acknowledge Drs Harb O and Roos DS (University of Pennsylvania) for providing the nucleotide sequence of *T. gondii* rDNA 18S. We are indebted to Drs A. Hakimi and D. Leprince for providing nucleosome and antibodies specific to core histones, respectively. Special thanks to Drs Gordon Langsley and Robert Walker for critical reading the manuscript and all members of the lab for fruitful discussions.

## **FUNDING**

This research was funded by the Centre National de la Recherche Scientifique (CNRS), Institut Médicale de la Recherche Scientifique (INSERM) and the Institut Pasteur de Lille (IPL). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## **COMPETING INTERESTS**

The authors have declared that no competing interests exist.

## **AUTHOR CONTRIBUTIONS**

Conceived and designed the experiments: ST, AV, CSR. Performed the experiments: AOL, ED, AH, EW, IC, CS, SD, TM. Wrote the paper: ST, CS, CSR.

## **REFERENCES**

1. Aravind L, Iyer LM, Wellems TE, Miller LH (2003) *Plasmodium* biology: genomic gleanings. Cell 115: 771-785.
2. Gardner MJ, Hall N, Fung E, White O, Berriman M, *et al.* (2002) Genome sequence of the human malaria parasite *Plasmodium falciparum*. Nature 419: 498-511.
3. Templeton TJ, Iyer LM, Anantharaman V, Enomoto S, Abrahante JE, *et al.* (2004) Comparative analysis of apicomplexa and genomic diversity in eukaryotes. Genome Res 14: 1686-1695.
4. Abrahamsen MS, Templeton TJ, Enomoto S, Abrahante JE, Zhu G, *et al.* (2004) Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*. Science 304: 441-445.
5. Callebaut I, Prat K, Meurice E, Mornon JP, Tomavo S (2005) Prediction of the general transcription factors associated with RNA polymerase II in *Plasmodium falciparum*: conserved features and differences relative to other eukaryotes. BMC Genomics 6: 100.

6. Meissner M, Soldati D (2005) The transcription machinery and the molecular toolbox to control gene expression in *Toxoplasma gondii* and other protozoan parasites. *Microbes Infect* 7: 1376-1384.
7. Saksouk N, Bhatti MM, Kieffer S, Smith AT, Musset K, *et al.* (2005) Histone-modifying complexes regulate gene expression pertinent to the differentiation of the protozoan parasite *Toxoplasma gondii*. *Mol Cell Biol* 25: 10301-10314.
8. Gissot M, Kelly KA, Ajioka JW, Grealley JM, Kim K (2007) Epigenomic modifications predict active promoters and gene structure in *Toxoplasma gondii*. *PLoS Pathog* 3: e77.
9. Hakimi M.-A, Deitsch KW (2007) Epigenetics in Apicomplexa: control of gene expression during cell cycle progression, differentiation and antigenic variation. *Current Opinion in Microbiology* 10: 357-362.
10. Balaji S, Babu MM, Iyer LM, Aravind L (2005) Discovery of the principal specific transcription factors of Apicomplexa and their implication for the evolution of the AP2-integrase DNA binding domains. *Nucleic Acids Res* 33: 3994-4006.
11. De Silva EK, Gehrke AR, Olszewski K, Leon I, Chahal JS, Bulyk ML, Llinas M. (2008) Specific DNA-binding by Apicomplexan AP2 transcription factors. *Proc Natl Acad Sci USA* 105: 8393-8398.
12. Yuda M, Iwanaga S, Shigenobu S, Mair GR, Janse CJ, Waters AP, Kato T, Kaneko I (2009) Identification of a transcription factor in the mosquito-invasive stage of malaria parasites. *Mol Microbiol* 71: 1402-1414.
13. Radke JR, Behnke MS, Mackey AJ, Radke JB, Roos DS, White MW (2005) The transcriptome of *Toxoplasma gondii*. *BMC Biol* 3:26.
14. Behnke MS, Radke JB, Smith AT, Sullivan WJ Jr, White MW (2008) The transcription of bradyzoite genes in *Toxoplasma gondii* is controlled by autonomous promoter elements. *Mol Microbiol* 68: 1502-1518.
15. Cleary MD, Singh U, Blader IJ, Brewer JL, Boothroyd JC (2002) *Toxoplasma gondii*: asexual development: identification of developmentally regulated genes and distinct patterns of gene expression. *Eukaryot Cell* 1:329-340.
16. Matrajt M, Donald RG, Singh U, Roos DS (2002) Identification and characterization of differentiation mutants in the protozoan parasite *Toxoplasma gondii*. *Mol Microbiol* 44: 735-747.
17. Singh U, Brewer JL, Boothroyd JC (2002) Genetic analysis of tachyzoite to bradyzoite differentiation mutants in *Toxoplasma gondii* reveals a hierarchy of gene induction. *Mol Microbiol* 44: 721-733.
18. Ellis J, Sinclair D, Morrison D (2004) Microarrays and stage conversion in *Toxoplasma gondii*. *Trends Parasitol* 20: 288-295.
19. Jerome ME, Radke JR, Bohne W, Roos DS, White MW (1998) *Toxoplasma gondii* bradyzoites

- form spontaneously during sporozoite-initiated development. *Infect Immun* 66: 4838-4844.
20. Yang S, Parmley AF (1997) *Toxoplasma gondii* expresses two distinct lactate dehydrogenase homologous genes during its life cycle in intermediate hosts. *Gene* 184: 1-12.
21. Manger ID, Hehl A, Parmley S, Sibley LD, Marra M, Hillier L, Waterston R, Boothroyd JC (1998) Expressed sequence tag analysis of the bradyzoite stage of *Toxoplasma gondii*: identification of developmentally regulated genes. *Infect Immun* 66: 1632-1637.
22. Dziarszinski F, Popescu O, Toursel C, Slomianny C, Yahiaoui B, Tomavo S (1999) The protozoan parasite *Toxoplasma gondii* expresses two functional plant-like glycolytic enzymes. Implications for evolutionary origin of apicomplexans. *J Biol Chem* 274: 24888-24895.
23. Dziarszinski F, Mortuaire M, Dendouga N, Popescu O, Tomavo S (2001) Differential expression of two plant-like enolases with distinct enzymatic and antigenic properties during stage conversion of the protozoan parasite *Toxoplasma gondii*. *J Mol Biol* 309: 1017-1027.
24. Delbac F, Sanger A, Neuhaus E, Stratmann R, Ajioka WJ, Toursel C, Herm-Gotz A, Tomavo S, Soldati T, Soldati D (2001) *Toxoplasma gondii* myosins B/C: one gene, two tails, two localizations, and a role in parasite division. *J Cell Biol* 155: 613-623.
25. Meissner M, Reiss M, Viebig N, Carruthers V, Toursel C, Tomavo S, Ajioka J, Soldati D. (2001) A family of transmembrane microneme proteins of *Toxoplasma gondii* contain EGF-like domains and function as escorts. *J Cell Sci* 115: 563-574.
26. Morris MT, Coppin A, Tomavo S, Carruthers VB. (2002) Functional analysis of *Toxoplasma gondii* protease inhibitor 1. *J Biol Chem* 277: 45259-45266.
27. Coppin A, Varre JS, Lienard L, Dauvillee D, Guerardel Y, Soyer-Gobillard MO, Buleon A, Ball S, Tomavo S (2005) Evolution of plant-like crystalline storage polysaccharide in the protozoan parasite *Toxoplasma gondii* argues for a red alga ancestry. *J Mol Evol* 60: 257-267.
28. Que X, Engel JC, Ferguson D, Wunderlich A, Tomavo S, Read SL (2007) Cathepsin Cs are key for the intracellular survival of the protozoan parasite, *Toxoplasma gondii*. *J Biol Chem* 282: 4994-5003.
29. Soldati D, Boothroyd JC (1995) A selector of transcription initiation in the protozoan parasite *Toxoplasma gondii*. *Mol Cell Biol*. 15: 87-93.
30. Mercier C, Lefebvre-Van Hende S, Garber GE, Lecordier L, Capron A, Cesbron Delauw MF (1996) Common cis-acting elements critical for the expression of several genes of *Toxoplasma gondii*. *Mol. Microbiol* 21: 421-428.
31. Kibe M, Coppin A, Dendouga N, Oria G, Meurice E, Mortuaire M, Madec E, Tomavo S (2005) Transcriptional regulation of two stage-specifically expressed genes in the protozoan parasite *Toxoplasma gondii*. *Nucleic Acids Res* 33: 1722-1736.

32. Bell SD, Botting CH, Wardleworth BN, Jackson SP White MF (2002) The interaction of Alba, a conserved archaeal chromatin protein, with Sir2 and its regulation by acetylation. *Science* 296: 148-151.
33. Jelinska, C., Conroy, M.J., Craven, C.J., Hounslow, A.M., Bullough, P.A., Waltho, J.P., Taylor, G.L., and White, M.F. (2005). Obligate heterodimerization of the archaeal Alba2 protein with Alba1 provides a mechanism for control of DNA packaging. *Structure* 13: 963-971.
34. Kuzuhara T, Horikoshi M (2004) A nuclear FK506-binding protein is a histone chaperone regulating rDNA silencing. *Nat Struct Mol Biol* 11: 275-283.
35. Ferguson DJ, Parmley SF, Tomavo S (2002) Evidence for nuclear localisation of two stage-specific isoenzymes of enolase in *Toxoplasma gondii* correlates with active parasite replication. *Int J Parasitol* 32: 1399-1410.
36. Degen WG, Agterbos MA, Muyrers JP, Bloemers HP, Swart GW (1999) memA/DRS, a putative mediator of multiprotein complexes, is overexpressed in the metastasizing human melanoma cell lines BLM and MV3. *Biochim Biophys Acta* 1444: 384-394.
37. Aravind L, Iyer LM, Anantharaman V (2002) The two faces of Alba: the evolutionary connection between proteins participating in chromatin structure and RNA metabolism. *Genome Biology* 4: R64.
38. Callebaut I, Labesse G, Durand P, Poupon A, Canard L, Chomilier J, Henrissat B, Mornon JP (1997) Deciphering protein sequence information through hydrophobic cluster analysis (HCA): current status and perspectives. *Cell Mol Life Sci* 53: 621-645.
39. Aravind L, Koonin EV (1998) Second family of histone deacetylases. *Science* 280: 1167-1168.
40. Frehlick LJ, Eirín-López JM, Ausió J (2007) New insights into the nucleophosmin/nucleoplasmin family of nuclear chaperones. *Bioessays* 29: 49-59.
41. Iyer LM, V A, Wolf M, Aravind L (2008) Comparative genomics of transcription factors and chromatin proteins in parasitic protists and other eukaryotes. *Int J Parasitol* 38: 1-31.
42. Dutta S, Akey IV, Dingwall C, Hartman KL, Laue T, Nolte RT, Head JF, Akey CW (2001) The crystal structure of nucleoplasmin-core: implications for histone binding and nucleosome assembly. *Mol Cell* 8: 841-853.
43. Namboodiri VM, Dutta S, Akey IV, Head JF, Akey CW (2003) The crystal structure of *Drosophila* NLP-core provides insight into pentamer formation and histone binding. *Structure* 11: 175-186.
44. Namboodiri VM, Akey IV, Schmidt-Zachmann MS, Head JF, Akey CW (2004) The structure and function of *Xenopus* NO38-core, a histone chaperone in the nucleolus. *Structure* 12: 2149-2160.
45. Zhou C, Labbe H, Sridha S, Wang L, Tian L, Latoszek-Green M, Yang Z, Brown D, Mik iB, Wu K (2004) Expression and function of HD2-type histone deacetylases in *Arabidopsis* development. *Plant J* 38: 715-724.

46. Boothroyd JC, Black M, Bonnefoy S, Hehl A, Knoll LJ, Manger ID, Ortega-Barria E, Tomavo S. (1997) Genetic and Biochemical analysis of development in *Toxoplasma gondii*. *Philos Trans R Soc Lond B Biol Sci* 352: 1347-1354.
47. Namboodiri HVM, Dutta S, Akey IV, Head JF, Akey CW (2003) The crystal structure of Drosophila NLP-core provides insight into pentamer formation and histone binding. *Structure* 11: 175-186.
48. Zhao K, Chai X, Marmorstein R (2003) Structure of a Sir2 substrate Alba, reveals a mechanism for deacetylation-induced enhancement of DNA-binding. *J Biol Chem* 278: 26071-26077.
49. Marsh VL, Peak-Chew SY, Bell SD (2005) Sir2 and the acetyltransferase, Pat, regulate the archaeal chromatin protein, Alba. *J Biol Chem* 280: 21122-21128.
50. Laskey RA, Honda BM, Mills AD, Finch JT (1978) Nucleosomes are assembled by an acidic protein which binds histones and transfers them to DNA. *Nature* 275: 416-420.
51. Huynh MH, Carruthers VB (2009) Tagging of endogenous genes in a *Toxoplasma gondii* strain lacking KU80. *Eukaryot Cell* 8: 530-539.
52. Saeij JP, Boyle JP, Collier S, Taylor S, Sibley LD, Brooke-Powell ET, Ajioka JW, Boothroyd JC (2006) Polymorphic secreted kinases are key virulence factors in toxoplasmosis. *Science* 314: 1780-1780.
53. Taylor S, Barragan A, Su C, Fux B, Fentress SJ, Tang K, Beatty WL, Hajj HE, Jerome M, Behnke MS, White M, Wootton JC, Sibley LD (2006) A secreted serine-threonine kinase determines virulence in the eukaryotic pathogen *Toxoplasma gondii*. *Science* 314: 1776-1780.
54. Akey CW, Luger K (2003) Histone chaperones and nucleosome assembly. *Current Opinion in Structural Biology* 13: 6-14.
55. Geer LY, Markey SP, Kowalak JA, Wagner L, Xu M, Maynard DM, Yang X, Shi W, Bryant SH (2004) Open mass spectrometry search algorithm. *J Proteome Res.* 3: 958-964.
56. Fauquenoy S, Morelle W, Hovasse A, Bednarczyk A, Slomianny C, Schaeffer C, Van Dorselaer A and Tomavo S (2008) Proteomics and glycomics analyses of N-glycosylated structures involved in *Toxoplasma gondii*-host cell interactions. *Mol Cell Proteomics* 7: 891-910.
57. Elias JE, Gygi SP (2007) Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods* 4: 207-214.
58. Peng J, Elias JE, Thoreen CC, Licklider LJ, Gygi SP (2003) Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. *J Proteome Res* 2: 43-50.
59. Bennett-Lovsey RM, Herbert AD, Sternberg MJE, Kelley LA (2008) Exploring the extremes of sequence/structure space with ensemble fold recognition in the program Phyre. *Proteins* 70: 611-625.

## FIGURE LEGENDS

**Figure 1.** Affinity purification of *T. gondii* nuclear factors. (A). A schematic diagram showing the experimental strategy devised for the purification of ENO1 promoter binding proteins. (B) Electrophoretic mobility shift assays (EMSA) showing DNA-protein complexes using the biotinylated DNA sequence corresponding to the ENO1 promoter and total nuclear extract of tachyzoites. Lane 1, unbound biotinylated probe alone. Lane 2, gel shift binding assays revealing the biotinylated DNA-protein complexes. Lane 3, specific competitor corresponding to unlabeled ENO1 promoter introduced simultaneously with labelled probe during binding assays. (C) SDS-PAGE and silver staining after a large scale affinity purification of ENO1 probe bound proteins using a total nuclear extract from *T. gondii* tachyzoites. The markers (MW) indicated in kilodaltons (kDa) are shown on the left.

**Figure 2.** Four randomly chosen candidate factors binding to ENO1 promoter are targeted into *T. gondii* nucleus after transient transfection. The full-length cDNA of 4 candidate factors designated TgNF1, TgNF2, TgNF3 and TgNF4 were PCR amplified and cloned in frame to YFP (panel A) or to HAFLAG (panel B) reporter in two different vectors, which are schematically presented on the top of each panel. The expression of these candidate factors was driven either by heterologous tubulin or dense granule 1 gene promoters. These two *T. gondii* vectors contain a chloramphenicol acetyltransferase (*CAT*) or bleomycin (*Bleo*) cassette whose expression is driven by surface antigen 1 (*SAG1*) or tubulin gene promoters. Direct fluorescence was detected for YFP expression (panel A) whereas indirect fluorescence assay (IFA) was performed using monoclonal antibody specific to the HA tag (panel B). The nuclei of intracellular tachyzoites were stained with the polyclonal antibodies specific to ENO2 or with a classical nuclear dye (DAPI), both known to label the whole parasite nuclear space except the area defining the parasite nucleolus. The scale bars are shown on these figures.

**Figure 3.** Analysis of the TgNF3 sequence, a member of the HD2/FKBP/NPM superfamily. (A) Hydrophobic cluster analysis (HCA) plots of the TgNF3 sequence. The sequence is shown on a duplicated  $\alpha$ -helical net, in which strong hydrophobic amino acids (VILFMYW) are contoured. These form hydrophobic clusters, which mainly correspond to regular secondary structures [38]. The way to read the sequence (1D) and secondary structures (2D) are indicated with arrows. The plot allows the delineation of globular domains (boxed), made of ~33 % of strong hydrophobic amino acids, which are distributed into clusters whose lengths are typical of those of regular secondary structures. Secondary structures indicated below the plot for the first globular domain were deduced from the similarities shared with the NPM family (see panel B). (B) Multiple alignment of the TgNF3/HD2/FKBP/NPM family. The protein alignments were constructed on the basis of data from sequence similarity and fold

recognition searches, and refined using HCA. UniProt identifiers are reported in front of the sequences. The secondary structures observed in the NPM experimental 3D structure (NPM\_XENLA, pdb identifier 1XE0) are shown above the alignment, with the outstanding features of the NPM sequences (A1 tract,  $\beta$ -hairpin) indicated. Conserved hydrophobic amino acids (VILFMYW) are indicated in green (light green for amino acids considered as valid substitutions (ACTS)), aromatic amino acids (FYW) in purple (pink for H, considered as a valid substitution), loop-forming (PGDNS) or small residues (AVC) in yellow. The two conserved amino acids described earlier for the HD2/FKBP family [39] are indicated with blue (H) and red (acidic amino acid) arrows.

**Figure 4. (A)** 3D-structure of the NPM Np core domain, highlighting the positions of hallmarks of the NPM and HD2/nuclear FKBP/TgNF3 families. The NPM Np core domain, as observed in the *Xenopus laevis* No38 (pdb identifier 1xe0) consists in an eight-stranded  $\beta$ -barrel, which is organized as a dimer of pentamers. On this template (two orthogonal views) are indicated the positions of; the A1 acidic tract (pink) and  $\beta$ -hairpin (orange), typical of the NPM family; the FKBP acidic tract (red), typical of nuclear FKBP; the HD2 conserved histidines (blue); and the acidic amino acids (red).

(B) Domain organization of SpFKBP39p and TgNF3. The NPM/HD2 domains in the two proteins are highlighted. An insertion of an acidic region is present in the middle of the yeast FKBP39 NPM/HD2 domain. Numbers indicate amino acid positions. The acidic and basic regions of both proteins are described.

**Figure 5.** Comparison of the expression of *T. gondii* TgNF3 transcript and protein between tachyzoites and bradyzoites. (A) RT-PCR analysis on total RNA prepared from tachyzoites cultivated *in vitro* and encysted bradyzoites isolated from brains of chronically infected mice. After first strand synthesis, the cDNA of TgNF3 as well as the actin cDNA (used as a control) was amplified by two specific primers as described in Materials and Methods. (B) Western blot analysis of the bacterially expressed and purified recombinant TgNF3-GST fusion protein (lane 1), proteolytically cleaved and purified recombinant TgNF3 non-fusion protein (lane 2), total SDS-protein extracts from *T. gondii* tachyzoites of 76K strain (lane 3), from non-infected human fibroblast cells (lane 4) and from non-infected mouse brain cells (lane 5). The immunoblot was incubated with the affinity purified anti-TgNF3 antibodies. (C) Western blot analysis of total SDS-protein extract prepared from purified tachyzoites and bradyzoites. In the upper panel, the immunoblot was incubated with the polyclonal antibodies anti-TgNF3 whereas the lower panel corresponds to duplicated blot loaded with identical amount of total SDS-protein extract analysed above and incubated with a monoclonal antibody anti-actin used to standardise equal amounts of parasite proteins. (D) Over-expression of YFP tagged to TgNF3 protein (TgNF3-YFP) in stably transformed tachyzoites. The chloramphenicol-resistant transfectants were

cloned and total protein extract of one positive clone shown on these blots (lane 7) was compared to that of wild type tachyzoites (lane 6). Note that the 70 kDa-protein, which corresponds to the TgNF3-YFP fusion protein, is exclusively detected in the transformed positive tachyzoites only. This was confirmed by the monoclonal antibody anti-GFP tested on the total SDS-extract protein of TgNF3-YFP over-expressers (lane 9) *versus* wild type tachyzoites (lane 8).

**Figure 6.** Expression of TgNF3 protein in the nucleus of extracellular tachyzoites. (A) The upper panel shows a low level of TgNF3 in the extracellular tachyzoites of the wild type 76K strain detected with the polyclonal antibodies anti-TgNF3. The lower panel corresponds to extracellular transgenic tachyzoites stably expressing higher level of TgNF3-YFP protein. (B) 3D-reconstruction based on a Z-stack confocal acquisition followed by a rotation around the y-axis (panel representing 5 different rotation angles) of other sets of extracellular tachyzoites of 76K wild type and TgNF3-YFP over-expressers. These 3D-confocal scans allow the visualization of the whole bodies and entire volume of parasite nucleus and confirmed that only a tiny amount of TgNF3 was detectable in tachyzoites of the wild type 76K strain whereas its corresponding transgenic over-expressers have a high level of TgNF3-YFP. Z-stacks were generated in increments of 0.6  $\mu\text{m}$  or 1.6  $\mu\text{m}$ .

**Figure 7.** Native TgNF3 and transgenic TgNF3-YFP proteins are preponderant nucleolar resident factors. (A) Confocal image of intracellular dividing tachyzoites stained with the purified polyclonal antibodies anti-TgNF3 and DAPI (upper panel). The lower panel corresponds to direct fluorescence captured by confocal microscopy on the fixed intracellular tachyzoites over-expressing TgNF3, which were also stained with DAPI. Note that the fluorescence signal is mostly present in the nucleolus, which is not stained by DAPI. (B) Time lapse imaging of fluorescence in living dividing tachyzoites, which stably over-express TgNF3-YFP. One vacuole containing four daughter parasites were selected at time 0 (0 h) and constantly followed during intracellular development and parasite replication for 24 h. In these experiments, no nuclear staining dye was used. The blue arrows indicate the presence of one or two smaller nucleoli close to a large nucleolus in the parasite nucleus. (C) Three separate sets of intracellular replicating tachyzoites, which stably over-express TgNF3-YFP in the presence of hydroethidine, a living nuclear dye staining, prior to direct imaging of fluorescence. The panel in red shows the nucleus labelled with HET. The green label corresponds to TgNF3-YFP signal. In the merged pictures, the red arrows indicate nucleolar fragmentation or its *de novo* synthesis during direct fluorescence experiment in which hydroethidine dye has been used to confirm the presence of smaller nucleoli with a large nucleolus in a single given parasite.

**Figure 8.** Ultrastructural analyses show that TgNF3-YFP over-expression enhanced considerably nucleolar size and changed its position. (A) This electron micrograph corresponds to cryo-immunogold detection of one longitudinal section of transgenic tachyzoite TgNF3-YFP over-expresser. The black

dots correspond to TgNF3-YFP detected with the monoclonal antibody anti-GFP. B) An enlarged picture of the parasite area containing the immunogold stained nucleus of the tachyzoite TgNF3 over-expresser shown in panel A. C) Enlarged section of unlabelled nucleus of parental wild type 76K tachyzoite. In addition to differences in the nucleolus size, which are approximately 4-5 folds smaller in the parental 76K tachyzoite relative to that of TgNF3-YFP over-expresser, its position is also more centrally located in the wild type tachyzoite. N=nucleus; Nu, nucleolus; arrow indicates the nuclear membrane; Rh, rhoptries; M, micronemes; DG, denses granules. The scale bar = 500 microns.

**Figure 9.** Over-expression of TgNF3 enhances parasite replication *in vitro* but drastically attenuates parasite virulence *in vivo*. (A) The percentage distribution of vacuole size (number of parasites/vacuole) was determined at 24h after infection with the tachyzoites of the parental 76K strain and transgenic 76K TgNF3-YFP over-expressers. The parasites were grown in human foreskin fibroblast (HFF) cells under classical culture conditions described in Materials and Methods, stained and counted. The number of parasites per vacuole was counted for 500 vacuoles in three different experiments ( $P < 0.001$ ). The relative time-to-death of mice infected with TgNF3-YFP over-expressers and parental wild type 76K strain were compared. A group of 10 females of BALB/c mice (B) or a group of 10 females of CBA/J (C) was separately inoculated with lethal doses ( $10^4$  parasites) of tachyzoites, and mortality was monitored over 6 weeks. All mice infected with TgNF3-YFP survived whereas all of those infected with wild type tachyzoites died. The surviving mice were then challenged at 42 days post-infection either with  $10^5$  tachyzoites from 76K strain or  $10^5$  tachyzoites of the highly virulent RH strain, and monitored for life expectancy and protection. Three independent experiments have been performed ( $n=3$ ) and  $P < 0.001$ . D) A group of mice were infected with  $10^4$ ,  $10^5$  and  $10^6$  TgNF3-YFP over-expressing tachyzoites, respectively. All infected mice survived as above and at 42 days post-infection, the serum of each surviving mouse was tested on total SDS-protein extract of *T. gondii* by Western blots. These experiments were performed twice with identical results.

**Figure 10.** TgNF3 is present exclusively in the cytoplasm of dormant bradyzoites and relocates in the nucleolus during bradyzoite to tachyzoite conversion. A) Superimposition of five confocal images of one representative cyst, which was isolated from the brain of the infected mice analyzed above, stained specifically with the Dolichol biflorus FITC-labeled lectin, which recognizes the cell wall. The fluorescence of bradyzoites over-expressing TgNF3-YFP was also simultaneously visualized (see also Video S3 and S4). Z-stacks were generated in increments of  $0.45 \mu\text{m}$ . B) Z-stack confocal acquisitions of one representative bradyzoite expressing YFP-TgNF3 released from cysts isolated from infected mice brain as described above, fixed and nuclear stained with DAPI. Z-stacks corresponding to 10 confocal images, which were generated in increments of  $0.6 \mu\text{m}$ . Note that the fluorescence signal is restricted to the cytoplasm of the bradyzoites. C) Relocation of TgNF3-YFP in the nucleolus of newly

converted tachyzoites, which are dividing inside human fibroblast cells inoculated by transgenic bradyzoites expressing TgNF3-YFP at 36h post-infection.

**Figure 11.** The presence of TgNF3 in the cytoplasm of dormant bradyzoites is accompanied by profound changes in parasite's nucleus. The bradyzoites released from wild type *T. gondii* 76K cysts isolated from chronically infected mice were formaldehyde-fixed and loaded on glass slides. The bradyzoites were incubated with the purified polyclonal antibodies specific to TgNF3 and processed for confocal imaging. Panels 1 to 12 correspond to 12 individual positive bradyzoites, which confirmed the presence of TgNF3 signal exclusively in the cytoplasm of these dormant bradyzoites. The video S5 and S6 show the rotation at 360° around the y axis, at a rate of 5 frames per second, which demonstrate the complete absence of fluorescence and TgNF3 protein in the nucleus of the encysted dormant bradyzoites; TgNF3 is detected only in the parasite cytoplasm. In addition, note that the nuclei of slowly dividing bradyzoites displayed a remarkable half reduction in size (panels 4, 6 and 7), probably due to nuclear condensation relative to that of rapidly replicating tachyzoites. All bradyzoites analyzed show a remarkable and unexpected lack of nucleolus. Some bradyzoites also contain degenerated nucleus (panels 8, 9 and 10) and few lack the whole nucleus (panels 11, 12). The dramatic reduction in TgNF3 protein level in the cytoplasm of bradyzoites (compare panels 1 and 9, 2 and 10) can be correlated with the degeneration and disappearance of nucleus in the bradyzoite form.

**Figure 12.** TgNF3 associates with core and nucleosome-associated histones *in vitro*. (A) The purified recombinant protein TgNF3-GST or GST alone were coupled to glutathione-Sepharose 4B and the equivalent of 0.5 µg and 2 µg of bound proteins were boiled and electrophoresed on SDS-PAGE, in order to verify that equal amounts of GST and TgNF3-GST bound protein were being used for the pull down assays. B) The quality of purified core histones containing histone H3, H4, H2A/H2B were checked by SDS-PAGE and coomassie blue staining. C) Silver staining of purified core histones of HeLa cells pulled down by TgNF3-GST. Lane 1, input corresponding to core histones; lane 2, TgNF3-GST incubated with core histones; lane 3, competition assay using non-fusion recombinant TgNF3 incubated with core histones prior to pull down with TgNF3-GST beads; lane 4, competition assay using GST alone incubated with core histones prior to pull down with TgNF3-GST beads; 5, GST bound beads incubated with core histones. D) Western blot after TgNF3-GST pull down using HeLa cell core histones and specific anti-histone H3 antibodies. Lane 1, GST alone incubated with core histones; lane 2, TgNF3-GST incubated with core histones; lane 3, competition assay using non-fusion recombinant TgNF3 incubated with core histones prior to pull down with TgNF3-GST beads; lane 4, competition assay using GST incubated with core histones prior to pull down with TgNF3-GST beads; lane 5, 25% of input corresponding to core histones. E) Western blots of total protein extract from purified *T. gondii* nucleosomes. Lane1, *T. gondii* nucleosomes probed with specific anti-TgNF3

antibodies; lane 2, *T. gondii* nucleosomes probed with specific anti-histone H3 antibodies. F) Western blot after TgNF3-GST pull down using *T. gondii* nucleosomes and specific anti-histone H3 antibodies. Lane 1, input corresponding to *T. gondii* nucleosomes; lane 2, the beads containing GST alone were incubated with *T. gondii* nucleosomes; lane 3, the TgNF3-GST beads were incubated with *T. gondii* nucleosomes; lane 4, competition assay using non-fusion recombinant TgNF3 incubated with TgNF3-GST beads prior to pull down with *T. gondii* nucleosomes. G) Western blot after immunoprecipitations using *T. gondii* nucleosomes and anti-TgNF3 antibodies. Lane 1, the anti-GST antibodies were incubated with *T. gondii* nucleosomes; lane 2, the anti-TgNF3 antibodies were incubated with *T. gondii* nucleosomes; lane 3, the anti-TgNF3 antibodies were incubated with non-fusion recombinant TgNF3 prior to addition of *T. gondii* nucleosomes. The immunoblot was revealed with the antibodies specific to histone H3.

**Figure 13.** Chromatin immunoprecipitation (ChIP) analyses of TgNF3 binding sites *in vivo*. (A) The anti-TgNF3 serum from five mice immunized with the recombinant TgNF3 protein was individually processed for ChIP experiments using tachyzoites of wild type 76K strain. Note that the stage-specific ENO1 promoter was specifically precipitated using either the anti-TgNF3 sera (IS1 to IS5) or the monoclonal anti-GFP antibody from chromatin extracts of wild type and TgNF3-YFP over-expresser tachyzoites, respectively. The irrelevant naive mouse serum (NS), which was used as control did not immunoprecipitate ENO1 promoter. (B) Both polyclonal antibody anti-TgNF3 (IS3 serum with the strongest signal, see panel A) and the monoclonal antibody anti-TgNF3 specifically immunoprecipitated ribosomal DNA 18S promoter using chromatin extracts from tachyzoites of wild type 76K strain and transgenic TgNF3-YFP over-expressers, respectively. (C) A chromatin sample with an irrelevant genomic DNA corresponding to the open reading frame (ORF) of ENO1 used as negative control. (D) No chromatin was immunoprecipitated when the tachyzoite-active ENO2 promoter was probed after ChIP assays using anti-GFP and anti-TgNF3 antibodies (IS3 serum). (E) No significant chromatin samples corresponding to bradyzoite-silent ENO1 promoter was immunoprecipitated using three epigenetic (acetylated and methylated) histone marks. The input corresponds to total chromatin extracts, which have not been immunoprecipitated by antibodies but were directly PCR amplified by the same set of primers tested. (F) Chromatin samples corresponding to the tachyzoite-active ENO2 promoter were specifically amplified when the same three distinct epigenetic histone marks were tested using ChIP assays. (G) Model of TgNF3 functions during promoter activity in *T. gondii*. TgNF3 binding to chromatin *in vivo* is specifically restricted to tachyzoite-silent ENO1 promoter whose core histones are not acetylated whereas the tachyzoite-active ENO2 promoter with epigenetic active marks through acetylated core histones does not interact with nuclear TgNF3. We propose that TgNF3 may be a *T. gondii* nuclear chaperone involved in alterations

in chromatin structure and function. This supports the notion that changes in chromatin structure and function may link the nuclear TgNF3 functions to the control of gene regulation in *T. gondii*.

#### **SUPPORTING INFORMATION LEGENDS**

**Figure S1.** Nucleotide sequence of the promoter of *T. gondii* bradyzoite-specific ENO1 gene used to purify novel nuclear factors after biotinylation and affinity purification. The underlined nucleotides have been previously identified as motifs that specifically bind to nuclear extract using gel retardation assays [31].

**Supplementary Table 1.** Protein identified by mass spectrometry using *nano* LC-MS/MS. The proteins, whose nuclear localization has been validated by reverse genetics using a yellow fluorescence protein tagged version or specific polyclonal antibodies, were designated TgNF for *T. gondii* nuclear factors.

**Video S1.** 3D-confocal reconstruction of TgNF3 signal in the wild type or parental parasites expressing native TgNF3 protein using specific polyclonal antibodies. Ten confocal images generated in increments of 0.6  $\mu\text{m}$ .

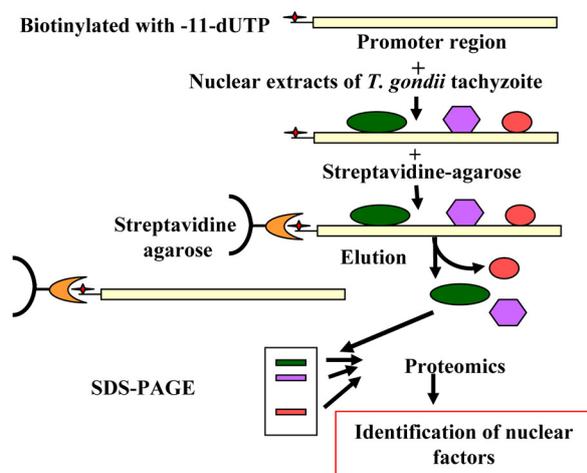
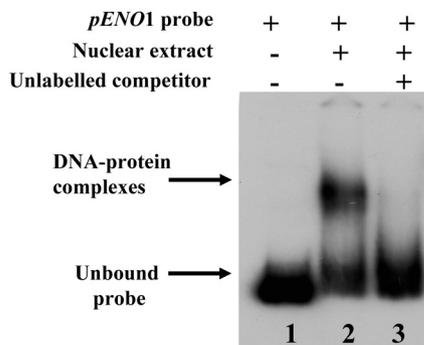
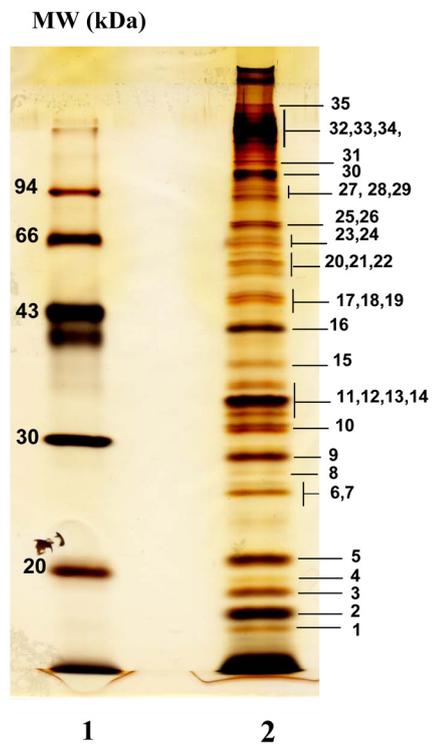
**Video S2.** 3D-confocal reconstruction of TgNF3-YFP signal in parasite over-expressing TgNF3 protein. Ten confocal images generated in increments of 0.8  $\mu\text{m}$ .

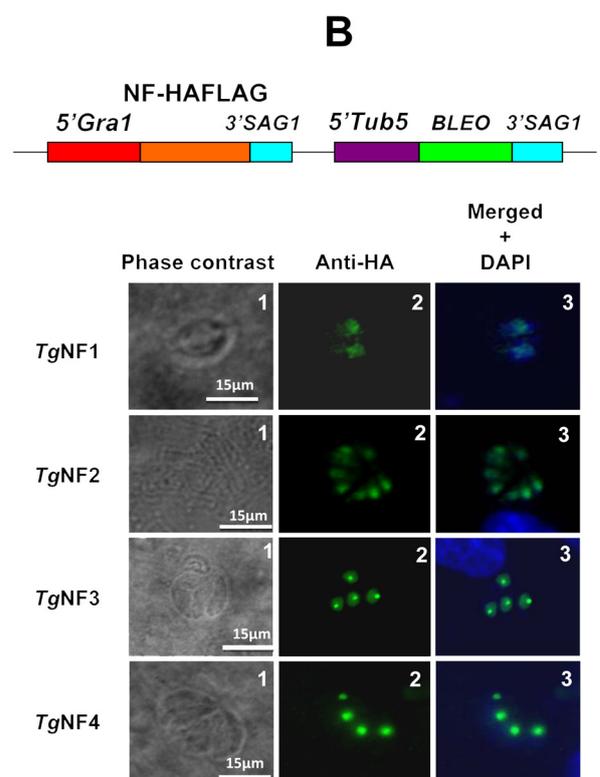
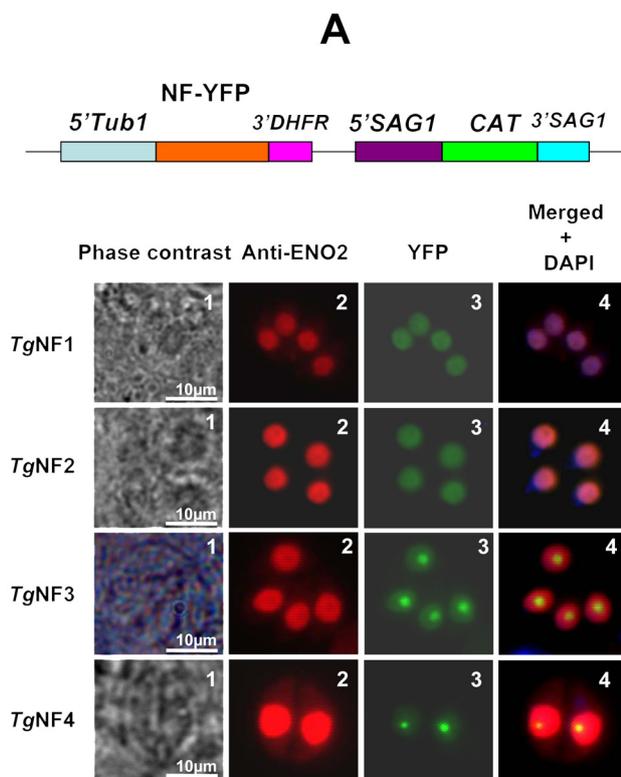
**Video S3.** Video representing the displacement along the z-axis (confocal Z-stack acquisition) created with the Zen software, at a rate of 5 frames per second showing the presence of TgNF3-YFP in bradyzoites (green dots appearing during the movie) present inside the cyst whose wall was stained with *Dolichos biflorus* lectins. The video was generated with 41 images in increments of 0.45  $\mu\text{m}$ .

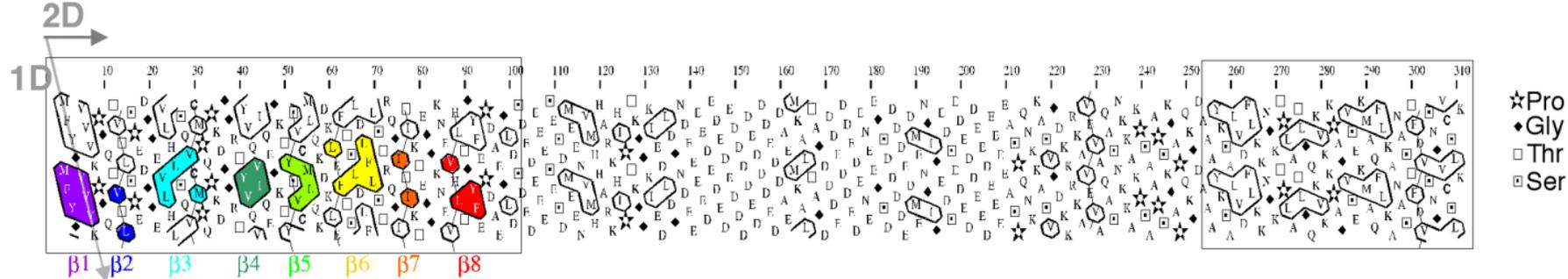
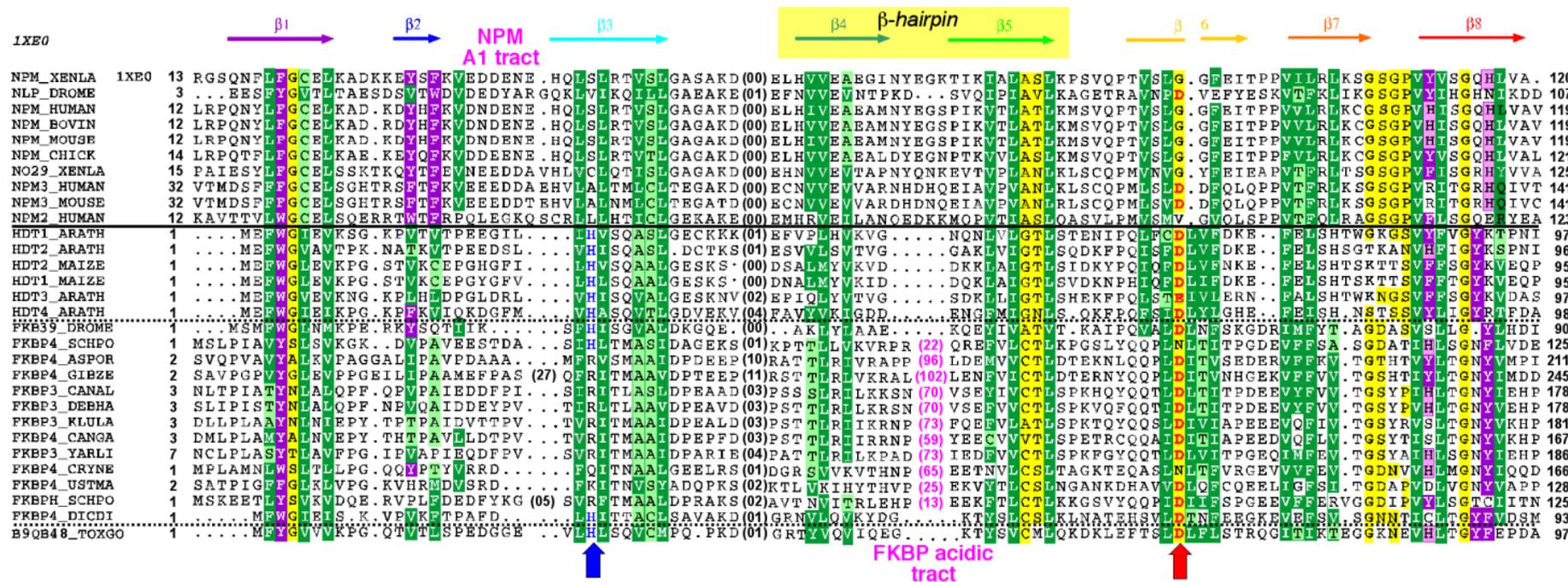
**Video S4.** Video representing the displacement along the z-axis (confocal Z-stack acquisition) created with the Zen software as above and showing the presence of TgNF3-YFP in bradyzoites (green dots appearing during the movie) inside the cyst whose wall was not stained with *Dolichos biflorus* lectins. The video was generated with 24 images in increments of 0.45  $\mu\text{m}$ .

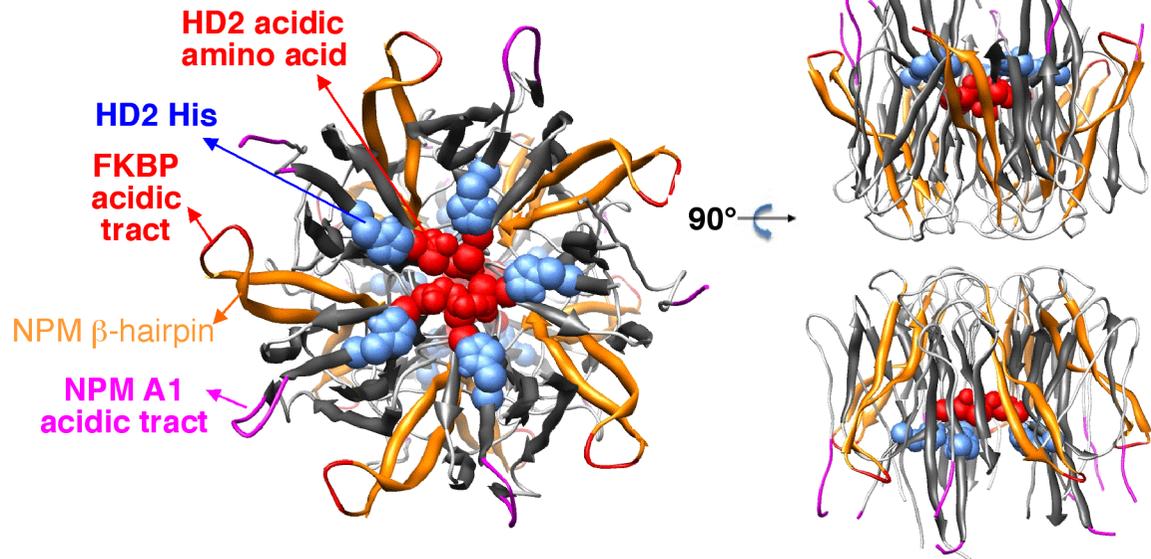
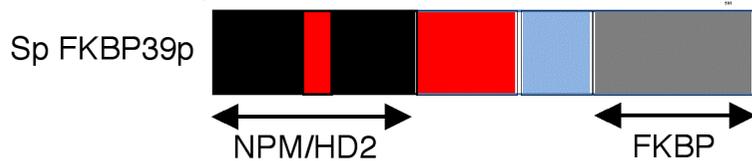
**Video S5.** 3D-confocal reconstruction of panel 1 from Figure 13 showing that native TgNF3 signal is exclusively localized to the cytoplasm of the bradyzoites. The video was generated with 7 images in increments of 0.8  $\mu\text{m}$ .

**Video S6.** Another example of 3D-confocal reconstruction of panel 2 from Figure 13 is shown. It also shows the presence of native TgNF3 signal only in cytoplasm of the bradyzoites. The video was generated with 7 images in increments of 0.8  $\mu\text{m}$ .

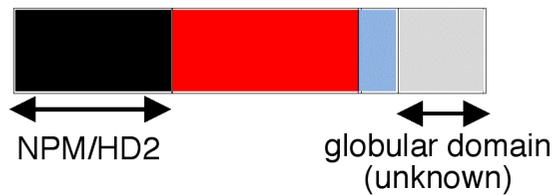
**A****B****C**



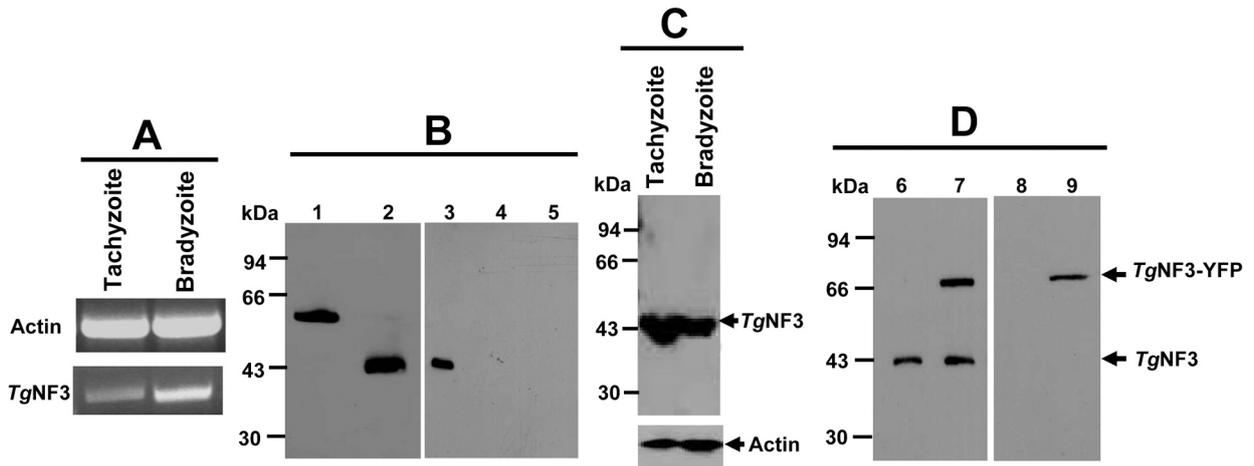
**A****B**

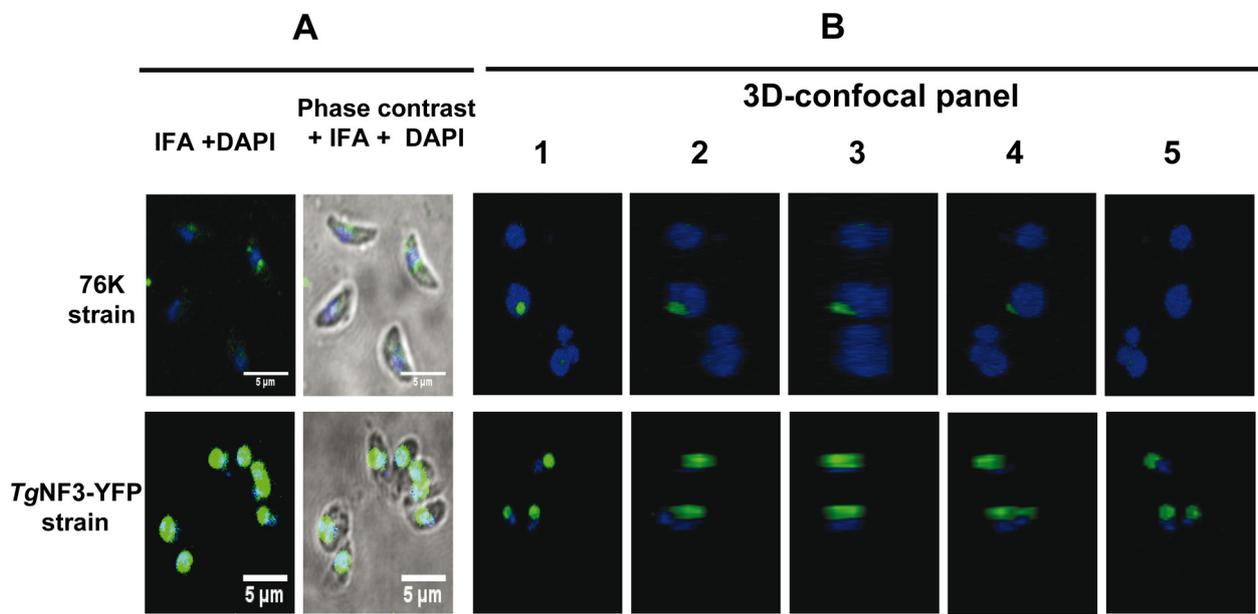
**A****B**

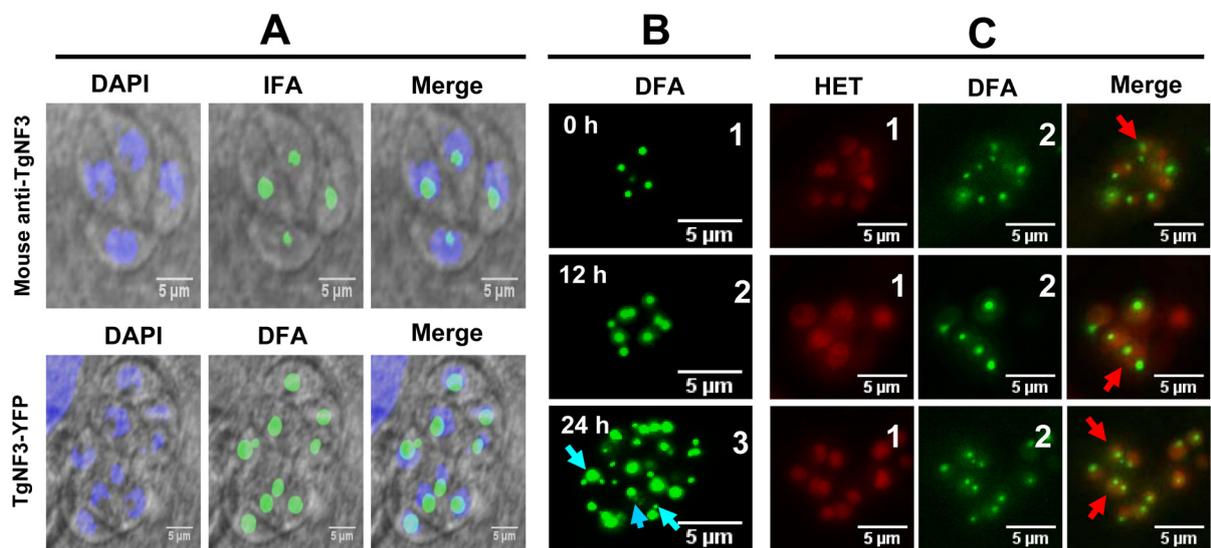
Tg NF3

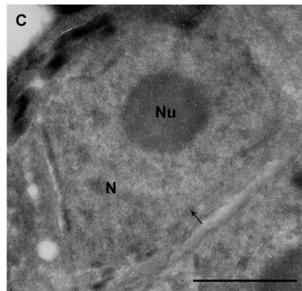
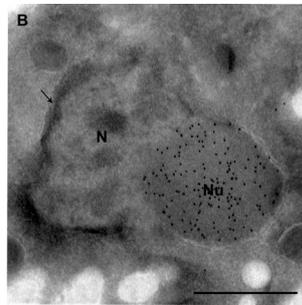
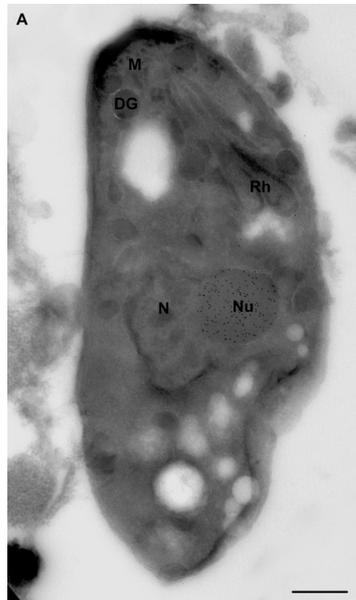


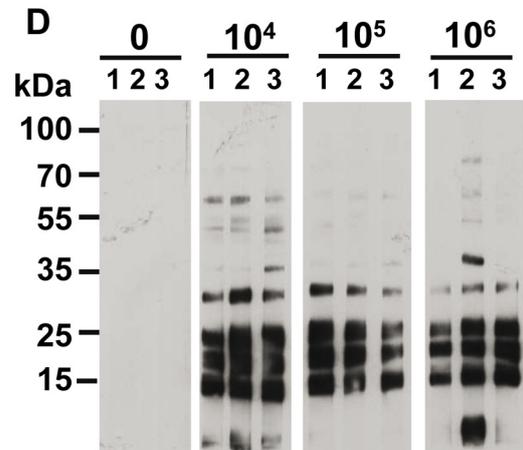
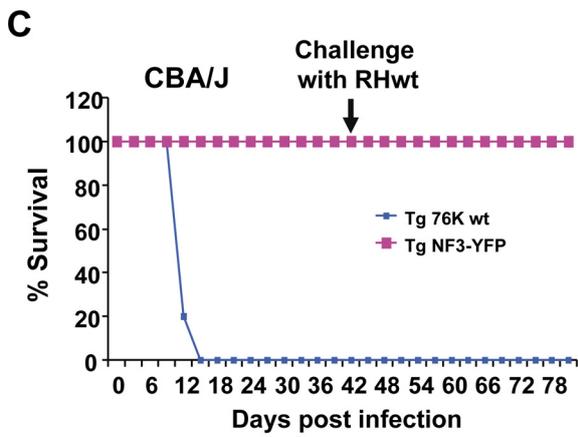
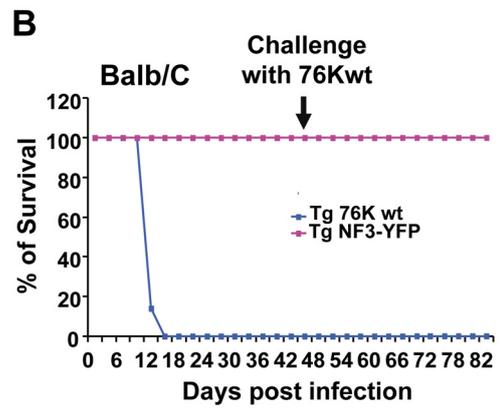
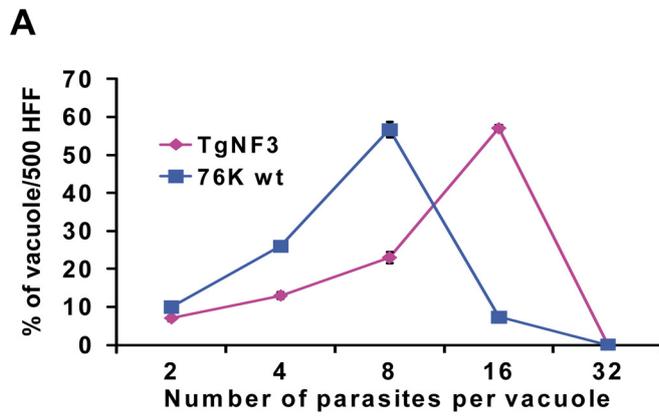
Acidic region  
Basic region

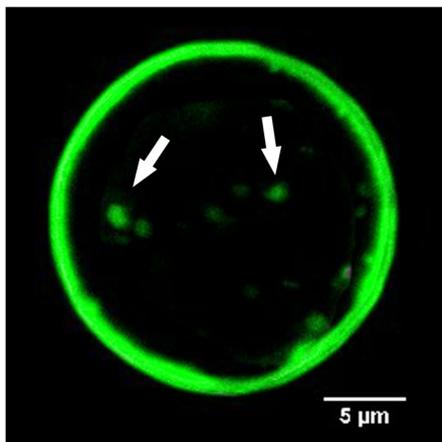
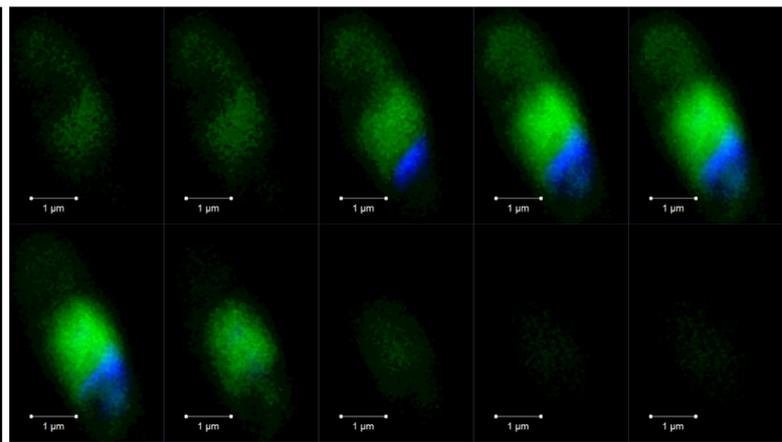
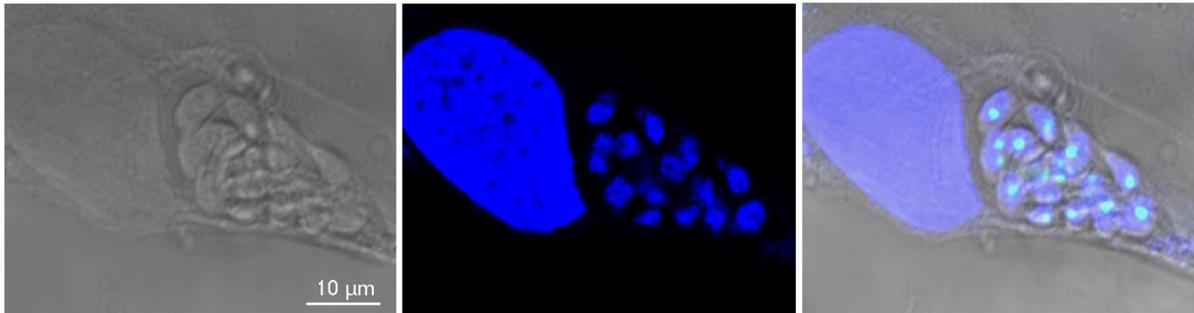


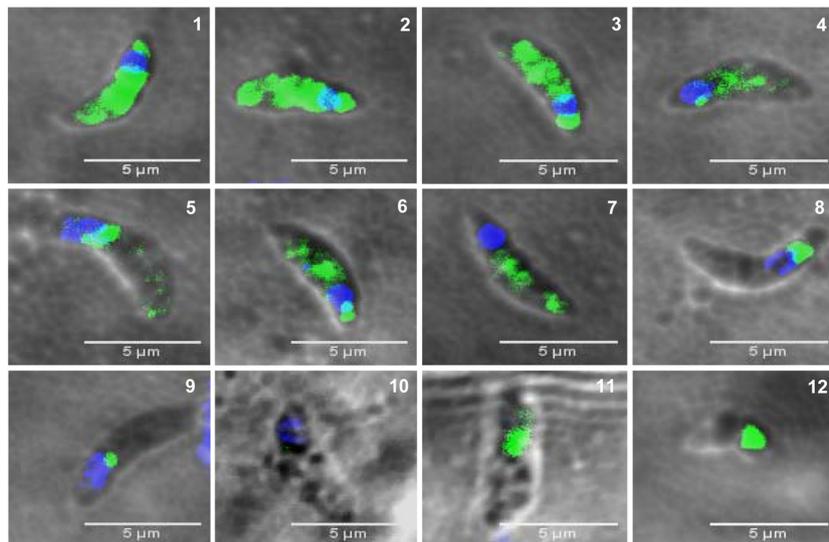


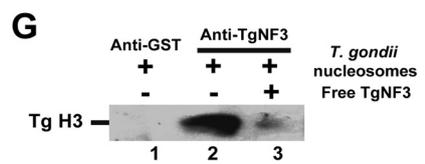
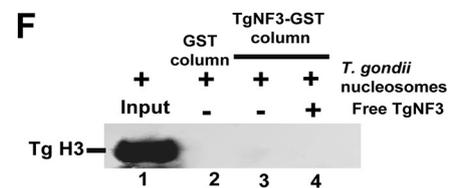
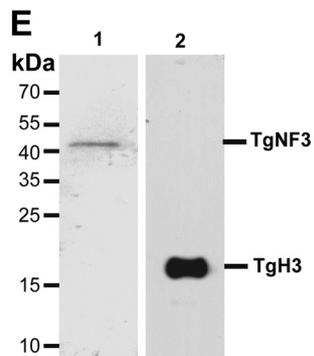
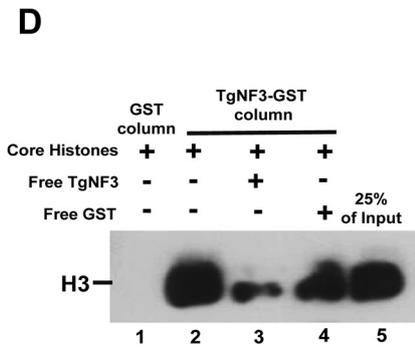
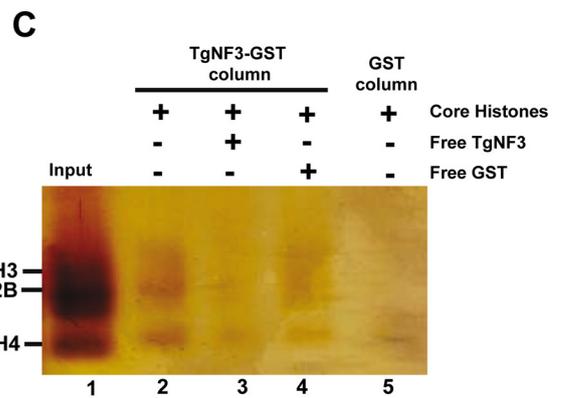
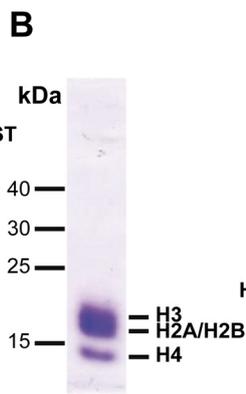
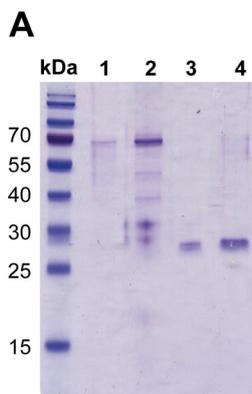


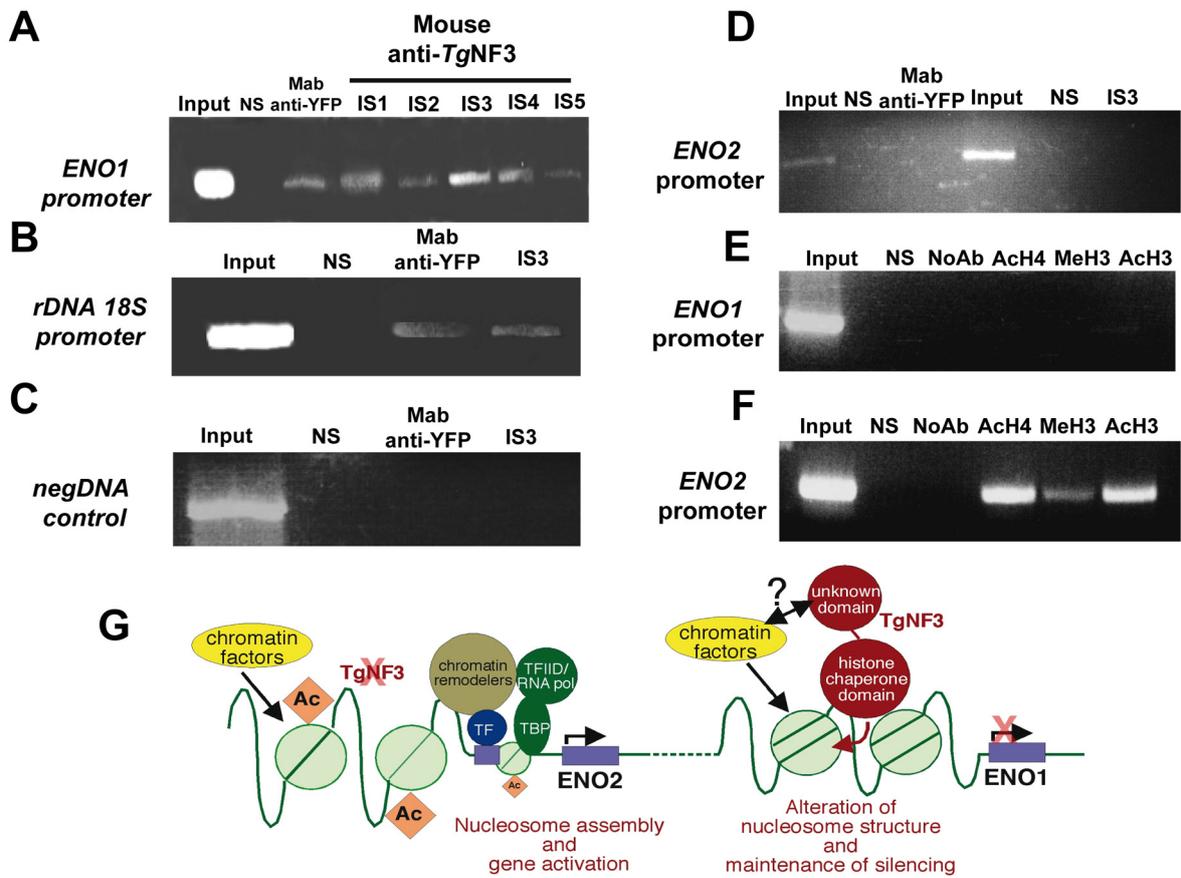




**A****B****C**







Supplementary Table 1

T. gondii nuclear factors	Protein name	Protein accession number	Protein molecular weight (Da)	Number of unique peptides	Percentage sequence coverage	Peptide sequence	Nuclear localization using YFP tagging and/or specific antibodies	Slice number
<b>Nuclear functions</b>								
<b>TgNF1</b>	DEAD/DEAH box helicase, putative [Toxoplasma gondii ME49]	gi 211966692	59 951,3	3	7.45%	DMIGIAETGSGK SGGFGGGNALPLGSRPY VTYLMDEADR	<b>YES</b>	<b>21</b>
<b>TgNF2</b>	pinin/SDK/memA/ domain-containing protein [Toxoplasma gondii ME49]	gi 211966969	36 006,7	9	40.9%	AMCTELDTSEPR ESSEVKAEGETVAAGEDER LEEVSTELSNK LEEVSTELSNKENELMR MFGALAGHLQR NMEVFIGTEAQPTLFWAPAK QLLMEQSTDLAHK RPPGELFSLDKPEASAPSLPQK VTYSFEGEEFQVEK	<b>YES</b>	<b>16</b>
<b>TgNF3</b>	46 kDa FK506-binding nuclear protein, putative [Toxoplasma gondii ME49]	gi 211962881	33 952,5	7	21.9%	ANASTFDVSNQVCSLK DKLEFTSLDLFLSTR TPLATLGQK TYSVCMLQK TYVQVIQEGK SALVDFLK SALVDFLKK	<b>YES</b>	<b>16</b>
	SWIB/MDM2 domain-containing protein [Toxoplasma gondii ME49]	gi 211963658	27 990,4	3	12.2%	DQVSMFELNK HIWDYIK MINADSTLRPLFQK		<b>1</b>
	Gbp1p protein, putative [Toxoplasma gondii ME49]	gi 211963026	31 742,6	4	17.1%	ADVFEFQGR ADVMTMPDGR GCGIVEYTNVEDAQK GVGTVLFSTPEGAQR		<b>9</b>
	Nucleolar protein family A, putative [Toxoplasma gondii ME49]	gi 211963509	20 913,6	2	16,20%	EEIGKVDEILGPINEMLFSVR GTLSQQVPPYFNDR		<b>9</b>
	RRM domain-containing protein [Toxoplasma gondii ME49]	gi 211966681	53 421,5	3	8,77%	QEMEFAAQGGDPR TTGEAYVQLPDPGLR TVMAAISGNTDSDLASR		<b>17-20</b>

## S1 (continued)

T. gondii nuclear factors	Protein name	Protein accession number	Protein molecular weight (Da)	Number of unique peptides	Percentage sequence coverage	Peptide sequence	Nuclear localization using YFP tagging and/or specific antibodies	Slice number
	Centromere/microtubule binding protein, putative [Toxoplasma gondii ME49]	gi 211968418	56 726,7	10	28.9%	ALETLTGALFQRPPVVAVK DHLVTMHDVLDAMHVVYDSTR DSCVNAICYGAK EATSDPGLVASAGTEGGK KTDNYAIEPNSSSPIDTSK LCGSEETENGEATEAAVKEEQPR LLDYDEQR MKEATSDPGLVASAGTEGGK RTIFPLELLLTGFPR TPASWLK		23-24
	Nucleolar phosphoprotein nucleolin, putative [Toxoplasma gondii ME49]	gi 211966199	73 100,9	3	6,10%	GIAFITFETEEAAQK GQAFIEFDSIESATK SLFEECGEIQEVR		28
	krr1 family, zinc finger-containing protein [Toxoplasma gondii ME49]	gi 211968343	99 371,4	2	4,00%	AFVDAAEQVEEGAFQK ALADDVAYAPLAEDEER		32
	sec63 domain-containing DEAD/DEAH box helicase, putative [Toxoplasma gondii ME49]	gi 211963689	247 618,1	19	12.7%	ALVAEQVQAFSQR ASNRYPVVNVFQVSK DTSTGVAPAVHLGEPTEGESLAGR DVGDWLGVSSSGLFNHPSVR EILQSEAEAVK GATVLDADVTEIFFYKPTTQQTR GLFFFGNHYPVPLK GTQVYLPEK IVAMANSLANAK LELAAFVQPLTR LFSLSSEFK LTAIDLLLQAAADDDPKK LVEQSQLILATPENWDFVSR LVGLSATLPNYDDVAVCLR LVIIDEIHLLHDAR MEQVYSGSFK WLLSESSLPISFK YENFELAK YNTMNEVTYEK		35

## S1 (continued)

T. gondii nuclear factors	Protein name	Protein accession number	Protein molecular weight (Da)	Number of unique peptides	Percentage sequence coverage	Peptide sequence	Nuclear localization using YFP tagging and/or specific antibodies	Slice number
<b>Enzymatic and chaperone functions</b>								
	Receptor for activated C kinase, RACK protein, putative [Toxoplasma gondii ME49]	gi 211968660	35 128,7	10	43.9%	DGVAMLWDVNEGK FSPSANKPLIVSCGWDK GVLEGHTDCVTAISTPSLK IWDLENK LWDLNAGVTVR LWNTLAECK NVLSEITPEK SFQGHTSDVNSVAFSPDNR SGAPWCTSLNWSHDGR VIVWTLNDNPDDSGSVGYAR		12-13
	Acid phosphatase, putative [Toxoplasma gondii ME49]	gi 211968360	46,589.6	3	11.6%	LAPADATEAAAAENHGYPK LVSGTTGETLYTHK WQSEFENVYSDANGALK		18
	NADP-specific glutamate dehydrogenase, putative [Toxoplasma gondii ME49]	gi 211967784	53 642,5	5	13,10%	AANAGGVAVSGLEMSQNAMR DVPAGDIGVGAR GFPLGQIQR IVAEGANMPTTR VQYSSAAGPYK		18
	delta-aminolevulinic acid dehydratase, putative [Toxoplasma gondii ME49]	gi 211967109	71 357,4	3	6,23%	AGADAVATYYAK AGADMVCPDMMMDGR GYISEKDTVLEVLK		23

## S1 (continued)

T. gondii nuclear factors	Protein name	Protein accession number	Protein molecular weight (Da)	Number of unique peptides	Percentage sequence coverage	Peptide sequence	Nuclear localization using YFP tagging and/or specific antibodies	Slice number
	Heat shock protein 70, putative [Toxoplasma gondii ME49]	gi 211963641	72 276,2	19	35.7%	AIQEALDWLDK ATAGDTHLGGEDFDNR DAGTIAGLSVLR EIAEAYLGK EVESVCTPIITK FDDPSVQSDMK FEELCMDYFR IQQLITDFFNGK KFDDPSVQSDMK LVDFCVQDFK NDAVEIANDQGGR NPENTIFDAK NQLAEKEEFEAK SINPDEAVAYGAAVQAAILK SNQITITNDK SQTFTTYADNQPGVLIQVYEGER SVSEVVLVGGSTR TTPSYVAFTDTER VIAGPGDKPLIEVTYQGEK		25-26-30
	Heat shock protein 70, putative [Toxoplasma gondii ME49]	gi 211962068	73 236,7	17	33.4%	DFFNGKEPNR DKDLLPYEIIINK DVEAVCNPIISK ETAEQFLGK FEELNSDLFQK IINEPTAAAAYGLDK ITPSYVAFTDDDR ITPSYVAFTDDDRK LTPEEIER MKETAEQFLGK NALEGYLHSMK NAVVTVPAYFNDAQR NEATINPTNTLFDVK NGILSVSAVDK VDIIPNDQGGR VEVENLMEGVDFSETLTR VTEAQEWLNTNPDADAEETR VYGQSGGPGAGGAAGGADDDDYGGHDEL		26

## S1 (continued)

T. gondii nuclear factors	Protein name	Protein accession number	Protein molecular weight (Da)	Number of unique peptides	Percentage sequence coverage	Peptide sequence	Nuclear localization using YFP tagging and/or specific antibodies	Slice number
	p97 protein [Toxoplasma gondii ME49]	gi 211969256	101 140,5	12	21.8%	AHPGPLAATLADPEGWALYR FAEWLEQQQQNPVEK GNTTVPATYAEWMK GPLHEVLQDPEGWK GPLTTVLADPDGWDLYK GQPVPDITYEEWR IEATYEEWMAK KAHEGPLEQLADPEGWK MVSEAPGVPSSPSTEVR NMVSSGASIEQTFEEWAQSK NVVDVPQSFEDFLQQPDQQK YEDWVSHPTVR		28
<b>Unknown functions</b>								
<b>TgNF4</b>	Hypothetical protein TGME49_005580 [Toxoplasma gondii ME49]	gi 211965453	51 203,0	11	26.6%	AFAVAPESAK EKPVPVYLTLQGTGAIK ETEEPEFDSAAMR GVPMVIYER KPEENANTEEKQEDGTGGATSPAVSNLK LEIPGFGLVSGGVDLVIK LHLFTMGR QEDGTGGATSPAVSNLK TRPPFAVDGR TVLVHLLR YGWCTAESR	<b>YES</b>	<b>23-24</b>
<b>TgNF5</b>	Hypothetical protein, conserved [Toxoplasma gondii ME49] ALBA 1	gi 211967631	27 211,5	7	25.1%	AVTLAEVLK AVTLAEVLKR DKEPIGSNEIR DPGYQAPIDESLVK EMKPEEITK LLTEQNMR VMNYVAYAAR	<b>YES</b>	<b>13</b>
<b>TgNF6</b>	Hypothetical protein, conserved [Toxoplasma gondii ME49] ALBA 2	gi 211968320	15 374,2	5	33.8%	DDKPACDEVIITGLGMATK TAIGASLLQR TPEFAEQIK TPEFAEQIKK VETSYFSSTR	<b>YES</b>	<b>2-4-5</b>

## S1 (continued)

T. gondii nuclear factors	Protein name	Protein accession number	Protein molecular weight (Da)	Number of unique peptides	Percentage sequence coverage	Peptide sequence	Nuclear localization using YFP tagging and/or specific antibodies	Slice number
TgNF7	Hypothetical protein TGME49_048810 [Toxoplasma gondii ME49]	gi 211964891	54 264,3	9	22.5%	DMSATEYELVVLK	?	21
						EGHMVVGDESAVITLK		
						FVGPAEEEVER		
						IISDIPEAEK		
						MFQDWLSVYATADR		
						SPHVAVGLSGGER		
						TDMCLFVK		
						VLPSSISQADWER		
						YRNPLAESSEEAK		
	Hypothetical protein TGME49_013050 [Toxoplasma gondii ME49]	gi 211968718	19 684,7	2	12,40%	EGVQSAANITEK VANLQQAMR		2
	Hypothetical protein, conserved [Toxoplasma gondii ME49]	gi 211968320	15,374.2	2	12.8%	TPEFAEQIK VETSYFSSTR		4
	Hypothetical protein TGME49_028630 [Toxoplasma gondii ME49]	gi 211964139	24,254.9	2	12.6%	GTLQATGASSPISSQK SSWGWSTPPDQWK		4
	Hypothetical protein TGME49_007440 [Toxoplasma gondii ME49]	gi 211967222	34 765,4	9	28.9%	EVMMIVMQR FEAGNMAMVTGGHNVGR GVPALITHDGR IAAPHHWMLDK ISNVFVIGK ISNVFVIGKGEK LRYALTYR TDQCYPAGFMDVISIEK VRTDQCYPAGFMDVISIEK		10
	Hypothetical protein, conserved [Toxoplasma gondii ME49]	gi 211967516	56 644,5	2	7,25%	GDYGGYGGPGSSSGGPQSDR TGAASTQAGSDAATMNSTR		15
	Hypothetical protein [Toxoplasma gondii RH]	gi 95007135	83 567,9	4	7.99%	EVASMEGASPAAFQAAQQR GAFLGAGATGIVTELIEK KLVDEHPFFQVVEDA LLLDPTHETR		16
	Hypothetical protein TGME49_107810 [Toxoplasma gondii ME49]	gi 211969572	44 152,9	2	7,80%	EAMTVEELQAAVHEAR GQDANAPESAAETAPPR		19
	Hypothetical protein TGME49_032440 [Toxoplasma gondii ME49]	gi 211965778	54,027.8	3	7.05%	LGFSSTEEIDKR SRLDELQAQMDK YQAFLAQQR		20

## S1 (continued)

T. gondii nuclear factors	Protein name	Protein accession number	Protein molecular weight (Da)	Number of unique peptides	Percentage sequence coverage	Peptide sequence	Nuclear localization using YFP tagging and/or specific antibodies	Slice number
	Hypothetical protein TGME49_040060 [Toxoplasma gondii ME49]	gi 211964299	88 515,2	4	8.30%	AVPSGEQPPPQPPPHSPPR GASTASLHSEPGAASMPGSPK GFDRPGTTGTTGPQVR KSPPLPLQEQGR		21
	Hypothetical protein TGME49_112490 [Toxoplasma gondii ME49]	gi 211962135	51 601,3	2	7,51%	SDLLSLLATDEDFYAEDR SSYVSEMILGEPDSTVR		22
	Hypothetical protein TGME49_055350 [Toxoplasma gondii ME49]	gi 211962527	66 010,3	4	8.01%	HMFETEIR SGIDQAAEGGGWHGSK TSMIVTTDAVEAR VQSDPFAPPSR		26
	Hypothetical protein TGME49_051170 [Toxoplasma gondii ME49]	gi 211965042	75 621,8	6	14.7%	ALLSSASGQVPEPLR EHPPPATGGPLPQR ELYVELGVAER IAEYMLEEHNPNPSSSEIR QQGQDAGEGTAATGPSAPSVSSASPMSR YTPVPTTTPPYPADVER		27-28-29
	Hypothetical protein TGME49_030930 [Toxoplasma gondii ME49]	gi 211965658	34,572.5	2	7.01%	LQQNALLSGR VQQAVAEIVEEK		28
	Conserved hypothetical protein [Toxoplasma gondii GT1]	gi 221485011	29,050.2	2	13.6%	LESMASEIAAK LLAAGVDLSHVPGDASLTPGASDQGPR		29
	Hypothetical protein TGME49_089380 [Toxoplasma gondii ME49]	gi 211966037	100 228,9	5	6.62%	ATVDYLDELGK AVAELAATNAGTEK GALPASAPVR QAVAATETTPFSSAQK VMFVGDSAER		29
	Hypothetical protein, conserved [Toxoplasma gondii ME49]	gi 211968066	142 579,4	5	5.96%	AAEQDAVGPALQASR GPSSGGGAAPVVLR IVILAETGSTAEDVTNR LQGEIEEGEIEEGEGEGQR VTLPAEGEGPR		31-33
	Hypothetical protein, conserved [Toxoplasma gondii ME49]	gi 211964776	245 852,7	3	1.87%	AVVYLQSQQAESAQYR GPLSPPYGSQR IEEGEEGGFHFYR		33



## Chapitre 4

# Application de la méthodologie protéomique développée à l'étude des protéines des vins de Champagne

---

Cette étude a fait l'objet d'une collaboration avec les docteurs Clara Cilindre et Sandrine Jégou du Laboratoire d'Œnologie et de Chimie Appliquée de l'Université de Reims.

L'objectif de ce travail était d'étudier les protéines des vins de Champagne et plus particulièrement celles qui sont impliquées dans leur moussabilité. L'effervescence des vins mousseux est en effet, un important critère de qualité pour les consommateurs. Bien que les protéines soient présentes en faible quantité [Marchal et al. 1997], elles font partie des composés essentiels du vin et contribuent à sa qualité [Waters et al. 1998, Ferreira et al. 2001]. Elles sont issues principalement de la vigne et des levures [Kwon 2004], et une partie d'entre elles sont glycosylées comme l'invertase de raisin par exemple [Dambrouck 2004].

Les protéines contenues dans les vins de Champagne sont reconnues comme ayant entre autre une influence sur leurs propriétés moussantes [Brissonnet et Maujean 1993, Clara Cilindre et al. 2007]. Cependant, l'identité des protéines impliquées reste encore à élucider. La moussabilité du Champagne est également dégradée en cas d'infection de la vigne par *Botrytis cinerea* [Marchal et al. 2001].

Nous avons, dans un premier temps, réalisé une étude protéomique globale afin d'identifier les protéines différentiellement exprimées entre un vin issu de raisins contaminés par le champignon *Botrytis cinerea* et un vin sain. Notre travail s'est ensuite focalisé sur la caractérisation des glycosylations de l'une de ces protéines : l'invertase vacuolaire de raisin.

# 1. Analyse protéomique différentielle des vins de Champagne contaminés par *Botrytis cinerea*

## 1.1. Contexte de l'étude.

*Botrytis cinerea* est un champignon qui s'attaque aux feuilles de vigne et aux raisins et qui provoque ce que l'on nomme la pourriture grise. Cette infection, qui peut atteindre 15 à 25% des pieds de vigne, entraîne une perte de rendement et une dégradation de la qualité du vin en provoquant une oxydation de sa couleur et l'apparition de notes de "moisie" et de "pourrie" à la dégustation.

Plusieurs études ont montré que la contamination de la vigne par *B. cinerea* provoque une diminution de la moussabilité des vins de Champagne [Marchal et al. 2001, Clara Cilindre et al. 2007]. Afin d'apporter des réponses pour mieux comprendre les mécanismes biologiques impliqués, nous avons réalisé la première étude protéomique des vins de Champagne. Pour cela, nous avons effectué une étude protéomique différentielle des protéines contenues dans le vin issu de raisins contaminés ou non par *B. cinerea*. Ce travail a fait l'objet d'une publication qui se trouve en annexe 3.

## 1.2. Stratégie protéomique

### 1.2.1. Préparation d'échantillon

L'ensemble des informations concernant la préparation d'échantillon et les expériences effectuées par nos collaborateurs (gel 2D, western blot, etc.) est décrit dans la partie matériel et méthode de la publication située en annexe 3.

Pour obtenir du vin de Champagne, le jus de raisin doit subir deux fermentations successives. Le vin obtenu à l'issue de la première fermentation est appelé le vin de base. C'est ce vin qui est utilisé pour les expériences réalisées au cours de cette étude. Les extraits protéiques issus des vins de base sain et contaminé sont chacun séparés sur un gel d'électrophorèse bidimensionnelle. Les protéines sont révélées à l'aide d'une coloration au bleu de Coomassie. Les deux gels obtenus sont comparés à l'aide d'un logiciel d'analyse d'image (PDQuest) et les spots différentiels sont découpés pour l'analyse protéomique.

### 1.2.2. Analyse protéomique

Après réduction et alkylation, les protéines sont digérées dans le gel avec de la trypsine. Les peptides extraits du gel sont ensuite analysés par nanoLC-MS/MS avec un système chromatographique nanoLC-Chip/MS (Agilent Technologies) couplé à une trappe ionique HCT Ultra

(Bruker Daltonics). Les données obtenues sont interprétées à l'aide du moteur de recherche Mascot en effectuant une recherche dans la banque protéique NCBI nr.

Les informations détaillées concernant ce travail sont disponibles dans la partie matériel et méthode de la publication située en annexe 3.

### 1.3. Résultats publiés

L'analyse différentielle des gels 2D correspondants aux vins sains et contaminés a permis de montrer que 75% des spots visibles sur le gel 2D correspondant au vin sain disparaissent dans le gel correspondant au vin botrytisé, et 9 spots sont uniquement visibles sur ce dernier gel. 37 spots différentiellement exprimés ont été sélectionnés (28 pour le vin sain et 9 pour le vin contaminé) puis analysés par nanoLC-MS/MS.

La majorité des protéines identifiées proviennent du raisin et peuvent être classées en deux groupes : les protéines impliquées dans le métabolisme saccharidique et les protéines de défense (PR). Les autres protéines identifiées proviennent de levures ou sont d'origine fongique.

L'invertase vacuolaire de raisin a été identifiée dans de nombreux spots. Cette protéine est impliquée dans la conversion du saccharose en glucose et en fructose mais aussi dans le développement et la croissance des plantes. L'invertase est principalement observée dans le gel correspondant au vin sain (16 spots) et très peu dans celui du vin contaminé (3 spots). La présence de cette protéine dans de nombreux spots s'explique par sa grande hétérogénéité. Elle possède en effet 12 sites potentiels de glycosylation et sa dégradation lors de la fermentation alcoolique [Okuda et al. 2006] génère des fragments de plus bas poids moléculaire.

De nombreuses protéines impliquées dans la défense des plantes ont également été identifiées comme la thaumatin-like et l'osmotin-like qui sont partiellement voire complètement modifiées au cours de l'infection par *B. cinerea*, mais aussi des chitinases et des endochitinases. Enfin, deux enzymes (endoPG4 et PME) sécrétées par *B. cinerea* ont également été observées pour la première fois dans un vin vinifié à partir de raisin contaminé par ce champignon.

### 1.4. Conclusion et perspectives

Ce travail constitue la première étude protéomique visant à identifier les protéines impliquées dans les propriétés moussantes des vins de Champagne. Cette étude préliminaire a permis de proposer une liste de protéines candidates. Dans un deuxième temps, ces protéines doivent être caractérisées afin de mieux comprendre leur rôle et leur implication dans les mécanismes moléculaires impliqués dans la moussabilité des vins de Champagne. C'est pourquoi, nous avons poursuivi cette étude en ciblant notre travail sur la caractérisation de l'une de ces protéines : l'invertase vacuolaire de raisin.



## **ANNEXE 3: Publication**

### **Proteomics Approach to Identify Champagne Wine Proteins as Modified by *Botrytis cinerea* Infection**

Clara Cilindre, Sandrine Jégou, Agnès Hovasse, Christine Schaeffer, Antonio J. Castro, Christophe Clément, Alain Van Dorsselaer, Philippe Jeandet, and Richard Marchal

J. Prot. Research, 2008, 7 (3), 1199-1208



## Proteomic Approach To Identify Champagne Wine Proteins as Modified by *Botrytis cinerea* Infection

Clara Cilindre,<sup>\*,†</sup> Sandrine Jégou,<sup>†</sup> Agnès Hovasse,<sup>‡</sup> Christine Schaeffer,<sup>‡</sup> Antonio J. Castro,<sup>§,||</sup> Christophe Clément,<sup>§</sup> Alain Van Dorsselaer,<sup>‡</sup> Philippe Jeandet,<sup>†</sup> and Richard Marchal<sup>†</sup>

Laboratoire d'Oenologie et Chimie Appliquée, Université de Reims, URVVC UPRES EA 2069, BP 1039, 51687 Reims Cedex 2, France, Laboratoire de Spectrométrie de Masse Bioorganique, Université de Strasbourg, UMR CNRS-ULP 7178, F-67087 Strasbourg, France, and Laboratoire Stress, Défenses et Reproduction des Plantes, Université de Reims, URVVC UPRES EA 2069, BP 1039, 51687 Reims Cedex 2, France

Received July 7, 2007; Accepted November 9, 2007

The presence of the fungal pathogen, *Botrytis cinerea*, in the vineyard causes reductions in both quality and quantity of grapes and wine. Because proteins are involved in the foam stabilization of sparkling wines, we have undertaken, for the first time, a thorough proteomic analysis of two champagne base wines prepared with either healthy or botrytized Chardonnay grapes, using two-dimensional electrophoresis (2DE) coupled with immunodetection and tandem mass spectrometry. Most of the identified proteins were from grape origin: invertase and pathogenesis-related (PR) proteins. The disappearance of numerous grape proteins was observed in the botrytized wine, suggesting that they were probably degraded or even repressed or the result of a differential expression of grape proteins upon fungal infection. On the other hand, two pectinolytic enzymes secreted by *B. cinerea* were found in the botrytized wine.

**Keywords:** Champagne • wine • two-dimensional gel electrophoresis • tandem mass spectrometry • grape • proteins • *Botrytis cinerea* • Western blot

### Introduction

Proteins are not the major components of wine, although they are essential compounds, contributing to its quality. Indeed, wine proteins are implied in the foaming properties of sparkling wines,<sup>1–3</sup> the stabilization of tartaric salt precipitation,<sup>4</sup> the interaction with wine volatile compounds,<sup>5</sup> the decrease in wine astringency and color stability,<sup>6</sup> and, unfortunately, in the formation of haze in white wines.<sup>7,8</sup>

The foam of sparkling wines is a very important criterion of quality, for consumers, as well as their effervescence.<sup>9,10</sup> The exact identity of wine proteins and those likely involved in foaming properties is still poorly elucidated. Previous studies have shown that wine proteins mainly originate from grape and yeast, as determined by one-dimensional electrophoresis coupled with immunodetection<sup>11</sup> or mass spectrometry.<sup>12</sup> Furthermore, a part of these wine proteins are considered as being glycosylated,<sup>13,14</sup> however, to date, no study has clearly identified the glycoproteins of wine, except for the grape invertase<sup>15</sup>

and for some yeast proteins, e.g., a mannoprotein.<sup>16</sup> In a recent work, Dambrouck et al.<sup>17</sup> established that a significant decrease in both the total protein and the grape invertase contents of Champagne base wines was correlated with the loss of wine foaming properties. Among grape proteins, another group is mostly important in terms of quality and quantity thanks to their proteolytic and low pH resistance, the pathogenesis-related (PR) proteins.<sup>7</sup> These proteins are constitutively present in grape berries from “véraison”, but their expression can be induced following biotic (pathogen) or abiotic (environment) stress. Waters et al.<sup>7,8</sup> showed that these proteins are in part responsible for the heat-induced haze in white wine and were identified as thaumatin-like proteins and chitinases. During alcoholic fermentation and aging on lees, heavily glycosylated proteins (i.e., mannoproteins) are released by yeast, such as *Saccharomyces cerevisiae*. These proteins were recently shown as having potential foam-active properties<sup>18</sup> and protective properties against protein haze,<sup>19</sup> as well.

Various factors are likely to modify the protein content of wine and, sometimes, its quality, as for example, the use of bentonite, that is, clays used for fining or riddling,<sup>17,20</sup> or the infection of grape berries by *Botrytis cinerea*. *B. cinerea* is indeed a widespread fungus, responsible for the gray mold disease, which degrades both must and wine proteins<sup>21,22</sup> by exhibiting proteolytic activity, as demonstrated by Marchal et al.<sup>23</sup> In addition to the moldy and earthy off flavors generally retrieved in botrytized wines,<sup>24</sup> this fungus could also induce important losses in the foaming properties of base wines<sup>22</sup> and Champagne wines.<sup>25</sup> Cilindre et al.<sup>22</sup> suggested that the decrease in

\* To whom correspondence should be addressed: Laboratoire d'Oenologie et Chimie Appliquée, Université de Reims, URVVC UPRES EA 2069, BP 1039, 51687 Reims Cedex 2, France. Telephone/Fax: +33-326-91-33-40. E-mail: clara.cilindre@univ-reims.fr.

<sup>†</sup> Laboratoire d'Oenologie et Chimie Appliquée, Université de Reims.

<sup>‡</sup> Laboratoire de Spectrométrie de Masse Bioorganique, Université de Strasbourg.

<sup>§</sup> Laboratoire Stress, Défenses et Reproduction des Plantes, Université de Reims.

<sup>||</sup> Present address: Departamento de Bioquímica, Biología Celular y Molecular de Plantas, Estación Experimental del Zaidín (CSIC), c/Professor Albareda 1, 18008 Granada, Spain.

base wine foam quality was probably linked to the degradation of some wine proteins, as shown by two-dimensional electrophoresis (2DE).

Few studies have attempted to precisely identify the wine proteins separated using one-dimensional electrophoresis and nano-liquid chromatography–tandem mass spectrometry (nano-LC–MS/MS)<sup>12</sup> or 2DE followed by N-terminal amino acid sequencing.<sup>26</sup> Other different mass spectrometry (MS) techniques were applied to analyze grape and wine proteins.<sup>27</sup> However, to date, 2DE coupled with nano-LC–MS/MS has never been employed to the identification of wine proteins.

To identify the wine proteins that are susceptible to be modified by an attack of *B. cinerea*, we have undertaken, for the first time, a precise protein analysis of two Champagne base wines, prepared with either healthy or botrytized grapes. A total of 37 spots of interest, separated by 2DE, were then identified using nano-LC–MS/MS and immunodetection analysis.

## Materials and Methods

**Wines.** Grape berries of the Chardonnay variety were hand-harvested in the same vineyard in the Champagne area (France), in September 2004. Wines were made from musts prepared from either healthy grapes or grapes naturally infected by *B. cinerea*. The rate of infection was expressed in the percent of infected grape berries versus the total number of grape berries on a bunch, and it was visually estimated at about 20%. Settled Chardonnay musts were racked and chaptalized with sucrose. The alcoholic fermentation was achieved by *S. cerevisiae* var. *bayanus* (active dry yeast, IOC 18-2007, Institut Oenologique de Champagne, Epernay, France) at 18 °C. Malolactic fermentation did not take place in these wines. Wines were centrifuged at 9500g for 10 min and filtered through a 0.45 µm membrane (Alltech) and stored at 4 °C.

**Protein Extraction.** Samples were concentrated 6 times and dialyzed 3 times against distilled water (0.2 L of concentrate plus 1.3 L of water). A hydrophilic polysulfone membrane (30 UFIB, Setric Génie Industriel, SGI, France) with a molecular-weight cutoff of 10 kDa was used. The crossflow filtration module is connected to the SGI Hi-Flow system (pumping system plus glass tank). Ultrafiltrations were carried out at 12 °C. Dialyzed retentates were aliquoted (15 mL) and stored at –20 °C before being used for alcoholic precipitation. Each sample aliquot (15 mL) was diluted with eight volumes of an ice-cold ethanol solution containing 15% (w/v) trichloroacetic acid (TCA). Proteins were then allowed to precipitate at 4 °C for 1 h. After centrifugation at 9500g for 8 min at 4 °C, pellets were washed in 60 mL of ice-cold ethanol as described above. Samples were washed again in 30 mL of ice-cold ethanol, and supernatants were removed. Resulting pellets were air-dried and then resolubilized in 1 mL of a sample buffer consisting of 7 M urea, 2 M thiourea, 4% (w/v) 3-[(3-cholamidopropyl)-dimethylammonio]-1-propanesulfonate (CHAPS), 0.5% (v/v) immobilized pH gradient (IPG) buffer 3-10, 60 mM 1,4-dithiothreitol (DTT), and traces of bromophenol blue.

**Two-Dimensional Electrophoresis (2DE).** For preparative 2DE analysis, precast dry polyacrylamide 18 cm length gels [ReadyStrip IPG (pH 3–6), Bio-Rad, Hercules, CA] were rehydrated with a mixture containing wine proteins diluted in sample buffer to a final volume of 350 µL. Isoelectric focusing (IEF) was conducted at 20 °C in an IPGphor unit (Amersham Pharmacia, Sweden) as follows: a linear increase from 50 to 8000 V to give a total of 70 000 V h. Focused proteins were reduced and subsequently alkylated according to Görg et al.<sup>28</sup>

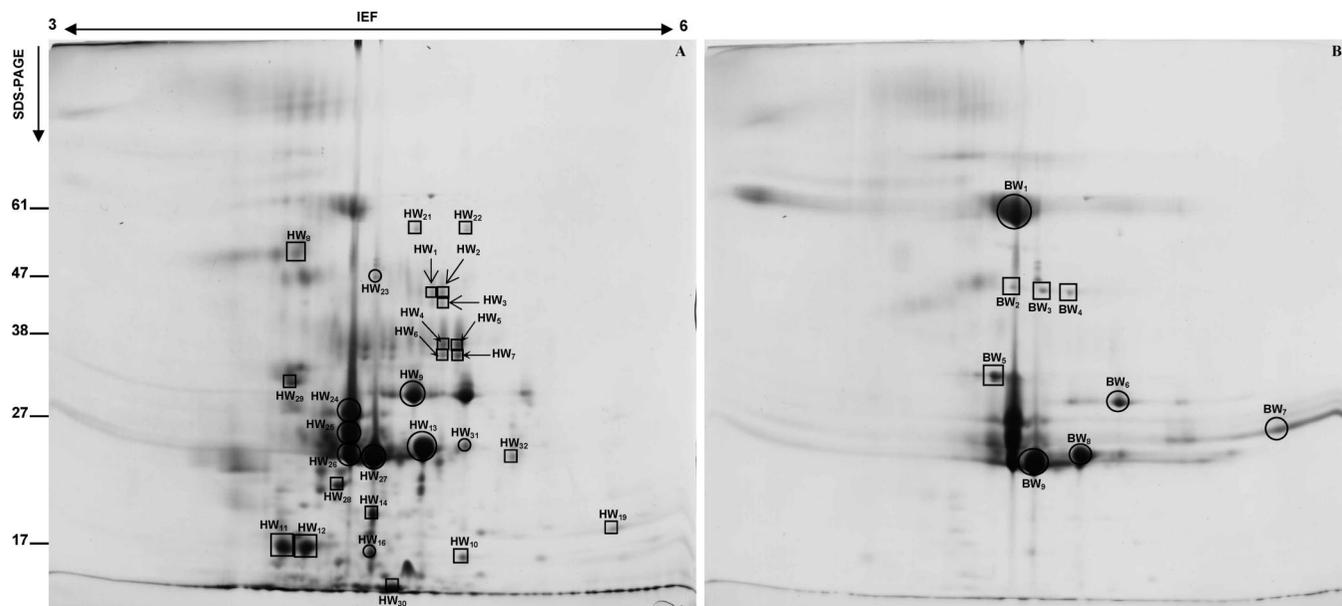
Gels were then placed on the top of vertical slabs of polyacrylamide (12% T and 2.6% C). The stacking gel was replaced by a layer of 1% (w/v) low melting point agarose, 0.15 M Bis-Tris/0.1 M HCl, and 0.2% (w/v) sodium dodecyl sulfate (SDS). Electrophoretic migration along the second dimension was performed using a Protean II xi Cell (Bio-Rad) at 10 °C under a constant current of 20 mA/gel for 1 h, followed by 40 mA/gel for 3 h. After completion of SDS–polyacrylamide gel electrophoresis (PAGE), gels were either stained with colloidal Coomassie Brilliant Blue according to Anderson et al.<sup>29</sup> or electrophoretically transferred to nitrocellulose membranes in a Hoefer Transphor unit (Amersham Pharmacia Biotech, Sweden) at 80 V for 1 h. Blotted membranes were then incubated with different polyclonal antibodies following the protocol described below.

**Polyclonal Antibodies.** Polyclonal antibodies raised against thaumatin-like (TL) protein and chitinase (CHV5), two pathogenesis-related proteins, purified from Pinot Noir grape berries, were kindly donated by Pr. F. Bailleul (University of Reims, France). Polyclonal antibodies against the grape berry invertase were provided by our laboratory and produced as described previously.<sup>17</sup>

**Western Blot and Immunodetection.** After transfer, membranes were blocked with 2% (w/v) skim milk powder (Merck, Germany) in Tris buffer saline (TBS), washed 3 times in TBS, and incubated for 3 h with the appropriate primary antibody diluted in TBS and 1% (w/v) skim milk powder. Membranes were then successively washed in TBS, TBS plus 0.05% (v/v) Tween 20, and TBS, before incubation with a goat anti-rabbit IgG peroxidase-conjugated antibody (Sigma-Aldrich, St. Louis, MO), and diluted  $1/_{1500}$  in TBS plus 1% (w/v) skim milk powder at 12 °C overnight. After further rinse in phosphate buffer saline (PBS), membranes were incubated with equal parts of luminol enhancer and peroxide solutions (SuperSignal West Pico Chemiluminescent Substrate, Pierce, Rockford, IL) for 8 min in the dark. Membranes were then placed in a plastic sheet protector and exposed to an autoradiographic film (Biomax ML, Kodak), which was fixed and developed using GBX solutions (Kodak, France).

**Image Analysis.** Digitized images at 63.5 µm resolution were obtained using the GS-800 scanner and Quantity One 4.6.2 software (Bio-Rad, Hercules, CA). Computerized 2D gel analysis, including spot detection and quantification, was performed using the PDQuest Basic 8.0.1 software (Bio-Rad, Hercules, CA). Reproducibility of 2D gels was assayed by running all samples at least 3 times. The relative molecular mass was calibrated with internal markers (2D Standards, Bio-Rad, Hercules, CA) after co-electrophoresis. Selected protein features were modeled as Gaussians, and the relative optical densities (OD), i.e., the feature OD divided by the total OD over the whole image, were computed. Means ± standard deviation (SD) ( $n = 3$ ) were finally calculated for each sample. The ratio of the expression level for each protein in the botrytized wine over the healthy one was also estimated. Values higher than 1.5 or lower than 0.5 were considered as significant.

**In-Gel Digestion and Nano-LC–MS/MS Analysis.** Gels bands of interest were cut out manually. Then, in-gel digestion was performed with an automated protein digestion system, MassPREP station (Waters, Milford, MA). The gel plugs were washed twice with 50 µL of 25 mM  $\text{NH}_4\text{HCO}_3$  and 50 µL of acetonitrile (ACN). The cysteine residues were reduced by 50 µL of 10 mM DTT at 57 °C and alkylated by 50 µL of 55 mM iodoacetamide at room temperature. After dehydration of the



**Figure 1.** Two-dimensional PAGE of proteins from a healthy (A) and botrytized (B) Champagne base wine stained with colloidal Coomassie Brilliant Blue. Healthy wine and botrytized wine proteins were separated by IEF in 18 cm long pH 3–6 IPG strips, followed by SDS–PAGE in vertical 12% gels. Numbered spots were excised, digested by trypsin, and analyzed by nano-LC–MS/MS. Spots common to both wines are circled, and those only present in the healthy or botrytized wine are squared. Molecular masses are displayed on the left of the gels.

gel bands with ACN, proteins were digested in gel with 25  $\mu\text{L}$  of 12.5 ng/ $\mu\text{L}$  modified porcine trypsin (Promega, Madison, WI) in 25 mM  $\text{NH}_4\text{HCO}_3$ , at room temperature overnight. The generated peptides were extracted with 60% ACN in 5% formic acid and analyzed by nano-LC–MS/MS.

Nano-LC–MS/MS analysis of the tryptic peptides was performed using an Agilent 1100 series HPLC–Chip/MS system (Agilent Technologies, Palo Alto, CA) coupled to an HCT Ultra ion trap (Bruker Daltonics). The voltage applied to the capillary cap was optimized to  $-1850$  V. For tandem MS experiments, the system was operated with automatic switching between MS and MS/MS modes. The three most abundant peptides, preferring doubly charged ions, were selected on each MS spectrum for further isolation and fragmentation. The MS/MS scanning was performed in the ultrascan resolution mode at a scan rate of  $m/z$  26,000 per second. A total of 6 scans were averaged to obtain a MS/MS spectrum. The complete system was fully controlled by ChemStation (Agilent Technologies) and Esquire-Control (Bruker Daltonics) software.

Mass data collected during nano-LC–MS/MS analysis were processed, converted into .mgf files, and interpreted using a local Mascot (Matrix Science, London, U.K.) server running on a 3 GHz Pentium IV processor. Searches were performed against the National Center for Biotechnology Information nonredundant (NCBI nr) database without any taxonomic, molecular weight, or isoelectric point restrictions. Searches were performed with a tolerance on mass measurements of 0.25 Da in both MS and MS/MS modes. One missed cleavage per peptide was allowed, and some variable modifications were taken into account, such as carbamidomethylation for cysteine and oxidation for methionine. Identifications were validated according to the established guidelines for proteomic data publication.<sup>30</sup> A complete list of the peptides identified for each protein is given in Tables S1 and S2 in the Supporting Information.

## Results and Discussion

**Influence of *B. cinerea* Infection on the 2D Pattern of the Chardonnay Base Wine.** An acidic pH range (pH 3–6) was chosen for the first dimension according to the results obtained by Cilindre et al.,<sup>22</sup> who have observed that almost all wine proteins were located between pH 3 and 5. Proteins from healthy and botrytized wines were thus analyzed with 2DE and revealed by colloidal Coomassie Brilliant Blue G-250, a sensitive and MS-compatible stain (parts A and B of Figure 1). Few spots were detected in the botrytized wine (Figure 1B) as compared to the healthy wine (Figure 1A), respectively, an average of 85 and 25 spots for the healthy and botrytized wines. A quantitative and comparative analysis of these spots showed that (i) almost 75% of the healthy wine spots disappeared in the botrytized wine, (ii) nine spots from the botrytized wine were absent in the healthy wine, and thus, (iii) 20 spots were present in both wines. More than half of these common spots were more or less intense in the botrytized wine, showing that a proteolysis<sup>23</sup> or a regulation of their expression has probably occurred following grape infection. Different authors reported the proteolytic activity exhibited by *B. cinerea* upon infection. Indeed, this fungus secretes various aspartic proteases, active in acid media,<sup>23,31,32</sup> such as grape and wine, that could thus degrade some proteins in mature grape berries. Our results clearly indicate that *B. cinerea* infection has a significant impact on base wine proteins, and raise a question: what kind of proteins are either vulnerable or, in contrast, resistant to *B. cinerea* or differentially expressed upon fungal attack?

Recently, Cilindre et al.<sup>22</sup> attempted to identify the origin of these differentially and nondifferentially expressed spots with the use of antibodies raised against either grape juice proteins or *B. cinerea*-secreted proteins; however, the exact identity of these wine proteins was not elucidated.

A total of 28 spots from the healthy wine and 9 spots from the botrytized wine were thus chosen for nano-LC–MS/MS

analysis, according to their presence or absence in the botrytized wine or to their change in abundance. Spots are numbered in parts A and B of Figure 1, and these are in accordance with those listed in Table 1.

**Identification of Proteins by Nano-LC–MS/MS.** The protein spots of interest (differentially and nondifferentially expressed) were excised from 2D gels, hydrolyzed with trypsin, and submitted to nano-LC–MS/MS analysis. A total of 37 spots from both base wines were identified using the NCBI nr database, and these are listed in Tables 1 and 2. Most of the identified wine proteins are from grape origin and can be classified in two main groups: (i) proteins involved in sugar metabolism (i.e., vacuolar invertase) and (ii) defense-related proteins (e.g., chitinase, TL proteins, and osmotin-like proteins). At the time of writing, only 1068 sequences from the genus *Vitis* were referenced in the NCBI nonredundant database. The other wine proteins are from yeast origin (*S. cerevisiae*), and fungal origin (*Botryotinia fuckeliana*, teleomorph of *B. cinerea*).

For several spots (e.g., spots HW<sub>13</sub>, HW<sub>23</sub>, and BW<sub>2</sub>), multiple proteins were identified using the database-searching program Mascot; these results might be due to “spot overlapping”. Indeed, this phenomenon occurs when two or more proteins have a similar electrophoretic behavior, as already reported by Westbrook et al.<sup>33</sup>

To validate the results of MS, three polyclonal antibodies raised against the major wine proteins identified [i.e., grape invertase, chitinase (CHV5), and TL] were used for a Western blot analysis (parts A–D of Figure 2).

**Proteins Associated with Sugar Metabolism: The Grape Vacuolar Invertase.** Numerous spots were identified as the vacuolar invertase 1, GIN1 (*Vitis vinifera*), which corresponds to the preferential expressed form encoded by GIN1 cDNA in berries.<sup>34</sup> Plant invertases, which are responsible for the conversion of sucrose into glucose and fructose, are involved in plant development and growth and also in biotic (pathogen) and abiotic (environment) interactions.<sup>35</sup>

Using nano-LC–MS/MS analysis, GIN1 was mainly retrieved in the healthy wine (16 spots) (Table 1) and only three spots among the botrytized-wine-identified proteins (Table 2). There was a reduction in the number of invertase-identified spots in the botrytized wine, as confirmed by 2D blots probed with anti-grape invertase antibodies (parts A and B of Figure 2). Indeed, 10 invertase-identified spots were absent in the botrytized wine.

GIN1 contains 642 amino acids and has a theoretical  $M_r$  of 75.1 kDa and  $pI$  4.6. This is far from the  $M_r$  estimated at 61 kDa, by Davies and Robinson,<sup>34</sup> for the predicted mature of GIN1. Furthermore, this protein has 12 potential sites of N-glycosylation, and the possible presence of glycans could explain the low coverage percentages obtained for these spots identified as GIN1. In a previous work, it was demonstrated that predeglycosylation and multiple proteasic digestions of a N-glycoprotein increase the amino acid sequence coverage.<sup>36</sup>

Thus, the heterogeneity of N-glycosylation may account for the presence of these numerous invertase-identified spots. Furthermore, few spots had the same matched peptides but with different  $M_r$  and  $pI$ , that is, HW<sub>2</sub>, HW<sub>5</sub>, and HW<sub>6</sub>. These spots could represent different glycoforms of the same invertase fragment.

Another reason for the presence of these multiple spots could be that the major spot of grape invertase (BW<sub>1</sub>) has undergone proteolytic degradation during the alcoholic fermentation, as suggested by Okuda et al.,<sup>26</sup> who also found multiple forms of invertase in a Chardonnay wine. These fragments (i.e.,

HW<sub>1</sub>–HW<sub>8</sub>) would be then more exposed to an attack by fungal proteases,<sup>23</sup> thus explaining their absence in the botrytized wine.

On the other hand, some spots are more resistant to the attack of *B. cinerea*, for example, HW<sub>24</sub>, HW<sub>25</sub>, HW<sub>26</sub>, and BW<sub>9</sub>. One spot, BW<sub>1</sub>, was even more intense (+2.0 fold) in the botrytized wine. Its expression was probably induced after *B. cinerea* infection, and this might be related with an increase in the invertase activity as previously reported in other plants.<sup>37</sup>

**Proteins Involved in Plant Defenses.** PR proteins are classified, to date, in 17 families and are present in many plant species.<sup>38</sup> In the grape berry, they accumulate during ripening and can be induced by biotic and/or abiotic stress. Some PR proteins [e.g., thaumatin-like (TL), class IV endochitinase, and lipid-transfer protein] are considered as allergens and have been found both in grape and wine.<sup>39</sup> On the other hand, the relevance of these proteins, particularly TL and chitinase, in haze has already been demonstrated.<sup>7,8</sup> However, most of these proteins have antifungal activities, and their elimination, by a repression of their expression or by genetically modified vines, would likely have deleterious effects on crop quality.

PR proteins identified in both healthy and botrytized wines belong to the following families: PR-2 ( $\beta$ -1,3-glucanase), PR-3 (class IV chitinase and endochitinase), PR-4 (chitinase), and PR-5 (TL and osmotin-like). Most of the PR proteins are encoded by multigene families, and this may account for the numerous spots identified as PR proteins in both wines. Post-translational modifications (PTMs), such as glycosylation or phosphorylation, may be consistent with the presence of multiple isoforms of PR proteins too.

**PR-5 Family: Thaumatin-like Proteins (TL) and Osmotin-like Proteins (OL).** A total of 14 spots from the healthy wine (including two common spots in the botrytized wine) were identified as VVTL1 (gil2213852) or putative TL protein (gil7406714 and gil7406716) (Table 1); both proteins belong to the TL family. Among these spots, 11 spots were probed with specific antibodies raised against a TL protein extracted from Pinot Noir grape berries (parts C and D of Figure 2) and sharing a N-terminal homology with VVTL1.<sup>40</sup> Four spots were absent in the botrytized wine, as for example, HW<sub>11</sub>. This spot has an experimental  $M_r$  around 18 kDa, which is lower than the theoretical  $M_r$  of VVTL1 (24 kDa). Identified tryptic peptides revealed that spot HW<sub>11</sub> is derived from the C-terminal part of VVTL1, thus corresponding to a proteolytic fragment or an isoform of VVTL1. In the healthy and botrytized wine, three (HW<sub>12</sub>, HW<sub>13</sub>, and HW<sub>31</sub>) and two spots (BW<sub>8</sub> and BW<sub>9</sub>) were respectively identified by MS as putative TL protein (gil7406716). This TL protein (gil7406716) and VVTL1 do not have a potential N-glycosylation site, contrary to other TL proteins (gil33329390 and gil7406714), which have two potential sites of N-glycosylation.

Proteins induced by osmotic stress are called osmotin-like proteins. They belong to the thaumatin-like family, with which they have a strong homology of sequence. Osmotin-like proteins (gil1839046) were identified in both wines: nine spots in the healthy wine and six spots in the botrytized wine. Among these spots, two (HW<sub>14</sub> and HW<sub>30</sub>) were only present in the healthy wine and four (BW<sub>2</sub>, BW<sub>3</sub>, BW<sub>4</sub>, and BW<sub>5</sub>) were only present in the botrytized wine.

For most of them, peptide sequences retrieved for each spot identified as osmotin were similar to two TL proteins: putative thaumatin-like protein (gil7406174) and thaumatin-like protein (gil33329390); therefore, precise identification between these

**Table 1.** Proteins from the Healthy Wine (HW) Identified by Nano-LC–MS/MS after IEF/SDS–PAGE

spot name	protein name (organism)	accession number <sup>a</sup>	number of peptides matched <sup>b</sup>	sequence coverage (%) <sup>c</sup>	trend in the tryptized wine
HW <sub>1</sub>	vacuolar invertase 1, GIN1 ( <i>Vitis vinifera</i> )	gil1839578	5	10	absent
HW <sub>2</sub>	vacuolar invertase 1, GIN1 ( <i>Vitis vinifera</i> )	gil1839578	4	9	absent
HW <sub>3</sub>	vacuolar invertase 1, GIN1 ( <i>Vitis vinifera</i> )	gil1839578	2	4	absent
HW <sub>4</sub>	VVTL1 ( <i>Vitis vinifera</i> )	gil2213852	4	26	absent
	vacuolar invertase 1, GIN1 ( <i>Vitis vinifera</i> )	gil1839578	4	9	
HW <sub>5</sub>	vacuolar invertase 1, GIN1 ( <i>Vitis vinifera</i> )	gil1839578	4	9	absent
HW <sub>6</sub>	vacuolar invertase 1, GIN1 ( <i>Vitis vinifera</i> )	gil1839578	4	9	absent
	class IV endochitinase ( <i>Vitis vinifera</i> )	gil2306813	2	15	
	class IV chitinase ( <i>Vitis vinifera</i> )	gil33329392	2	15	
HW <sub>7</sub>	vacuolar invertase 1, GIN1 ( <i>Vitis vinifera</i> )	gil1839578	3	6	absent
HW <sub>8</sub>	vacuolar invertase 1, GIN1 ( <i>Vitis vinifera</i> )	gil1839578	3	7	absent
HW <sub>9</sub>	class IV endochitinase ( <i>Vitis vinifera</i> )	gil2306813	2	16	decrease
	class IV chitinase ( <i>Vitis vinifera</i> )	gil33329392	2	16	
HW <sub>10</sub>	class IV chitinase ( <i>Vitis vinifera</i> )	gil33329392	3	15	absent
HW <sub>11</sub>	VVTL1 ( <i>Vitis vinifera</i> )	gil2213852	8	36	absent
HW <sub>12</sub>	putative thaumatin-like protein ( <i>Vitis vinifera</i> )	gil7406716	5	33	absent
	VVTL1 ( <i>Vitis vinifera</i> )	gil2213852	5	33	
HW <sub>13</sub>	putative thaumatin-like protein ( <i>Vitis vinifera</i> )	gil7406716	9	42	decrease
	osmotin-like protein ( <i>Vitis vinifera</i> )	gil1839046	2	13	
	putative thaumatin-like protein ( <i>Vitis vinifera</i> )	gil7406714	2	13	
	thaumatin-like protein ( <i>Vitis vinifera</i> )	gil33329390	2	13	
	vacuolar invertase 1, GIN1 ( <i>Vitis vinifera</i> )	gil1839578	2	6	
	$\beta$ -1,3-glucanase ( <i>Vitis vinifera</i> )	gil1549400	1	10	
HW <sub>14</sub>	VVTL1 ( <i>Vitis vinifera</i> )	gil2213852	9	50	absent
	vacuolar invertase 1, GIN1 ( <i>Vitis vinifera</i> )	gil1839578	3	4	
	osmotin-like protein ( <i>Vitis vinifera</i> )	gil1839046	2	13	
	putative thaumatin-like protein ( <i>Vitis vinifera</i> )	gil7406714	2	13	
	thaumatin-like protein ( <i>Vitis vinifera</i> )	gil33329390	2	13	
HW <sub>16</sub>	VVTL1 ( <i>Vitis vinifera</i> )	gil2213852	7	36	decrease
	class IV endochitinase ( <i>Vitis vinifera</i> )	gil2306811	3	11	
	class IV endochitinase ( <i>Vitis vinifera</i> )	gil2306813	3	11	
	class IV chitinase ( <i>Vitis vinifera</i> )	gil33329392	3	11	
	vacuolar invertase 1, GIN1 ( <i>Vitis vinifera</i> )	gil1839578	6	11	
	osmotin-like protein ( <i>Vitis vinifera</i> )	gil1839046	3	17	
	putative thaumatin-like protein ( <i>Vitis vinifera</i> )	gil7406714	3	17	
	thaumatin-like protein ( <i>Vitis vinifera</i> )	gil33329390	3	17	
HW <sub>19</sub>	class IV chitinase ( <i>Vitis vinifera</i> )	gil33329392	3	21	absent
HW <sub>21</sub>	class IV chitinase ( <i>Vitis vinifera</i> )	gil33329392	4	24	absent
HW <sub>22</sub>	class IV chitinase ( <i>Vitis vinifera</i> )	gil33329392	5	32	absent
HW <sub>23</sub>	vacuolar invertase 1, GIN1 ( <i>Vitis vinifera</i> )	gil1839578	5	8	increase
	osmotin-like protein ( <i>Vitis vinifera</i> )	gil1839046	4	22	
	putative thaumatin-like protein ( <i>Vitis vinifera</i> )	gil7406714	4	22	
	thaumatin-like protein ( <i>Vitis vinifera</i> )	gil33329390	4	22	
	VVTL1 ( <i>Vitis vinifera</i> )	gil2213852	2	11	
	YJU1 ( <i>Saccharomyces cerevisiae</i> )	gil4814	1	6	
	class IV endochitinase ( <i>Vitis vinifera</i> )	gil2306811	2	23	
	class IV endochitinase ( <i>Vitis vinifera</i> )	gil2306813	2	23	
	class IV chitinase ( <i>Vitis vinifera</i> )	gil33329392	2	23	
HW <sub>24</sub>	osmotin-like protein ( <i>Vitis vinifera</i> )	gil1839046	6	32	no change
	putative thaumatin-like protein ( <i>Vitis vinifera</i> )	gil7406714	6	38	
HW <sub>25</sub>	osmotin-like protein ( <i>Vitis vinifera</i> )	gil1839046	6	32	no change
	putative thaumatin-like protein ( <i>Vitis vinifera</i> )	gil7406714	6	38	
	vacuolar invertase 1, GIN1 ( <i>Vitis vinifera</i> )	gil1839578	4	6	
HW <sub>26</sub>	vacuolar invertase 1, GIN1 ( <i>Vitis vinifera</i> )	gil1839578	7	17	no change
	putative thaumatin-like protein ( <i>Vitis vinifera</i> )	gil7406714	6	41	
	osmotin-like protein ( <i>Vitis vinifera</i> )	gil1839046	6	32	
	VVTL1 ( <i>Vitis vinifera</i> )	gil2213852	5	33	
HW <sub>27</sub>	putative thaumatin-like protein ( <i>Vitis vinifera</i> )	gil7406714	6	30	no change
	vacuolar invertase 1, GIN1 ( <i>Vitis vinifera</i> )	gil1839578	5	10	
	VVTL1 ( <i>Vitis vinifera</i> )	gil2213852	4	22	
HW <sub>28</sub>	putative thaumatin-like protein ( <i>Vitis vinifera</i> )	gil7406714	5	37	absent
HW <sub>29</sub>	vacuolar invertase 1, GIN1 ( <i>Vitis vinifera</i> )	gil1839578	9	19	absent
HW <sub>30</sub>	PR-4 type protein ( <i>Vitis vinifera</i> )	gil3511147	5	56	absent
	osmotin-like protein ( <i>Vitis vinifera</i> )	gil1839046	3	17	
	putative thaumatin-like protein ( <i>Vitis vinifera</i> )	gil7406714	3	17	

**Table 1.** Continue.

spot name	protein name (organism)	accession number <sup>a</sup>	number of peptides matched <sup>b</sup>	sequence coverage (%) <sup>c</sup>	trend in the botrytized wine
HW <sub>31</sub>	thaumatin-like protein ( <i>Vitis vinifera</i> )	gil33329390	3	17	increase
	VVTL1 ( <i>Vitis vinifera</i> )	gil2213852	13	59	
	putative thaumatin-like protein ( <i>Vitis vinifera</i> )	gil7406716	12	59	
	osmotin-like protein ( <i>Vitis vinifera</i> )	gil1839046	3	15	
	putative thaumatin-like protein ( <i>Vitis vinifera</i> )	gil7406714	3	15	
HW <sub>32</sub>	thaumatin-like protein ( <i>Vitis vinifera</i> )	gil33329390	3	15	absent
	class IV chitinase ( <i>Vitis vinifera</i> )	gil33329392	6	32	

<sup>a</sup> Protein identification number provided by the NCBI nr database. <sup>b</sup> Number of peptides identified by the MASCOT search engine, which matched with the identified protein. <sup>c</sup> Percentage ratio of all amino acids from valid peptides matched to the total number of amino acids in the reference proteins.

**Table 2.** Proteins from the Botrytized Wine (BW) Identified by Nano-LC–MS/MS after IEF/SDS–PAGE

spot name	protein name (organism)	accession number <sup>a</sup>	number of peptides matched <sup>b</sup>	sequence coverage (%) <sup>c</sup>	trend in the healthy wine
BW <sub>1</sub>	vacuolar invertase 1, GIN1 ( <i>Vitis vinifera</i> )	gil1839578	11	14	decrease
	osmotin-like protein ( <i>Vitis vinifera</i> )	gil1839046	3	17	
	putative thaumatin-like protein ( <i>Vitis vinifera</i> )	gil7406714	3	17	
	thaumatin-like protein ( <i>Vitis vinifera</i> )	gil33329390	3	17	
BW <sub>2</sub>	vacuolar invertase 1, GIN1 ( <i>Vitis vinifera</i> )	gil1839578	5	7	absent
	osmotin-like protein ( <i>Vitis vinifera</i> )	gil1839046	4	22	
	putative thaumatin-like protein ( <i>Vitis vinifera</i> )	gil7406714	4	22	
	thaumatin-like protein ( <i>Vitis vinifera</i> )	gil33329390	4	22	
	YJU1 ( <i>Saccharomyces cerevisiae</i> )	gil4814	3	18	
BW <sub>3</sub>	endopolygalacturonase 4 ( <i>Botryotinia fuckeliana</i> )	gil3282222	2	8	absent
	osmotin-like protein ( <i>Vitis vinifera</i> )	gil1839046	4	22	
	putative thaumatin-like protein ( <i>Vitis vinifera</i> )	gil7406714	4	22	
	thaumatin-like protein ( <i>Vitis vinifera</i> )	gil33329390	4	22	
BW <sub>4</sub>	pectin methyl esterase ( <i>Botryotinia fuckeliana</i> )	gil12964194	4	13	absent
	pectin methyl esterase ( <i>Botryotinia fuckeliana</i> )	gil12964194	6	20	
	osmotin-like protein ( <i>Vitis vinifera</i> )	gil1839046	3	17	
BW <sub>5</sub>	putative thaumatin-like protein ( <i>Vitis vinifera</i> )	gil7406714	3	17	absent
	thaumatin-like protein ( <i>Vitis vinifera</i> )	gil33329390	3	17	
	osmotin-like protein ( <i>Vitis vinifera</i> )	gil1839046	2	13	
BW <sub>6</sub>	putative thaumatin-like protein ( <i>Vitis vinifera</i> )	gil7406714	2	13	absent
	thaumatin-like protein ( <i>Vitis vinifera</i> )	gil33329390	2	13	
BW <sub>7</sub>	class IV chitinase ( <i>Vitis vinifera</i> )	gil33329392	5	31	increase
BW <sub>8</sub>	class IV chitinase ( <i>Vitis vinifera</i> )	gil33329392	5	25	decrease
BW <sub>9</sub>	putative thaumatin-like protein ( <i>Vitis vinifera</i> )	gil7406716	10	48	increase
BW <sub>9</sub>	putative thaumatin-like protein ( <i>Vitis vinifera</i> )	gil7406716	5	21	no change
	osmotin-like protein ( <i>Vitis vinifera</i> )	gil1839046	3	17	
	putative thaumatin-like protein ( <i>Vitis vinifera</i> )	gil7406714	3	17	
	thaumatin-like protein ( <i>Vitis vinifera</i> )	gil33329390	3	17	
	vacuolar invertase 1, GIN1 ( <i>Vitis vinifera</i> )	gil1839578	3	6	

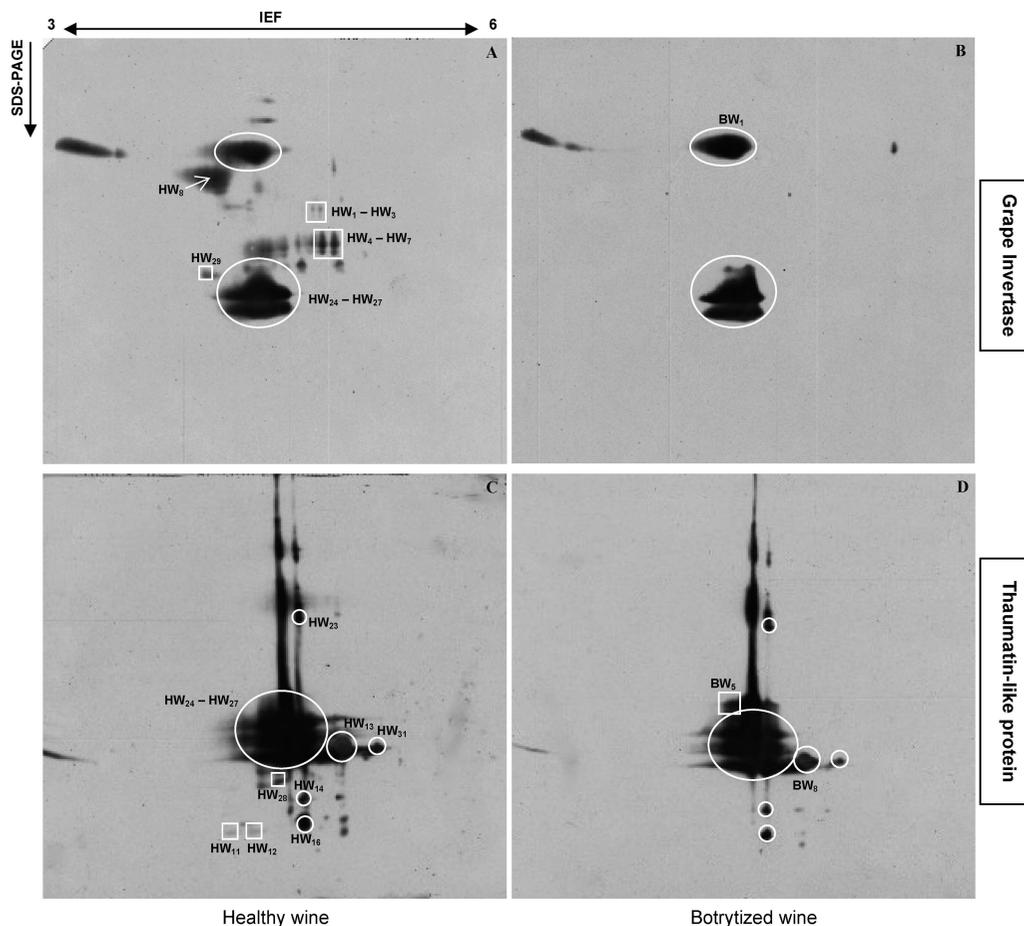
<sup>a</sup> Protein identification number provided by the NCBI nr database. <sup>b</sup> Number of peptides identified by the MASCOT search engine, which matched with the identified protein. <sup>c</sup> Percentage ratio of all amino acids from valid peptides matched to the total number of amino acids in the reference proteins.

three possibilities was more difficult. Nevertheless, three spots present in both wines (HW<sub>24</sub>, HW<sub>25</sub>, and HW<sub>26</sub>) were also identified as the osmotin-like protein (gil1839046) and the putative thaumatin-like protein (gil7406714), without any similar peptides.

Manteau<sup>40</sup> reported that N-terminal sequencing of the 24 kDa protein (extracted by SDS–PAGE) used for TL antibodies production, provided the following sequence: ATFNIQNHX-SYTVWAAAVPGGGM. Using BLAST search, they concluded that this protein was a thaumatin-like protein similar to VVTL1.<sup>41</sup> Because databases are unceasingly updated, we remade a BLAST analysis with this sequence and found a high homology with an osmotin-like protein (gil1839046) and a thaumatin-like protein (gil33329390); VVTL1 had a lower score. Therefore, differentiation between osmotin and thaumatin-like protein remained difficult, even with the use of MS and Western blot

analysis. Proteins from the PR-5 family were present in both wines, with a slight reduction of the number of spots, together with the appearance of one spot (BW<sub>5</sub>) in the botrytized wine. BW<sub>2</sub>, BW<sub>3</sub>, and BW<sub>4</sub> were not recognized by TL antibodies, such that they were not considered as TL or OL proteins, despite the MS results. Numerous spots were recognized by TL antibodies and were mainly common to both wines, with an exception for some spots with molecular masses below 23 kDa.

**PR-2 Family.**  $\beta$ -1,3-Glucanases are pathogen-induced enzymes that inhibit fungal growth by hydrolyzing  $\beta$ -1,3-glucans in fungal cell walls. Curiously, this type of grape protein was only identified in one spot (HW<sub>13</sub>) among the 28 spots analyzed in the healthy wine. Furthermore, this protein was not identified among the spots analyzed in the botrytized wine, even in BW<sub>8</sub> that was common to HW<sub>13</sub> but with a significant decrease in intensity. Different reasons could explain the absence of



**Figure 2.** Two-dimensional immunoblots of healthy and botrytized wine proteins probed with either grape invertase (A, healthy wine; B, botrytized wine) or thaumatin-like protein (C, healthy wine; D, botrytized wine) polyclonal antibodies. Spots common to both wines are circled, and those only present in the healthy or botrytized wine are squared.

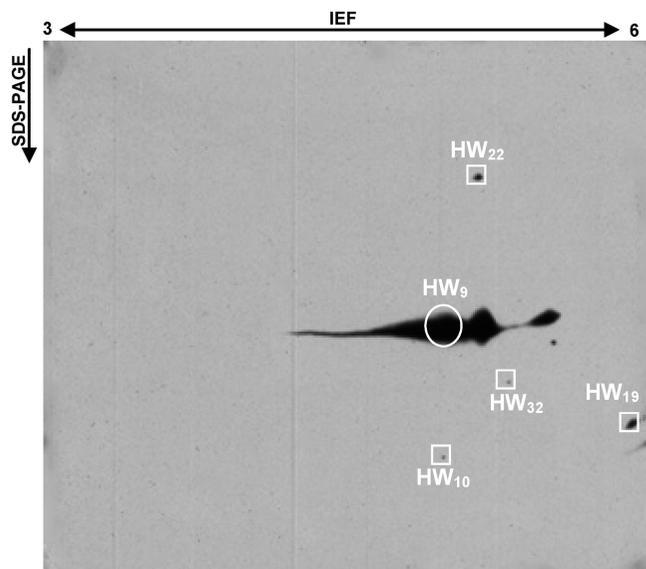
$\beta$ -1,3-glucanase in the botrytized wine, such as its degradation by fungal proteases<sup>23</sup> or an inhibition of its expression upon fungal infection. Another simple explanation is that this protein might be present among the nonidentified spots of the botrytized wine. Its presence in the healthy wine suggests that this PR protein is constitutively present in healthy grape berries. Indeed, the healthy wine had a nil gluconic acid content (healthy wine, 0.0 g/L; botrytized wine, 0.10 g/L), a compound known as a marker for *B. cinerea* infection. This ensures that what we called the “healthy wine” was truly prepared with healthy grape berries. On the contrary, Kwon<sup>12</sup> also identified a basic extracellular  $\beta$ -1,3-glucanase precursor (*Vitis vinifera*, gil451201) in a Chardonnay wine, but this wine was likely made with botrytized grapes because one fungal protein (i.e., laccase 2, *Botryotinia fuckeliana*) was also identified. Very few studies have shown the presence of glucanase (mRNA or protein) in healthy grapes.<sup>42,43</sup> It may be possible that other biotic or abiotic factors induced  $\beta$ -1,3-glucanases.<sup>44</sup>

**PR-3 and PR-4 Families: Chitinases and Endochitinases.** Chitinases and endochitinases are members of four PR families (PR3, PR4, PR8, and PR11), which are gathered in seven classes.<sup>38</sup> These proteins have chitinolytic activities and can also exhibit antifungal activities in *in vitro* assays, such as the inhibition of *B. cinerea* growth.<sup>45</sup>

Less spots were identified as PR3 and PR4 proteins, in both wines, than PR5 proteins. In the healthy wine, nine spots were identified as PR3 protein and one spot was identified as PR4

protein. Among the PR3-identified spots, six were absent in the botrytized wine (HW<sub>6</sub>, HW<sub>10</sub>, HW<sub>19</sub>, HW<sub>21</sub>, HW<sub>22</sub>, and HW<sub>32</sub>). The PR4-identified spot (HW<sub>30</sub>) was only present in the healthy wine. Furthermore, two spots identified as class IV chitinase (gil33329392) (BW<sub>6</sub> and BW<sub>7</sub>) were present in both wines but with different relative intensities in the healthy wine: a lower intensity for BW<sub>6</sub> and a higher intensity for BW<sub>7</sub>. For some spots, peptides matched were similar to multiple PR3 proteins (HW<sub>6</sub>, HW<sub>9</sub>, HW<sub>16</sub>, and HW<sub>23</sub>) as observed for some TL proteins.

Derckel et al.<sup>46</sup> revealed the presence of at least five chitinase induced isoforms in grapevine leaves and showed that these proteins are constitutively present in other tissues. In 1998, these authors<sup>45</sup> purified a chitinase isoform (CHV5) from mature grape berries that exhibited strong homology with the class IV chitinase isoform purified by Waters et al.<sup>7</sup> in a white wine and cloned by Robinson et al.<sup>47</sup> Polyclonal antibodies raised against CHV5 were used for Western blot analysis, but few proteins were recognized in the healthy wine (Figure 3) and none were recognized in the botrytized wine. Nevertheless, with Western blot analysis results (Figure 3), five proteins were confirmed as PR3 proteins in the healthy wine (HW<sub>9</sub>, HW<sub>10</sub>, HW<sub>19</sub>, HW<sub>22</sub>, and HW<sub>32</sub>). Almost all PR3 proteins identified in the healthy wine were absent or partially degraded (HW<sub>9</sub>, HW<sub>16</sub>, and HW<sub>23</sub>) in the botrytized wine. These results suggest that PR3 and PR4 proteins are less resistant to an attack by *B. cinerea*, as was proposed by Girbau et al.<sup>48</sup> These authors found



**Figure 3.** Immunoblot of healthy wine proteins probed with anti-chitinase (CHV 5) polyclonal antibodies. Spots common to both wines are circled, and those only present in the healthy wine are squared.

that the level of chitinases was significantly reduced in juices from grapes infected with *B. cinerea*; they suggested that proteolytic degradation of these proteins was the most consistent explanation.

It is well-known however that numerous PR proteins accumulate in ripening grape berries, especially chitinases and  $\beta$ -1,3-glucanases. Interestingly, constitutive expression of a class IV chitinase was reported to coincide with the onset of grape ripening.<sup>47</sup> In a similar way, Vincenzi and Curioni<sup>49</sup> reported the occurrence of chitinases in the mesocarp of grape berries in the absence of any pathogen attack. More recently, Deytieux et al.<sup>50</sup> have confirmed that several PR proteins (chitinases and  $\beta$ -1,3-glucanases) are constitutively expressed in grape skins during ripening. Thus, we cannot exclude the hypothesis that a differential expression in PR proteins may take place upon *B. cinerea* infection of mature grapes, resulting in changes in the electrophoretic patterns (PR3 and PR4 proteins) between the two wines analyzed.

**Yeast Proteins.** Two spots (HW<sub>23</sub> and BW<sub>2</sub>) were identified in both wines as a protein from *S. cerevisiae*, named YJU1 (gil4814). BW<sub>2</sub> was not present in the healthy wine, suggesting that this protein probably came from HW<sub>23</sub> and is partially modified by fungal proteases. This result suggests that *B. cinerea* secretes proteases responsible for a partial hydrolysis of some wine proteins (e.g., YJU1).<sup>22</sup>

YJU1 is a cell-wall mannoprotein, with a theoretical molecular mass of 21.8 kDa and a *pI* of 4.42. Two other yeast proteins matched the same peptides: cell-wall mannoprotein (gil6322754) and Cwp1p (gil45719936). In our experiment, both spots were located at a *M<sub>r</sub>* close to 55 kDa, thus definitely different from the theoretical *M<sub>r</sub>* of YJU1. This deviation could result from an anomalous electrophoretic behavior of YJU1 because of SDS-PAGE conditions and/or PTMs (e.g., glycosylation).

Kwon<sup>12</sup> has recently reported the presence of 12 yeast proteins in a Sauvignon Blanc wine, also including YJU1. In our experiment, a selected group of differentially expressed spots were identified by MS; other yeast proteins might likely be present in the unidentified spots from our sample wines.

**Proteins Secreted by *B. cinerea*.** Two pectinolytic enzymes from *Botryotinia fuckeliana* (teleomorph of *B. cinerea*) were identified in the botrytized wine: endopolygalacturonase 4 (EndoPG 4) (gil3282222, spot BW<sub>2</sub>) and pectin methylesterase (PME) (gil12964194, spots BW<sub>3</sub> and BW<sub>4</sub>). The presence of EndoPG 4 and PME, secreted by *B. cinerea*, in botrytized juices or wines has never been published before, even if these proteins are known to be implicated in the infection process of the fungus. Indeed, these enzymes act consecutively to degrade the plant cell wall. PME are implied in the degradation of pectin in pectic acid and methanol. The PME identified in the botrytized wine is encoded by Bcpme1 cDNA. In a recent study, it was demonstrated that *B. cinerea* strains have multiple PME-encoding genes, but Bcpme1 is mostly essential for the virulence of *B. cinerea*.<sup>51</sup> Then, after the action of PME, endopolygalacturonase, a depolymerizing enzyme, hydrolyzes pectic acid in oligogalacturonates. PME and EndoPG4 correspond to spots identified with anti-*B. cinerea*-secreted protein antibodies.<sup>22</sup>

Numerous grape proteins are absent in the botrytized wine as compared to the healthy wine pattern. Thus, fungal proteins (i.e., PME and EndoPG4) seem to be resistant during the vinification process. This is an important fact because these proteins could be used as markers of rotten harvests.

## Conclusion

This work is the first thorough proteomic analysis of some Champagne base wine proteins, which are considered as “foam-active” agents. Two-dimensional electrophoresis coupled with MS analysis and immunodetection were employed, for the first time, to identify and characterize proteins from two base wines, prepared with either healthy or botrytized grapes. The 37 spots analyzed by nano-LC-MS/MS were chosen among three main groups revealed by a comparative analysis of 2D gels: (i) proteins modified by *B. cinerea* infection, (ii) proteins resistant to *B. cinerea* infection, and (iii) novel proteins in the botrytized wine. Our study showed that both base wines contain multiple isoforms of grape proteins (i.e., invertase and PR proteins) and most of these spots are completely or partially modified by *B. cinerea* infection, particularly, the grape invertase, chitinase, and PR-5 proteins. However, some isoforms of grape PR proteins (e.g., TL) were present in the botrytized wine, and a major spot of the grape invertase (BW<sub>1</sub>) was even induced after fungal infection. On the other hand, two cell-wall-degrading enzymes (endoPG4 and PME) secreted by *B. cinerea* were identified, for the first time, in a base wine vinted from botrytized grape berries. Interestingly, our data also revealed that most of the wine proteins below 23 kDa disappeared following fungal infection. Further research is needed to elucidate the role of those identified proteins in the Champagne wine foaming properties and, particularly, PR proteins (chitinase and thaumatin-like) that are heat-induced haze active proteins.

**Abbreviations:** HW, healthy wine; BW, botrytized wine; 2D, two-dimensional; PAGE, polyacrylamide gel electrophoresis; LC, liquid chromatography; MS/MS, tandem mass spectrometry; IEF, isoelectric focusing; IPG, immobilized pH gradient; SDS, sodium dodecyl sulfate; NCBI, National Center for Biotechnology Information.

**Acknowledgment.** We express our gratitude to the Europol'Agro Institute (Reims, France) and the Association Recherche Oenologie Champagne et Université (Reims,

France) for financial support. We thank the Comité Inter-professionnel du Vin de Champagne (Epernay, France) for providing enological experiments and Champagne Moët and Chandon (Epernay, France) for its collaboration.

**Supporting Information Available:** Proteins from the healthy wine (HW; Table S1) and botrytized wine (BW; Table S2) identified by nano-LC–MS/MS after IEF/SDS–PAGE. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## References

- (1) Brissonnet, F.; Maujean, A. Identification of some foam-active compounds in Champagne base wines. *Am. J. Enol. Vitic.* **1991**, *42* (2), 97–102.
- (2) Malvy, J.; Robillard, B.; Duteurtre, B. Influence des protéines sur le comportement de la mousse des vins de Champagne. *Sci. Aliments* **1994**, *14*, 87–98.
- (3) Moreno-Arribas, V.; Pueyo, E.; Nieto, F. J.; Martin-Alvarez, P. J.; Polo, M. C. Influence of the polysaccharides and the nitrogen compounds on foaming properties of sparkling wines. *Food Chem.* **2000**, *70* (3), 309–317.
- (4) Moine-Ledoux, V.; Perrin, A.; Paladin, I.; Dubourdieu, D. First result of tartaric stabilization by adding mannoproteins (Mannostab). *J. Int. Sci. Vigne Vin* **1997**, *31* (1), 23–31.
- (5) Lubbers, S.; Charpentier, C.; Feuillat, M.; Voilley, A. Influence of yeast walls on the behavior of aroma compounds in a model wine. *Am. J. Enol. Vitic.* **1994**, *45* (1), 29–33.
- (6) Escot, S.; Feuillat, M.; Dulau, L.; Charpentier, C. Release of polysaccharides by yeasts and the influence of released polysaccharides on colour stability and wine astringency. *Aust. J. Grape Wine Res.* **2001**, *7* (3), 153–159.
- (7) Waters, E. J.; Shirley, N. J.; Williams, P. J. Nuisance proteins of wine are grape pathogenesis-related proteins. *J. Agric. Food Chem.* **1996**, *44* (1), 3–5.
- (8) Waters, E. J.; Hayasaka, Y.; Tattersall, D. B.; Adams, K. S.; Williams, P. J. Sequence analysis of grape *Vitis vinifera* berry chitinases that cause haze formation in wines. *J. Agric. Food Chem.* **1998**, *46* (12), 4950–4957.
- (9) Liger-Belair, G. The physics and chemistry behind the bubbling properties of Champagne and sparkling wines: A state-of-the-art review. *J. Agric. Food Chem.* **2005**, *53* (8), 2788–2802.
- (10) Liger-Belair, G.; Religieux, J.-B.; Fohanno, S.; Vialatte, M.-A.; Jeandet, P.; Polidori, G. Visualization of mixing flow phenomena in Champagne glasses under various glass-shape and engraving conditions. *J. Agric. Food Chem.* **2007**, *55* (3), 882–888.
- (11) Dambrouck, T.; Marchal, R.; Marchal-Delahaut, L.; Parmentier, M.; Maujean, A.; Jeandet, P. Immunodetection of proteins from grapes and yeast in a white wine. *J. Agric. Food Chem.* **2003**, *51* (9), 2727–2732.
- (12) Kwon, S. W. Profiling of soluble proteins in wine by nano-high-performance liquid chromatography/tandem mass spectrometry. *J. Agric. Food Chem.* **2004**, *52* (24), 7258–7263.
- (13) Yokotsuka, K.; Singleton, V. L. Glycoproteins: Characterization in a hybrid grape variety (Muscat Bailey A) juice, fermenting must, and resultant red wine. *Am. J. Enol. Vitic.* **1997**, *48* (1), 100–114.
- (14) Marchal, R.; Bouquelet, S.; Maujean, A. Purification and partial biochemical characterization of glycoproteins in a Champenois Chardonnay wine. *J. Agric. Food Chem.* **1996**, *44* (7), 1716–1722.
- (15) Dambrouck, T. Contribution à l'étude des protéines des moûts et des vins de Champagne: Détermination de leur origine; Etude d'une glycoprotéine majeure, l'invertase de raisin. Ph. D. Thesis, Université de Reims, Reims, France, 2004.
- (16) Waters, E. J.; Pellerin, P.; Brillouet, J.-M. *Saccharomyces* mannoprotein that protects wine from protein haze. *Carbohydr. Polym.* **1994**, *23* (3), 185–191.
- (17) Dambrouck, T.; Marchal, R.; Cilindre, C.; Parmentier, M.; Jeandet, P. Determination of the grape invertase content (using PTA–ELISA) following various fining treatments versus changes in the total protein content of wine. Relationships with wine foamability. *J. Agric. Food Chem.* **2005**, *53* (22), 8782–8789.
- (18) Nunez, Y. P.; Carrascosa, A. V.; Gonzalez, R.; Polo, M. C.; Martinez-Rodriguez, A. Isolation and characterization of a thermally extracted yeast cell wall fraction potentially useful for improving the foaming properties of sparkling wines. *J. Agric. Food Chem.* **2006**, *54* (20), 7898–7903.
- (19) Gonzalez-Ramos, D.; Gonzalez, R. Genetic determinants of the release of mannoproteins of enological interest by *Saccharomyces cerevisiae*. *J. Agric. Food Chem.* **2006**, *54* (25), 9411–9416.
- (20) Vanrell, G.; Canals, R.; Esteruelas, M.; Fort, F.; Canals, J. M.; Zamora, F. Influence of the use of bentonite as a riddling agent on foam quality and protein fraction of sparkling wines (Cava). *Food Chem.* **2007**, *104* (1), 148–155.
- (21) Marchal, R.; Berthier, L.; Legendre, L.; Marchal-Delahaut, L.; Jeandet, P.; Maujean, A. Effects of *Botrytis cinerea* infection on the must protein electrophoretic characteristics. *J. Agric. Food Chem.* **1998**, *46* (12), 4945–4949.
- (22) Cilindre, C.; Castro, A. J.; Clement, C.; Jeandet, P.; Marchal, R. Influence of *Botrytis cinerea* infection on Champagne wine proteins (characterized by two-dimensional electrophoresis/immunodetection) and wine foaming properties. *Food Chem.* **2007**, *103* (1), 139–149.
- (23) Marchal, R.; Warchol, M.; Cilindre, C.; Jeandet, P. Evidence for protein degradation by *Botrytis cinerea* and relationships with alteration of synthetic wine foaming properties. *J. Agric. Food Chem.* **2006**, *54* (14), 5157–5165.
- (24) La Guerche, S.; Dauphin, B.; Pons, M.; Blancard, D.; Darriet, P. Characterization of some mushroom and earthy off-odors microbially induced by the development of rot on grapes. *J. Agric. Food Chem.* **2006**, *54* (24), 9193–9200.
- (25) Marchal, R.; Tabary, I.; Valade, M.; Moncomble, D.; Viaux, L.; Robillard, B.; P., J. Effects of *Botrytis cinerea* infection on Champagne wine foaming properties. *J. Sci. Food Agric* **2001**, *81* (14), 1371–1378.
- (26) Okuda, T.; Fukui, M.; Takayanagi, T.; Yokotsuka, K. Characterization of major stable proteins in chardonnay wine. *Food Sci. Technol. Res.* **2006**, *12* (2), 131–136.
- (27) Flamini, R.; De Rosso, M. Mass spectrometry in the analysis of grape and wine proteins. *Expert Rev. Proteomics* **2006**, *3* (3), 321–331.
- (28) Görg, A.; Postel, W.; Weser, J.; Günther, S.; Strahler, J. R.; Hanash, S. M.; Somerlot, L. Elimination of point streaking on silver stained two-dimensional gels by addition of iodoacetamide to the equilibration buffer. *Electrophoresis* **1987**, *8* (2), 122–124.
- (29) Anderson, N. L.; Esquer-Blasco, R.; Hofmann, J.-P.; Anderson, N. G. A two-dimensional gel database of rat liver proteins useful in gene regulation and drug effects studies. *Electrophoresis* **1991**, *12* (11), 907–930.
- (30) Wilkins, M. R.; Appel, R. D.; Van Eyk, J. E.; Chung, M. C. M.; Görg, 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.; Dunn, M. J. Guidelines for the next 10 years of proteomics. *Proteomics* **2006**, *6* (1), 4–8.
- (31) Manteau, S.; Abouna, S.; Lambert, B.; Legendre, L. Differential regulation by ambient pH of putative virulence factor secretion by the phytopathogenic fungus *Botrytis cinerea*. *FEMS Microbiol. Ecol.* **2003**, *43* (3), 359–366.
- (32) ten Have, A.; Dekkers, E.; Kay, J.; Phylip, L. H.; van Kan, J. A. L. An aspartic proteinase gene family in the filamentous fungus *Botrytis cinerea* contains members with novel features. *Microbiology* **2004**, *150* (7), 2475–2489.
- (33) Westbrook, J. A.; Yan, J. X.; Wait, R.; Welson, S. Y.; Dunn, M. J. Zooming-in on the proteome: Very narrow-range immobilised pH gradients reveal more protein species and isoforms. *Electrophoresis* **2001**, *22* (14), 2865–2871.
- (34) Davies, C.; Robinson, S. P. Sugar accumulation in grape berries (cloning of two putative vacuolar invertase cDNAs and their expression in grapevine tissues). *Plant Physiol.* **1996**, *111* (1), 275–283.
- (35) Roitsch, T.; Gonzalez, M.-C. Function and regulation of plant invertases: Sweet sensations. *Trends Plant Sci.* **2004**, *9* (12), 606–613.
- (36) Mills, P. B.; Mills, K.; Johnson, A. W.; Clayton, P. T.; Winchester, B. G. Analysis by matrix assisted laser desorption/ionisation-time of flight mass spectrometry of the post-translational modifications of  $\alpha_1$ -antitrypsin isoforms separated by two-dimensional polyacrylamide gel electrophoresis. *Proteomics* **2001**, *1* (6), 778–786.
- (37) Sturm, A. Invertases. Primary structures, functions, and roles in plant development and sucrose partitioning. *Plant Physiol.* **1999**, *121* (1), 1–8.
- (38) van Loon, L. C.; Rep, M.; Pieterse, C. M. J. Significance of inducible defense-related proteins in infected plants. *Annu. Rev. Phytopathol.* **2006**, *44* (1), 135–162.
- (39) Pastorello, E. A.; Farioli, L.; Pravettoni, V.; Ortolani, C.; Fortunato, D.; Giuffrida, M. G.; Perono Garoffo, L.; Calamari, A. M.; Brenna, O.; Conti, A. Identification of grape and wine allergens as an

- endochitinase 4, a lipid-transfer protein, and a thaumatin. *J. Allergy Clin. Immunol.* **2003**, *111* (2), 350–359.
- (40) Manteau, S. Interaction vigne-Botrytis: Etude des facteurs de virulence de *Botrytis cinerea* et des protéines de défense de la baie. Ph. D. Thesis, Université de Reims Champagne-Ardenne, France, 2003.
- (41) Tattersall, D. B.; van Heeswijck, R.; Hoj, P. B. Identification and characterization of a fruit-specific, thaumatin-like protein that accumulates at very high levels in conjunction with the onset of sugar accumulation and berry softening in grapes. *Plant Physiol.* **1997**, *114* (3), 759–769.
- (42) Sarry, J. E.; Sommerer, N.; Sauvage, F. X.; Bergoin, A.; Rossignol, M.; Albagnac, G.; Romieu, C. Grape berry biochemistry revisited upon proteomic analysis of the mesocarp. *Proteomics* **2004**, *4* (1), 201–215.
- (43) Kraeva, E.; Tesniere, C.; Terrier, N.; Romieu, C.; Sauvage, F. X.; Bierne, J.; Deloire, A. Transcription of a  $\beta$ -1,3-glucanase gene in grape berries in a late developmental period, or earlier after wounding treatments. *Vitis* **1998**, *37* (3), 107–111.
- (44) Jacobs, A. K.; Dry, I. B.; Robinson, S. P. Induction of different pathogenesis, related cDNAs in grapevine infected with powdery mildew and treated with ethephon. *Plant Pathol.* **1999**, *48* (3), 325–336.
- (45) Derckel, J.-P.; Audran, J.-C.; Haye, B.; Lambert, B.; Legendre, L. Characterization, induction by wounding and salicylic acid, and activity against *Botrytis cinerea* of chitinases and  $\beta$ -1,3-glucanases of ripening grape berries. *Physiol. Plant.* **1998**, *104* (1), 56–64.
- (46) Derckel, J.-P.; Legendre, L.; Audran, J.-C.; Haye, B.; Lambert, B. Chitinases of the grapevine (*Vitis vinifera* L.): Five isoforms induced in leaves by salicylic acid are constitutively expressed in other tissues. *Plant Sci.* **1996**, *119* (1–2), 31–37.
- (47) Robinson, S. P.; Jacobs, A. K.; Dry, I. B. A class IV chitinase is highly expressed in grape berries during ripening. *Plant Physiol.* **1997**, *114* (3), 771–778.
- (48) Girbau, T.; Stummer, B. E.; Pocock, K. F.; Baldock, G. A.; Scott, E. S.; Waters, E. J. The effect of *Uncinula necator* (powdery mildew) and *Botrytis cinerea* infection of grapes on the levels of haze-forming pathogenesis-related proteins in grape juice and wine. *Aust. J. Grape Wine Res.* **2004**, *10* (2), 125–133.
- (49) Vincenzi, S.; Curioni, A. Anomalous electrophoretic behavior of a Chitinase isoform from grape berries and wine in glycol chitin-containing sodium dodecyl sulfate–polyacrylamide gel electrophoresis gels. *Electrophoresis* **2005**, *26* (1), 60–63.
- (50) Deytieu, C.; Geny, L.; Lapaillerie, D.; Claverol, S.; Bonneau, M.; Doneche, B. Proteome analysis of grape skins during ripening. *J. Exp. Bot.* **2007**, *58* (7), 1851–1862.
- (51) Valette-Collet, O.; Cimerman, A.; Reignault, P.; Levis, C.; Boccard, M. Disruption of *Botrytis cinerea* pectin methylesterase gene Bcpme1 reduces virulence on several host plants. *Mol. Plant-Microbe Interact.* **2003**, *16* (4), 360–367.

PR070419P

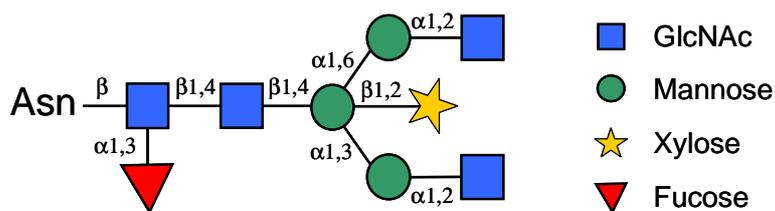
## 2. Caractérisation des glycosylations de l'invertase vacuolaire de raisin

### 2.1. Contexte de l'étude

L'invertase vacuolaire de raisin est l'une des protéines majoritaires du vin [Dambrouck 2004, C. Cilindre et al. 2008] et est responsable de la transformation du saccharose en glucose et fructose dans les plantes. De plus, la première étude protéomique que nous avons réalisée a permis de montrer que cette glycoprotéine est impliquée dans les propriétés moussantes du vin de Champagne [C. Cilindre et al. 2008]. Afin de mieux comprendre les mécanismes moléculaires impliqués dans ce phénomène de moussabilité, nous avons voulu caractériser cette glycoprotéine.

L'invertase possède douze séquences consensus pouvant être potentiellement glycosylées. Une étude a déjà été menée pour caractériser les glycosylations de cette protéine [Dambrouck 2004] en utilisant une technique classique de déglycosylation enzymatique. Cette technique a permis de caractériser les glycanes libres issus de la déglycosylation de l'invertase. En revanche ni les sites de glycosylation, ni la micro-hétérogénéité n'ont pu être identifiés.

Ainsi, les résultats des expériences réalisées par nos collaborateurs, ont montré que l'invertase porte des glycanes de type paucimannosidique (Figure 1). Ce type de glycosylation est composé du noyau complet ( $\text{Man}_3\text{GlcNAc}_2$ ) ou réduit ( $\text{Man}_2\text{GlcNAc}_2$ ) des N-glycanes, sur lequel vient se fixer un  $\alpha(1,3)$ -fucose et/ou un  $\beta(1,2)$ -xylose lié respectivement sur la GlcNAc proximale et le résidu  $\beta$ 3-mannose. Ce type de glycosylation est caractéristique des glycoprotéines de plantes.



**Figure 1.** Structure d'une glycosylation de type paucimannosidique observée chez l'invertase vacuolaire de raisin

Dans un premier temps, nos collaborateurs ont mis au point une méthode de purification de l'invertase qui a donné lieu à une publication (annexe 4). Ensuite, nous avons cherché à identifier les sites de glycosylations et l'hétérogénéité de l'invertase vacuolaire de raisin. Cette étude a été réalisée au début de ma thèse avant le développement de la méthodologie pour la caractérisation des glycoprotéines. De plus nous ne disposons pas encore d'un instrument ayant une précision de mesure de masse suffisante. C'est pourquoi les glycopeptides ont été caractérisés à l'aide d'une analyse nanoLC-MS/MS sur une trappe ionique.

## **2.2. Méthodologie**

### **2.2.1. Purification de l'invertase vacuolaire de raisin**

Une méthode de purification de l'invertase vacuolaire de raisin basée sur de la chromatographie échangeuse d'anions a été développée par nos collaborateurs. La protéine a ensuite été déposée sur un gel d'électrophorèse monodimensionnel SDS-PAGE afin d'en vérifier sa pureté. Une fois révélée par une coloration à l'argent, la bande obtenue a été découpée. La protéine est réduite/alkylée, digérée à la trypsine et analysée par nanoLC-MS/MS sur un couplage nanoLC-Chip/HCT Ultra (Agilent Technologies/Bruker Daltonics). L'ensemble des protocoles et des paramètres expérimentaux est décrit dans la partie matériel et méthode de l'annexe 4.

### **2.2.2. Caractérisation des glycosylations de l'invertase vacuolaire de raisin**

#### **2.2.2.1. Analyse des glycopeptides intacts**

Après purification, l'invertase vacuolaire de raisin a été déposée sur un gel d'électrophorèse bidimensionnel coloré au bleu de Coomassie. Le spot correspondant à la glycoprotéine est découpé. La glycoprotéine est digérée à la trypsine après réduction et alkylation. Les peptides et les glycopeptides extraits sont analysés par nanoLC-MS/MS avec un couplage nanoLC-Chip/HCT Plus (Agilent Technologies/Bruker Daltonics). Les spectres MS/MS correspondants à des glycopeptides ont été recherchés à l'aide des ions diagnostiques (Chapitre 2) et ont été interprétés manuellement.

#### **2.2.2.2. Analyse de la protéine déglycosylée**

L'invertase a été déglycosylée avec de la PNGase A qui coupe au niveau de la liaison entre le GlcNAc proximal et l'asparagine. La protéine a ensuite été dialysée dans de l'eau. Après réduction et alkylation elle a été digérée à la chymotrypsine puis analysée par nanoLC-MS/MS sur un couplage nanoLC-Chip/amaZon (Agilent Technologies/Bruker Daltonics). Les données obtenues ont été interprétées à l'aide du moteur de recherche Mascot, par recherche dans une banque de donnée restreinte à la taxonomie *Vitis vinifera*, extraite de NCBI nr.

## **2.3. Résultats**

### **2.3.1. Purification de l'invertase vacuolaire de raisin**

La pureté de la protéine a été vérifiée par gel 1D puis une analyse protéomique a permis d'identifier l'invertase vacuolaire GIN1 (gi|1839578) avec un pourcentage de recouvrement de 14%. Cette protéine a un poids moléculaire de 71 kDa ce qui peut expliquer ce faible taux de recouvrement.

De plus, l'invertase est fortement glycosylée, et nous savons que les chaînes glycaniques peuvent gêner voire empêcher l'action des endoprotéases à cause du fort encombrement stérique qu'elles provoquent.

### **2.3.2. Caractérisation des glycosylations de l'invertase vacuolaire de raisin**

#### **2.3.2.1. Analyse des glycopeptides intacts**

Lors de l'étude réalisée précédemment par nos collaborateurs [Dambrouck 2004], seuls les glycanes libres issus de la déglycosylation de l'invertase avaient été caractérisés. Notre objectif a donc été de déterminer les sites de glycosylations et la micro-hétérogénéité au niveau de ces sites.

Ce travail a été réalisé au début de ma thèse, et la méthodologie pour la caractérisation des glycoprotéines qui est détaillée dans le Chapitre 2 n'était pas encore développée. C'est pourquoi la caractérisation de la glycoprotéine a été réalisée à partir d'une analyse nanoLC-MS/MS sur une trappe ionique dont le paramétrage a été optimisé pour sélectionner et fragmenter des ions de hautes masses et donc favoriser les glycopeptides.

Cette analyse a permis d'identifier la protéine nommée : Predicted, hypothetical protein (gi|225466093) avec un peptide discriminant entouré en noir sur la Figure 2. Cette protéine possède une très forte homologie de séquence avec l'invertase vacuolaire GIN1 (gi|1839578) et possède les mêmes sites consensus de glycosylation (Figure 2).

PREDICTED: hypothetical protein [Vitis vinifera]:gi|225466093

vacuolar invertase 1, GIN1 [Vitis vinifera]:gi|1839578

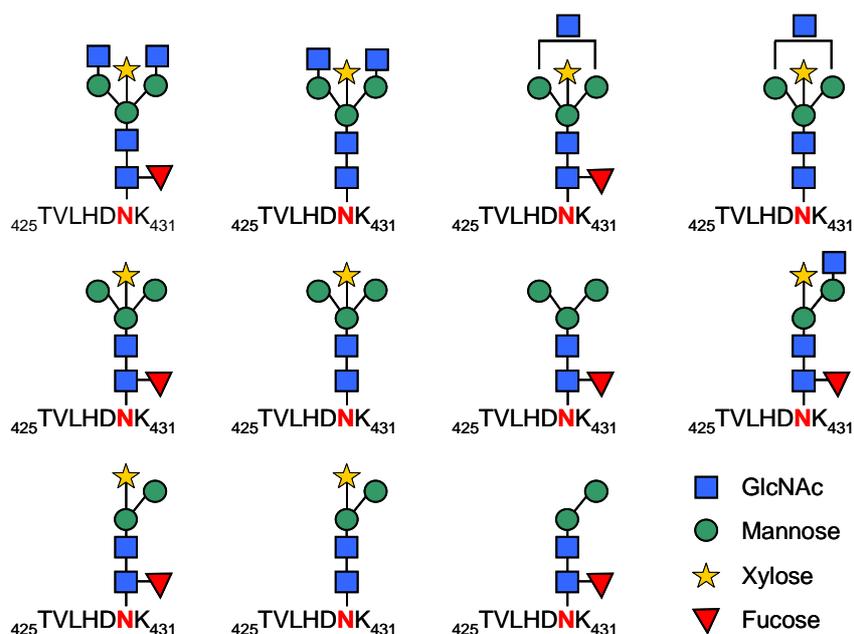
```

hypothetical      1MDTHHATSHD LLHDEAYAPL PYSPLPDHPV PAGAPALGQR RPLKVFAVTL ASLMFLASLV ALVIHQSPQT PVTVDNDGPS
GIN1              MDTHHATSHD LLHDEAYAPL PYSPLPDHPV PAGAPALGQR RPLKV CVTL SLMFLVSLV SFI      RV hVTVDNDGPS
*****          *****          *****          *****          *****          *****          *****          *****
hypothetical     81MEREGRGVAQ GVSEKSFLGF SGRRLSYNWT TAMFAWQRTA FHFQPEKNWM NDPDGPLFHM GWYHLFYQYN PDSAVWGNIT
GIN1             MEREGRGVAQ GVSEKSFLGF SGRRLSYNWT TAMFAWQRTA FHFQPEKNWM NDPDGPLFHM GWYHLFYQYN PDSAVWGNIT
*****          *****          *****          *****          *****          *****          *****          *****
hypothetical     161WGHAVSRDMI HWLYLPLAMV PDRWFDLNGV WTGSATILPN QIIMLYTGD TNDSVQVQNL AYPANLSDPL LLHWIKYENN
GIN1            WGHAVSRDMI HWLYLPLAMV PDRWFDLNGV WTGSATILPN QIIMLYTGD TNDSVQVQNL AYPANLSDPL LLHWIKYENN
*****          *****          *****          *****          *****          *****          *****          *****
hypothetical     241PVMVPPAGIS SDDFRDPTTM WVGADGNWRV AVGSLVNTTG IVLVFQTTNF TDFELLDGEL HGVPGTGMWE CVDFYPVSIN
GIN1            PVMVPPAGIS SDDFRDPTTM WVGADGNWRV AVGSLVNTTG IVLVFQTTNF TDFELLDGEL HGVPGTGMWE CVDFYPVSIN
*****          *****          *****          *****          *****          *****          *****          *****
hypothetical     321GVYGLDTSAH GPGIKHVLKA SMDDNRHDYY ALGEYDPMTD TWPDDPELD VGIGLRLDYGRKYYASKTFYD QVKKRRILYG
GIN1            GVYGLDTSAH GPGIKHVLKA SMDDNRHDYY ALGEYDPMTD TWPDDPELD VGIGLRLDYGRKYYASKTFYD QVKKRRILYG
*****          *****          *****          *****          *****          *****          *****          *****
hypothetical     401WISEGDIESD DLKKGWASLQ SIPRTVLHN KTGTLLLWP IEEVESLRTN STEFEDVLLE PGSVVPLDIG SASQLDIVAE
GIN1            WISEGDIESD DLKKGWASLQ SIPRTVLHN KTGTLLLWP IEEVESLRTN STEFEDVLLE PGSVVPLDIG SASQLDIVAE
*****          *****          *****          *****          *****          *****          *****          *****
hypothetical     481FEVDNETLEA MVEADVIYNC STSAGAAGRG ALGPFGLLVL ADDTLSELTP IYFYIAKDTD GSYKTFFCTD LSRSSLAVDD
GIN1            FEVDNETLEA MVEADVIYNC STSAGAAGRG ALGPFGLLVL ADDTLSELTP IYFYIAKDTD GSYKTFFCTD LSRSSLAVDD
*****          *****          *****          *****          *****          *****          *****          *****
hypothetical     561VDQRIYGSIV PALDDEKPTM RVLVDHSIVE GFSQGRSCI TTRVYPTEAI YGAARLFLFNN NATGVNVTAS IKIWEMASAD
GIN1            VDQRIYGSIV PALDDEKPTM RVLVDHSIVE GFSQGRSCI TTRVYPTEAI YGAARLFLFNN NATGVNVTAS IKIWEMASAD
*****          *****          *****          *****          *****          *****          *****          *****
hypothetical     641IHPYPLDQP
GIN1            IHPYPLDQP
*****

```

**Figure 2.** Alignement de séquences entre la protéine identifiée (hypothetical protein) et la protéine recherchée (vacuolar invertase1, GIN1); les sites consensus de glycosylation sont surlignés en bleu; les différences de séquences sont écrit en vert

Afin de déterminer les zones d'éluion des glycopeptides, les ions diagnostiques ont été recherchés dans les spectres MS/MS. Ainsi, un site de glycosylation a pu être identifié (Asn<sup>430</sup>) et 11 glycoformes ont été caractérisées au niveau de ce site (Figure 3).



**Figure 3.** Structures des glycopeptides identifiées par MS/MS au niveau du site Asn<sup>430</sup> de la protéine "Predicted : hypothetical protein"

L'analyse des glycopeptides intacts n'a donc permis d'identifier qu'un seul site de glycosylation sur les douze potentiels. Le peu de site identifié peut être lié à la composition des glycosylations (Figure 1). En effet, les ions oxoniums générés par ce type de glycosylation ont des masses assez basses ( $m/z$  163 et 204). Les analyses ayant été effectuées sur une trappe ionique, ces ions sont peu intenses dans les spectres MS/MS. Les spectres correspondant aux glycopeptides sont donc difficiles à extraire. De plus, la trypsine génère des peptides de taille relativement élevée, et certains portent plusieurs sites consensus. Les glycopeptides recherchés sont donc de haut poids moléculaire et les ions correspondants sont difficiles à piéger et à fragmenter dans la trappe ionique. Enfin, il est possible que certains sites ne soient que partiellement glycosylés, les glycopeptides correspondant sont alors largement minoritaires dans le digestat.

Afin de générer des glycopeptides plus petits, l'invertase pourrait être digérée avec de la chymotrypsine. Ensuite en analysant le digestat avec un Q-TOF, les ions diagnostiques de bas poids moléculaire seraient plus facilement observables et les glycopeptides de haut poids moléculaire pourraient être isolés et fragmentés plus efficacement.

Cette expérience n'ayant permis d'identifier qu'un seul site de glycosylation, nous avons poursuivi l'étude en utilisant une méthode classique de déglycosylation enzymatique afin de focaliser notre recherche sur l'identification des sites de glycosylation. Ce travail a pu être réalisé car nous avons à notre disposition une quantité suffisante de matériel biologique.

### 2.3.2.2. Analyse de la protéine déglycosylée

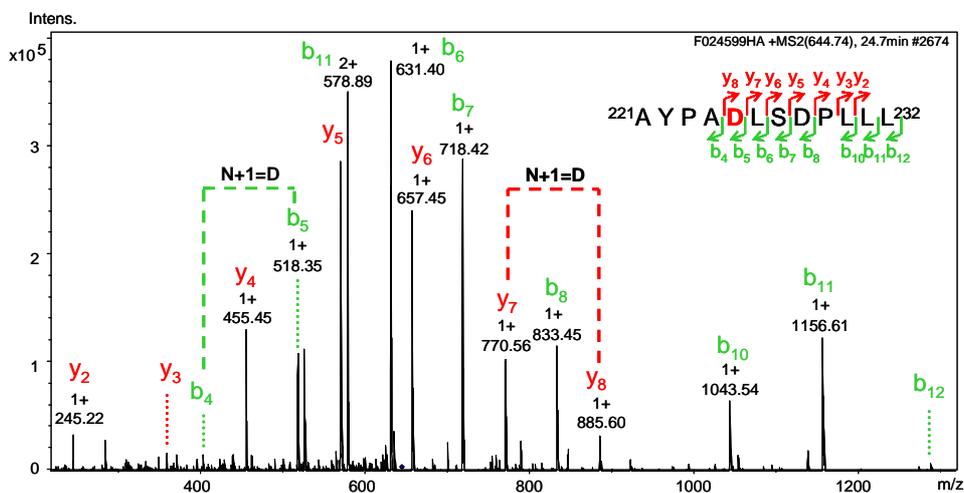
Les glycanes portés par l'invertase ont déjà été identifiés par nos collaborateurs en utilisant une déglycosylation enzymatique à la PNGase A [Dambrouck 2004]. Ainsi, afin de compléter cette étude,

et identifier les sites de glycosylation, nous avons analysé la protéine déglycosylée. La digestion par la PNGase déamide l'asparagine portant la glycosylation en acide aspartique et provoque ainsi une différence de masse de +1 Da, ce qui permet d'identifier les peptides ayant porté une glycosylation.

Afin d'obtenir un ensemble de peptides de taille homogène et ne comportant qu'un seul site de glycosylation, la protéine a été digérée avec de la chymotrypsine. Le digestat a été analysé par nanoLC-MS/MS avec la trappe ionique amaZon (Bruker Daltonics). Cet instrument possède une précision de mesure de masse de  $\pm 0,2$  Da en MS (mode "Standard Enhanced") et en MS/MS (mode "Ultra Scan") ce qui est suffisant pour observer une différence de masse de 1 Da pour des peptides portant moins de deux charges. La trappe amaZon a surtout pour avantage de générer des spectres de fragmentation d'excellente qualité, ce qui est particulièrement important lorsque l'on veut caractériser par MS/MS une modification au niveau de la séquence en acides aminés d'un peptide.

Cette analyse a permis d'identifier la protéine nommée Predicted, hypothetical protein (gij225466093) avec 70% de recouvrement avec 5 peptides discriminant (entourés en vert sur la Figure 2) par rapport à l'invertase GIN1 (gij1839578). L'interprétation des données a été effectuée à l'aide du logiciel Mascot par recherche dans une banque restreinte à la taxonomie *Vitis vinifera* extraite de NCBI. La modification Asn→Asp a été spécifiée comme modification variable. Chaque spectre MS/MS identifié par Mascot comme correspondant à un peptide modifié a été validé manuellement (Figure 4).

Ainsi, cinq peptides sur les 12 potentiels ont été identifiés comme portant la modification Asn→Asp au niveau des sites consensus connus, soit en position 225, 277, 289, 430 et 499.



**Figure 4.** Spectre MS/MS obtenu avec la trappe ionique amaZon du peptide portant Asn<sup>225</sup> transformé en Asp<sup>225</sup> après déglycosylation avec la PNGase A et ayant permis l'identification du site de glycosylation Asn<sup>225</sup>

Des analyses complémentaires ont été réalisées sur un Q-TOF maXis, qui permet une précision de mesure de masse de 5 ppm en MS et de 0,02 Da en MS/MS, et ont confirmé les résultats obtenus avec la trappe ionique. Par conséquent quatre sites de glycosylation supplémentaires ont été identifiés par rapport à l'étude sur les glycopeptides intacts qui avait permis de caractériser le site Asn<sup>430</sup>.

Cependant, une différence de masse de 1 Da peut également correspondre à une déamidation *in vitro* qui provoque des faux positifs [Robinson et Robinson 2001]. En effectuant la déglycosylation à la PNGase dans de l'eau lourde ( $H_2O^{18}$ ), la différence de masse générée par la transformation de l'asparagine en acide aspartique est de 3 Da [Gonzalez et al. 1992, Küster et Mann 1999, Kaji et al. 2003] ce qui permet de lever l'ambiguïté. Cette expérience pourrait être réalisée afin de confirmer les résultats.

## 2.4. Conclusion et perspectives

L'objectif de cette étude était la caractérisation des glycosylations de l'invertase vacuolaire de raisin. Deux approches complémentaires ont été utilisées. Dans un premier temps, l'analyse des glycopeptides intacts a permis de déterminer la micro-hétérogénéité, ensuite l'analyse des peptides déglycosylés a mis en évidence la macro-hétérogénéité de la glycoprotéine.

Grâce à l'analyse des glycopeptides intacts, le site Asn<sup>430</sup> a été identifié comme étant glycosylé. Onze glycoformes ont été caractérisées par MS/MS, mettant ainsi en évidence la micro-hétérogénéité au niveau de ce site de glycosylation. Les recherches doivent être poursuivies pour caractériser plus de glycosylations en utilisant d'autres enzymes et d'autres spectromètres de masse (chymotrypsine, Q-TOF) afin d'obtenir des résultats complémentaires.

L'identification des sites d'attachement des glycanes sur la chaîne peptidique à l'aide d'un clivage enzymatique a permis de caractériser cinq sites de glycosylation dont Asn<sup>430</sup> (Asn<sup>225</sup>, Asn<sup>277</sup>, Asn<sup>289</sup>, Asn<sup>430</sup> et Asn<sup>499</sup>). La déamidation d'une asparagine portant une glycosylation en acide aspartique constitue une preuve indirecte de la présence d'un site de N-glycosylation.

Une autre approche utilisable pour déterminer les sites de glycosylation est l'utilisation d'une endoglycosydase qui permet une déglycosylation partielle en laissant le GlcNAc proximal fixé à l'asparagine [Segu et al. 2010]. L'utilisation de ce type d'enzyme permet d'obtenir une preuve plus directe de la présence d'un site de glycosylation en générant une différence de masse de 203 Da (masse d'un GlcNAc) par rapport à l'asparagine nue. Cependant, il n'existe pas d'enzyme compatible avec tous les types de glycosylation. Dans notre cas, de nombreux glycanes portent un fucose en  $\alpha$ 1-3 du GlcNAc proximal qui gêne l'action des endoglycosydases les plus courantes (endo H, endo M). L'enzyme endo F2 est compatible avec le type de glycosylations portées par l'invertase mais son utilisation reste délicate.

Enfin la plupart des analyses réalisées n'ont pas permis d'identifier directement l'invertase vacuolaire de raisin GIN1, mais une isoforme appelée "hypothetical protein". Des analyses doivent donc être effectuées afin de caractériser la séquence exacte de cette protéine. Ces expériences devront être réalisées sur la protéine déglycosylée car l'hétérogénéité engendrée par les nombreuses glycosylations rend l'analyse de la glycoprotéine intacte particulièrement complexe.



# **ANNEXE 4: Publication**

## **One step purification of the grape vacuolar invertase**

Sandrine Jégou, Alexandra Conreux, Sandra Villaume, Agnès Hovasse, Christine Schaeffer, Clara Cilindre, Alain Van Dorsselaer, Philippe Jeandet

Analytica Chimica Acta, 2009, 75-78





## One step purification of the grape vacuolar invertase

Sandrine Jégou<sup>a,\*</sup>, Alexandra Conreux<sup>a</sup>, Sandra Villaume<sup>a</sup>, Agnès Hovasse<sup>b</sup>, Christine Schaeffer<sup>b</sup>, Clara Cilindre<sup>a</sup>, Alain Van Dorsselaer<sup>b</sup>, Philippe Jeandet<sup>a</sup>

<sup>a</sup> Laboratoire d'Oenologie et Chimie Appliquée, Université de Reims Champagne-Ardenne, URVVC UPRES EA 2069, BP 1039, 51687 Reims Cedex 2, France

<sup>b</sup> Laboratoire de Spectrométrie de Masse Bio-Organique, Institut Pluridisciplinaire Hubert Curien IPHC-DSA, ULP, CNRS, UMR 7178, ECPM, 25 rue Becquerel, 67087 Strasbourg Cedex 2, France

### ARTICLE INFO

#### Article history:

Received 4 November 2008  
Received in revised form 28 January 2009  
Accepted 2 February 2009  
Available online 11 February 2009

#### Keywords:

Grape must  
Protein  
Invertase  
Purification  
Ion exchange chromatography  
Mass spectrometry

### ABSTRACT

Invertase is a major protein of grape juice and wine. Accordingly, in order to study the biochemical and structural characteristics of this protein and for a better understanding of its physico-chemical properties, large amounts of the pure protein are needed. A simple method for the purification of the grape vacuolar invertase in a preparative-scale is described in this work. The grape protein was isolated and purified from must by ultrafiltration and anion exchange chromatography. The identification and purity determination of the grape invertase fraction were assessed by SDS-PAGE, and were then confirmed using nanoLC-chip-MS/MS analysis. The laboratory fractionation procedure presented in this work generated large quantities of pure grape vacuolar invertase from must.

© 2009 Published by Elsevier B.V.

### 1. Introduction

Paradoxically, proteins are present at low levels in wine, their concentration usually ranging from 4 to 20 mg L<sup>-1</sup> [1], but they widely contribute to its quality. Indeed, proteins are of great importance in the winemaking process, playing a role in wine stability and clarity [2,3] as well as in the foaming properties of sparkling wines and Champagne [4,5].

The invertase originating from grape, responsible for the conversion of sucrose into glucose and fructose, is believed to be one of the most abundant proteins in wine (from 9 to 14% of the total protein content of a Chardonnay wine) [6,7]. The grape vacuolar invertase ( $\beta$ -D-fructofuranoside fructohydrolase, EC 3.2.1.26) is a N-glycoprotein [8–10], which has a molecular mass estimated at 60–65 kDa by SDS-PAGE analysis, or 72 kDa determined by gel filtration [11], and a pI 3.9 [10]. A recent study has shown that a significant decrease in both the total protein and the grape invertase contents of Champagne base wines is correlated with the decrease of wine foaming properties [6]. This last work underlies the role

played by the grape invertase in Champagne wine foaming properties although other related works reported that this protein is not implied in foam stability of Champagne [12]. The grape invertase was previously purified from wine or must obtained from different grape varieties, but, to date, the protein is not entirely characterized [6,8–11,13,14]. The fractionation schemes included numerous steps such as affinity chromatography with Concanavalin A, preparative isoelectric focusing, preparative SDS-PAGE, ion exchange or gel filtration chromatography. Unfortunately, scarce information were available from these studies about the yields of the purification procedures or low quantities of pure grape invertase were isolated and reported [6,8,11].

In this work, a simple method for separation and purification of the grape invertase from must is presented. The preparative-scale procedure described allows the isolation of pure grape vacuolar invertase and the preparation of large amounts of this protein, using Q-Sepharose Fast Flow chromatography, for subsequent studies relating to enological fining treatments or wine foaming properties.

### 2. Experimental

#### 2.1. Chemicals

Tris(hydroxymethyl)aminomethane, chloridric acid, tetramethylethylenediamine, ammonium persulfate and silver nitrate were purchased from Sigma–Aldrich (Steinheim, Germany). Sodium chloride, formaldehyde and sodium dodecyl sulfate were

*Abbreviations:* SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis; LC, liquid chromatography; MS/MS, tandem mass spectrometry; NCBI, National Center for Biotechnology Information; EDTA, ethylenediaminetetraacetic acid; DTT, dithiothreitol.

\* Corresponding author. Tel.: +33 326 91 86 14; fax: +33 326 91 86 14.

E-mail address: [sandrine.jegou@univ-reims.fr](mailto:sandrine.jegou@univ-reims.fr) (S. Jégou).

obtained from Acros Organics (New Jersey, USA). Acetic acid and ethanol were purchased from Fisher Scientific (Leicestershire, UK) and Charbonneaux-Brabant (Reims, France), respectively. Glycine and sodium thiosulfate were obtained from VWR Pro-labo (Fontenay-sous-Bois, France), and acrylamide solution was obtained from Bio-Rad (USA). The chromatography column was packed with Q-Sepharose Fast Flow (GE Healthcare, Uppsala, Sweden). All chemicals and reagents for mass spectrometry analysis were of analytical grade and purchased from Sigma–Aldrich (Steinheim, Germany). Porcine trypsin was purchased from Promega (Madison, WI, USA). Deionized water produced from a Milli-Q Water System™ (Millipore, Guyancourt, France) or an Arium 611 UV (Sartorius, Palaiseau, France) was used for the preparation of the buffers and solvents.

## 2.2. Plant material and must fraction

Grape berries of *Vitis vinifera* cv. Chardonnay were hand-harvested in the Champagne area (France), in September 2007. Grapes were pressed with a pneumatic press, applying a pressure between 1.5 and 2 bars. Sulphur dioxide ( $7 \text{ g h L}^{-1}$ ) was added to the free run juice and the must was collected after static settling for 24 h.

## 2.3. Macromolecule isolation

The settled Chardonnay must (41.6 L) was centrifuged at 8000 g for 10 min. The supernatant was separated and filtered through a  $0.45 \mu\text{m}$  membrane (Alltech, Deerfield, USA). Must samples were concentrated seven times and dialyzed three times against distilled water. A hydrophilic polysulfone membrane with a molecular weight cutoff of 10 kDa (30 UFIB, Setric Génie Industriel, SGI, France) was used. The crossflow filtration module was connected to the SGI Hi-Flow system (pumping system plus glass tank) and ultrafiltrations were carried out at  $4^\circ\text{C}$ . Dialyzed retentates were freeze-dried (SERAIL CS 5L) before purification.

## 2.4. Purification of the grape vacuolar invertase

The grape invertase was purified from the concentrated must. Chromatography experiments were conducted using an ÄKTA™ Prime Plus system (GE Healthcare, Uppsala, Sweden). The lyophilized ultraconcentrate was solubilized in 50 mM Tris–HCl buffer pH 7.5 and loaded on a Q-Sepharose Fast Flow anion exchange column ( $170 \times 50 \text{ mm}$ ) at room temperature. The column was equilibrated with 50 mM Tris–HCl buffer pH 7.5 and proteins were eluted at a flow rate of  $30 \text{ mL min}^{-1}$  with a gradient from 0 to 500 mM NaCl in Tris–HCl buffer pH 7.5. Eluted proteins were detected by measurement of absorbance at 280 nm.

The collected fractions were analyzed by SDS-PAGE and the fractions containing proteins with molecular mass around 60 kDa were pooled. The eluted protein was desalted and concentrated using Microcon YM-10 cartridges (Millipore, Bedford, MA), and freeze-dried.

## 2.5. SDS-PAGE analysis

The protein composition of samples during macromolecule isolation and purification of the grape vacuolar invertase was assessed by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE). Vertical SDS-PAGE was performed in a Mini-Protean III electrophoresis cell (Bio-Rad, USA) according to the method of Laemmli [15], using a 4% stacking gel and a 13% resolving gel. Proteins were separated under a constant voltage of 150 V. Standard protein markers (Precision Plus Protein™ Standards, Bio-Rad, USA) were simultaneously loaded at each run. The relative molecular masses ( $M_r$ ) of proteins were calculated from the linear regres-

sion equation of  $\log M_r$  vs. mobility. After electrophoresis, gels were silver-stained [16].

## 2.6. Reduction/alkylation of proteins and nanoLC-chip-MS/MS analysis

Analysis, identification and purity determination of the grape vacuolar invertase was performed by nanoLC-chip-MS/MS.

### 2.6.1. Reduction-alkylation-digestion

The lyophilized must protein fraction ( $250 \mu\text{g}$ ) was solubilized in 8 M urea and 0.1 M  $\text{NH}_4\text{HCO}_3$  pH 8 buffer for a final concentration of  $8.3 \text{ pmol } \mu\text{L}^{-1}$ . The cysteine residues were reduced by incubating the protein solution with 700 mM DTT for 30 min at  $37^\circ\text{C}$ . Alkylation of cysteines was performed by adding 700 mM iodoacetamide; the reaction occurred in the dark at room temperature within 1 h. The sample was then diluted to 1 M urea with 0.1 M  $\text{NH}_4\text{HCO}_3$ . The digestion was made by  $2.5 \mu\text{g}$  of trypsin at  $37^\circ\text{C}$  overnight. The tryptic digest was analyzed by nanoLC-chip-MS/MS.

### 2.6.2. Chromatography

The LC system used was an Agilent 1100 series nano-HPLC-Chip system (Agilent Technologies, Palo Alto, USA). The chip was composed of a Zorbax 300SB-C18 ( $150 \text{ mm} \times 75 \mu\text{m}$ , with a  $5 \mu\text{m}$  particle size) analytical column and a Zorbax 300SB-C18 ( $40 \text{ nL}$ ,  $5 \mu\text{m}$ ) enrichment column. The solvent system consisted of 2% acetonitrile, 0.1% formic acid in water (solvent A) and 2% water, 0.1% formic acid in acetonitrile (solvent B). The sample was loaded into the enrichment column at a flow rate set to  $3.75 \mu\text{L min}^{-1}$  with an  $\text{H}_2\text{O}/2\% \text{ acetonitrile}/0.1\% \text{ HCOOH}$  solution. Elution was performed at a flow rate of  $300 \text{ nL min}^{-1}$  with a 8–40% linear gradient (solvent B) over 30 first minutes, followed by a 70% stage (solvent B) over 3 min before the reconditioning of the column at 92% of solvent A.

### 2.6.3. Mass spectrometry

The nano-HPLC-Chip system was coupled with an HCT Ultra ion trap (Bruker Daltonics, Bremen, Germany). The complete system was fully controlled by ChemStation (Agilent Technologies) and EsquireControl (Bruker Daltonics) softwares.

The voltage applied to the capillary cap was optimized to  $-1750 \text{ V}$  and the mass range was 250–2500  $m/z$ . For tandem MS experiments, the system was operated with automatic switching between MS and MS/MS modes. The three most abundant peptides of each  $m/z$  were selected to be further isolated and fragmented. The MS/MS scanning was performed in the ultrascan resolution mode at a scan rate of 26,000  $m/z$  per second. A total of six scans were averaged to obtain an MS/MS spectrum.

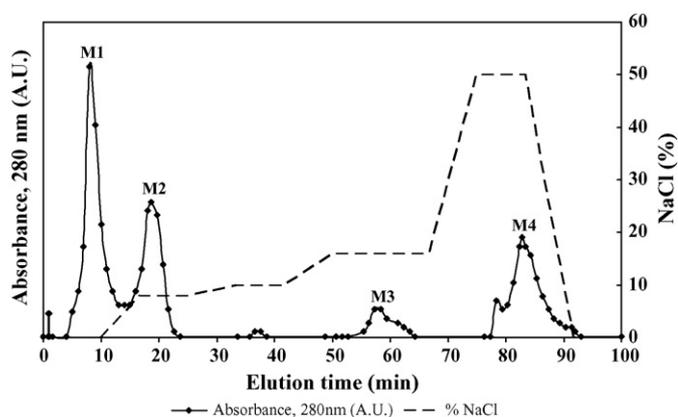
Mass data collected during nanoLC-chip-MS/MS analysis were processed, converted into \*.mgf files and were interpreted using a local Mascot (Matrix Science, London, UK) server running on an Intel Xeon biprocessor quad core 3 GHz. Searches were performed against the NCBIInr database without any taxonomic, molecular weight or isoelectric point restrictions.

The tolerance on mass measurements was of 0.25 Da both in MS and MS/MS modes. One missed cleavage per peptide was allowed and some variable modifications were taken into account such as carbamidomethylation for cysteine and oxidation for methionine.

## 3. Results and discussion

### 3.1. Protein fractionation and purification

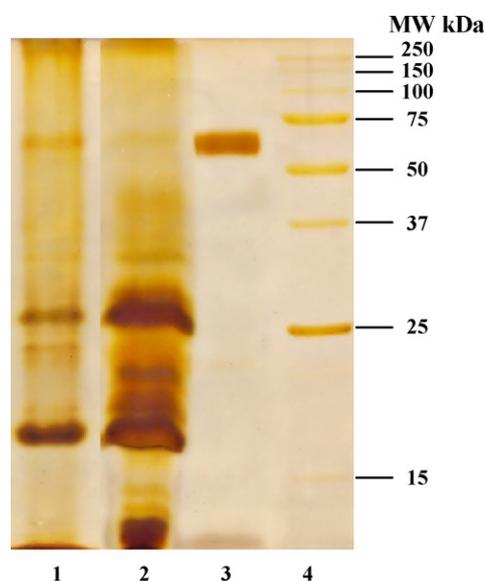
The fractionation of the crude Chardonnay must extract obtained after ultrafiltration was performed on a Q-Sepharose Fast Flow anion exchange column and a NaCl gradient from 0 to 500 mM was applied to elute proteins. The anion exchange chromatogram



**Fig. 1.** Chromatographic pattern of eluted proteins from the ultraconcentrated must on Q-Sepharose Fast Flow anion exchange column. The column was equilibrated with 50 mM Tris-HCl buffer pH 7.5 and protein elution was performed with a linear gradient from 0 to 500 mM NaCl in Tris-HCl buffer pH 7.5.

profile showed four main peaks M1, M2, M3 and M4, eluted at 80, 100, 160 and 500 mM NaCl, respectively (Fig. 1). Each peak was collected as several fractions which were then pooled for subsequent analysis. The must protein content after centrifugation and 10 kDa ultrafiltration, and the fractions eluted by anion exchange chromatography were analyzed by SDS-PAGE (Fig. 2); protein detection was achieved by silver-staining, a more sensitive stain procedure than Coomassie brilliant blue staining. The must protein pattern (Fig. 2, lane 1) showed several major bands with apparent molecular masses ranging from 16 to 69 kDa. After 10 kDa ultrafiltration, the electrophoretic profile of must proteins revealed ten bands (Fig. 2, lane 2), of which five bands with apparent molecular masses of 12, 17.5, 19.5, 21.7, and 27 kDa, respectively, appeared with a high coloration intensity. In contrast, the band at 60–65 kDa was less intense in the must ultraconcentrate than in the total protein fraction. These results are somewhat different compared to previous data obtained from the analysis of Chardonnay must proteins isolated by a 10 kDa ultrafiltration, after Coomassie brilliant blue staining [4]. The affinity of silver and Coomassie brilliant blue for each protein and the sensitivity of the two stain procedures could account for the different electrophoretic patterns observed.

After preparative fractionation on the Q-Sepharose Fast Flow anion exchange column, the electrophoretic profile of the fractions M1, M2 and M4 revealed the presence of different grape proteins (data not shown). The fractions corresponding to peak M3 were pooled, concentrated and desalted before freeze-drying. Analysis of the lyophilized fraction M3, as performed by SDS-PAGE, showed a single band of apparent molecular masses in the 59–69 kDa range (Fig. 2, lane 3). At first, this protein band was attributed to a grape



**Fig. 2.** SDS-PAGE analysis of the Chardonnay must protein fractions visualized by silver-staining. Lane 1, must total proteins after settling by centrifugation; lane 2, must proteins isolated by a 10 kDa ultrafiltration; lane 3, protein fraction M3 purified on a Q-Sepharose Fast Flow anion exchange column (Fig. 1); lane 4, molecular weight standard proteins (kDa). The molecular mass was assigned to each band using the Quantity One software.

glycoprotein identified as the vacuolar invertase [8,9], and the SDS-PAGE analysis indicated that the purified fraction M3 contained a single protein.

### 3.2. Identification and purity determination

In order to assess purity and precisely identify proteins, fraction M3 was hydrolyzed with trypsin and submitted to nanoLC-chip-MS/MS analysis. Protein composition of the lyophilized fraction M3 is presented in Table 1. The protein was identified as the vacuolar invertase 1, GIN1 (*Vitis vinifera*, gi|1839578), which corresponds to the preferential expressed form encoded by GIN1 cDNA in berries [17]; the putative amino acid sequence of GIN1 deduced from cDNA is shown in Fig. 3. Nine tryptic digested peptides were identified by nanoLC-chip-MS/MS analysis (Table 1), which corresponded to well-distributed peptides along the protein sequence. These data thus suggest that the grape vacuolar invertase purified according to the fractionation procedure described in this study is not a truncated protein. This is in agreement with the apparent molecular mass of the protein in the 59–69 kDa range, as estimated by SDS-PAGE analysis. NanoLC-chip-MS/MS analysis allows us to confirm that the initial identification of the single band from fraction

**Table 1**

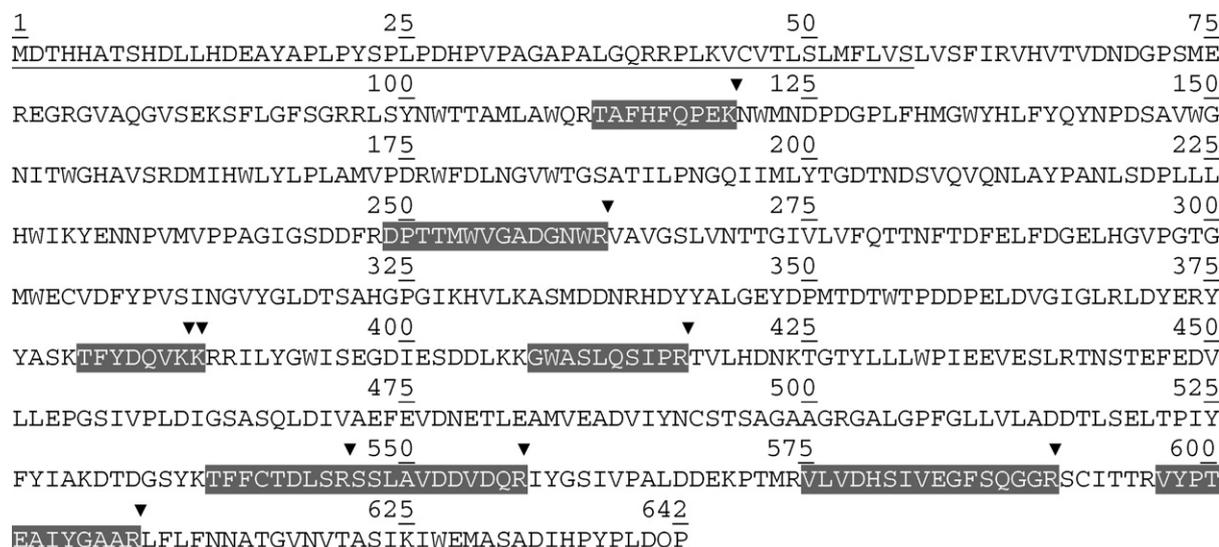
Protein identified by nanoLC-chip-MS/MS analysis from the lyophilized fraction M3 purified by Q-Sepharose Fast Flow anion exchange chromatography.

Protein name (organism)	NCBI accession number <sup>a</sup>	MW (theoretical)	pI	Number of peptides matched <sup>b</sup>	Sequence coverage <sup>c</sup>	Start–end	Peptide sequence
Vacuolar invertase 1, GIN1 ( <i>Vitis vinifera</i> )	gi 1839578	71.5 kDa	4.60	9	14%	112–120	TAFHFQPEK
						249–262	DPTTMWVGADGNWR
						380–386	TFYDQVK
						380–387	TFYDQVKK
						408–417	GWASLQSIPR
						538–546	TFCTDLR
						547–557	SSLAVDDVDQR
						575–590	VLVDHSIVEGFSQGGK
						597–608	VYPTEAIYGAAR

<sup>a</sup> Protein identification number provided by the NCBI database.

<sup>b</sup> Number of peptides identified by the MASCOT search engine, that matched with the identified protein.

<sup>c</sup> Percentage ratio of all amino acids from valid peptides matched to the total number of amino acids in the reference protein.



**Fig. 3.** Amino acid sequences of putative grape invertase GIN1 [17] and peptides identified by the MASCOT search engine from digestion by trypsin and nanoLC-chip-MS/MS analysis. Residue numbers refer to the protein including the signal peptide. The putative signal peptide region is indicated by single underlining. Peptides matching with the grape vacuolar invertase sequence are shaded, and tryptic cleavage sites are pointed by arrows.

M3 was correct. GIN1 consists of 642 amino acid residues and is characterized by a theoretical molecular mass of 71.5 kDa and a *pI* of 4.6. Moreover, the protein displays twelve potential sites of *N*-glycosylation. The large size of the grape invertase and the possible presence of glycans on the protein backbone, restricting the digestion by trypsin, could explain the low sequence coverage percentage of 14% obtained for GIN1.

From nanoLC-chip-MS/MS analysis and database-searching program Mascot, a second protein was identified as VVTL1 (*Vitis vinifera*, gi|2213852) for fraction M3, with a theoretical molecular mass around 24 kDa (data not shown). VVTL1 is a pathogenesis-related (PR) protein from grape origin belonging to the thaumatin-like family (TL). Waters et al. [2,18] showed that VVTL1 was one of the major heat-unstable proteins in white wine. The identification of VVTL1 in fraction M3 is not in agreement with the absence of a band at 24 kDa on the silver-stained gel (Fig. 2, lane 3) and on a western-blot membrane, using a specific polyclonal antibody raised against TL protein (data not shown). Considering these experimental results and due to the high-sensitivity of the nanoLC-chip-MS/MS technique, which enables to detect very low abundant peptides, one can conclude that the pooled fraction M3 contains the highly purified grape vacuolar invertase.

The purification steps and protein yields are summarized in Table 2. From 41.6 L of Chardonnay must, around 12.2 mg of lyophilized grape invertase were obtained. Most interestingly, the amount of the pure invertase, produced by the one step purification method presented in this work, is at least ten times higher than quantities reported in previous studies as determined by the invertase activity assay or the Enzyme-Linked Immunosorbent Assay [6,8,11].

**Table 2**  
Purification of the grape vacuolar invertase from Chardonnay must.

Purification step	Volume or concentration <sup>a</sup>
Grape must	41.6 L
Ultrafiltration	381.5 mg L <sup>-1</sup>
Freeze-dried total colloids >10 kDa	
Q-Sepharose Fast Flow anion exchange chromatography	12.18 mg L <sup>-1</sup>
Freeze-dried fraction M3, grape vacuolar invertase	

<sup>a</sup> Concentrations were expressed as milligrams of lyophilized fractions per liter of must used for purification.

#### 4. Conclusion

In this work, large amounts of pure vacuolar grape invertase were purified from grape juice using a preparative chromatography procedure. Analysis of the purified grape invertase by SDS-PAGE and nanoLC-chip-MS/MS confirmed both the identity and the purity of the protein. Considering the high quantities of the grape invertase produced by the simple method developed here, physico-chemical and structural studies are now undertaken on the pure protein.

#### Acknowledgements

We express our gratitude to the Association Recherche Oenologie Champagne et Université (Reims, France) for financial support. We thank Pr F. Baillieul (University of Reims Champagne-Ardenne, France) for the generous gift of polyclonal antibodies raised against thaumatin-like protein. We gratefully thank the wine coop of Nogent l'Abbesse (France) for providing must samples.

#### References

- [1] R. Marchal, V. Seguin, A. Maujean, *Am. J. Enol. Vitic.* 48 (1997) 303.
- [2] E.J. Waters, N.J. Shirley, P.J. Williams, *J. Agric. Food Chem.* 44 (1996) 3.
- [3] E.J. Waters, G. Alexander, R. Muhlack, K.F. Pocock, C. Colby, B.N. O'Neill, P.B. Høj, P.R. Jones, *Aust. J. Grape Wine Res.* 11 (2005) 215.
- [4] T. Dambrouck, R. Marchal, L. Marchal-Delahaut, M. Parmentier, A. Maujean, P. Jeandet, *J. Agric. Food Chem.* 51 (2003) 2727.
- [5] C. Cilindre, A.J. Castro, C. Clément, P. Jeandet, R. Marchal, *Food Chem.* 103 (2007) 139.
- [6] T. Dambrouck, R. Marchal, C. Cilindre, M. Parmentier, P. Jeandet, *J. Agric. Food Chem.* 53 (2005) 8782.
- [7] C. Cilindre, S. Jégou, A. Hovasse, C. Schaeffer, A.J. Castro, C. Clément, A. Van Dorsseleer, P. Jeandet, R. Marchal, *J. Proteome Res.* 7 (2008) 1199.
- [8] K. Nakanishi, K. Yokotsuka, *J. Ferment. Bioeng.* 69 (1990) 16.
- [9] H.P. Ruffner, M. Hürlimann, R. Skrivan, *Plant Physiol. Biochem.* 33 (1995) 25.
- [10] R. Marchal, S. Bouquet, A. Maujean, *J. Agric. Food Chem.* 44 (1996) 1716.
- [11] W. Porntaveewat, T. Takayanagi, K. Yokotsuka, *J. Ferment. Bioeng.* 4 (1994) 288.
- [12] N. Puff, R. Marchal, V. Aguié-Béghin, R. Douillard, *Langmuir* 17 (2001) 2206.
- [13] K. Nakanishi, W. Wu, K. Yokotsuka, *J. Ferment. Bioeng.* 71 (1991) 66.
- [14] L.P. Dreier, J.J. Hunter, H.P. Ruffner, *Plant Physiol. Biochem.* 36 (1998) 865.
- [15] U.K. Laemmli, *Nature* 227 (1970) 680.
- [16] T. Rabilloud, L. Vuillard, C. Gilly, J.J. Lawrence, *Cell. Mol. Biol.* 40 (1994) 57.
- [17] C. Davies, S.P. Robinson, *Plant Physiol.* 111 (1996) 275.
- [18] E.J. Waters, Y. Hayasaka, D.B. Tattersall, K.S. Adams, P.J. Williams, *J. Agric. Food Chem.* 46 (1998) 4950.

## Chapitre 5

### Quantification de HbA<sub>2</sub> et étude des mutations de l'hémoglobine humaine

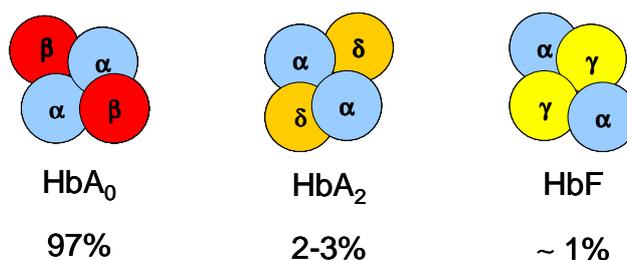
---

#### 1. Introduction

Les méthodologies de détermination de séquences de peptides et de protéines utilisées et développées dans les chapitres précédents se sont également révélées très utiles dans le travail présenté ci-dessous, qui porte sur des pathologies qui font intervenir des hémoglobines.

L'hémoglobine est le constituant essentiel des globules rouges et a pour fonction le transport de l'oxygène dans le sang. C'est un tétramère protéique dont la forme majoritaire est constituée de 2 sous-unités de globine  $\alpha$  et de 2 sous-unités de globine  $\beta$  ( $\alpha_2\beta_2$ ). Néanmoins, les globules rouges contiennent plusieurs formes d'hémoglobines correspondant à des séquences différentes des chaînes polypeptidiques qui sont nommées  $\alpha$ ,  $\beta$ ,  $\delta$ ,  $\gamma$ , etc. De part leur origine ancestrale commune, ces chaînes de globines ont des homologies de séquence très élevées (supérieures à 90% pour les chaînes de globines  $\delta$  et  $\beta$ ).

L'hémoglobine normale chez l'adulte est composée à 97% d'hémoglobine HbA<sub>0</sub>, de 2-3% d'HbA<sub>2</sub> et d'environ 1% d'HbF (Figure 1).



**Figure 1.** Composition d'un hémolysât d'une personne saine

Les chaînes de globines peuvent être affectées par des mutations. A ce jour, plus de 1000 formes d'hémoglobines anormales ont été décrites (HbVar: a Database of Human Hemoglobin Variants and Thalassemias. <http://globin.cse.psu.edu/hbvar/menu.html>; Accessed April 21, 2010). La plupart de ces variants n'ont pas de répercussion clinique [Wild et Bain 2004] mais certains sont responsables de signes cliniques sévères comme les thalassémies [Weatherall et Clegg 2001] ou encore la drépanocytose [Serjeant et Serjeant 2001]. Les variants S (HbS), C (HbC), et E (HbE) correspondent à des mutations de la chaîne beta et sont à l'origine des hémoglobinopathies les plus courantes [Clarke et Higgins 2000]. La mutation HbS est responsable de l'anémie falciforme appelée également drépanocytose. L'hémoglobinose C (HbC) homozygote ou associée à une beta-thalassémie, provoque des hémolyses, des anémies et une hypertrophie de la rate. L'hémoglobinose E (HbE) est une maladie très courante en Asie et offre une protection relative contre la malaria.

La plupart des variants sont détectés dans les laboratoires d'hématologie hospitaliers lors des analyses sanguines réalisées avec des méthodes de chromatographie (HPLC échangeuse de cations ou en phase inverse) ou d'électrophorèse et d'isoelectrofocalisation [Wajcman et al. 2001, Joutovsky et al. 2004]. Cependant, les identifications non ambiguës de certains variants nécessitent parfois une analyse d'ADN ou un séquençage de protéines complémentaires. Par ailleurs, ces méthodes d'analyse de routine ne permettent pas d'identifier de nouveaux variants.

Des méthodes basées sur la spectrométrie de masse ont émergé les 20 dernières années pour caractériser la nature des mutations et des modifications de l'hémoglobine provenant de patients qui présentaient des signes cliniques spécifiques [Wild et al. 2001].

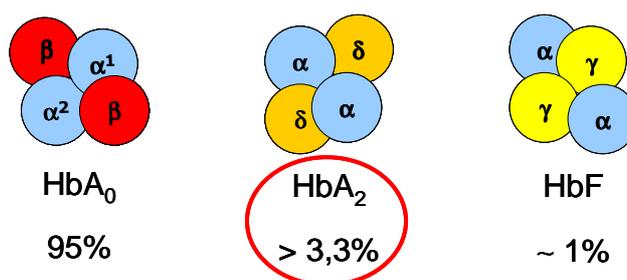
Dans ce chapitre, nous présentons trois études dans lesquelles la spectrométrie de masse joue un rôle central. La première porte sur une méthodologie permettant de quantifier l'hémoglobine A2 dans le but de certifier des standards de référence. Dans la seconde étude, nous avons mis au point, une stratégie permettant de caractériser de nouvelles mutations, qui ne pouvaient pas être identifiées par les méthodes classiques. La troisième étude est en fait une présentation des possibilités de la spectrométrie de masse pour doser HbA<sub>2</sub> et rechercher les mutations simultanément. En intégrant une analyse LC-MS/MS à la méthodologie développée pour la quantification de HbA<sub>2</sub>, nous avons montré qu'il était possible d'identifier les variants les plus courants (HbS, E, C et A2').

## 2. Quantification de HbA<sub>2</sub>

Cette étude a été réalisée dans le cadre d'un groupe de travail mandaté par l'International Federation of Clinical Chemistry and Laboratory Medicine (IFCC). Le coordinateur de ce groupe de travail est le Dr Andrea Mosca de l'Université de Milan, Italie.

### 2.1. Contexte de l'étude

L'hémoglobinopathie la plus grave est la beta-thalassémie, maladie génétique très fréquente caractérisée par une forte anémie liée à une réduction de la synthèse des chaînes  $\beta$  de globines [Weatherall et Clegg 2001]. Le déficit de production de la chaîne beta conduit à une augmentation du pourcentage en HbA<sub>2</sub> (Figure 2) ce qui constitue le marqueur de cette maladie. Le dosage de HbA<sub>2</sub> permet donc de diagnostiquer les syndromes thalassémiques.



**Figure 2.** Composition d'un hémolysât d'un patient atteints de beta-thalassémie

L'hémoglobine A<sub>2</sub> (HbA<sub>2</sub>) est composée de deux chaînes alpha et deux chaînes delta ( $\alpha_2\delta_2$ ) et est produite à un taux d'environ 2,5% à partir de la période néo-natale. La variation entre le taux normal (2-3%) et les valeurs pathologiques (>3,3%) est faible, ce qui implique de disposer de méthodes de dosage d'une grande précision pour établir le bon diagnostic.

Plusieurs méthodes de dosage de routine de HbA<sub>2</sub> existent comme l'électrophorèse ou l'HPLC [Clarke et Higgins 2000, Stephens 2005] mais aucun standard de référence n'était disponible sur le marché au début de ce projet. L'instrumentation médicale conçoit des appareils relativement peu coûteux qui peuvent doser HbA<sub>2</sub>, mais ces appareils ont besoin d'être étalonnés. C'est dans ce cadre, que l'IFCC a pour rôle d'aider au développement des méthodes analytiques qui permettent de garantir le contenu d'échantillons de référence. L'IFCC a approuvé la constitution d'un groupe de travail qui avait pour but de développer un système métrologique complet pour doser HbA<sub>2</sub> qui comprend : une méthode de référence, le matériel primaire (constitué de fractions pures de HbA<sub>0</sub> et HbA<sub>2</sub>) et le matériel secondaire que nous appellerons standards de référence ou calibrants par la suite. Au sein de ce groupe de laboratoires, nous avons cherché à développer une méthode permettant de garantir

le contenu de HbA<sub>2</sub> par rapport à Hb<sub>0</sub> avec un coefficient de variation proche de 5% et une précision de 0,3 %, pour des standards de référence allant de 1 à 10 % de HbA<sub>2</sub>.

## 2.2. Stratégie

L'approche envisagée est le dosage, par chromatographie liquide couplée à la spectrométrie de masse, des peptides issus de la digestion enzymatique par une endoprotéase de l'hémolysât total qui contient les chaînes de globines α, β et δ. Le dosage de HbA<sub>2</sub> est effectué en quantifiant la chaîne delta.

Le ratio d'un peptide spécifique de HbA<sub>2</sub> (chaîne delta) par rapport à un peptide spécifique de HbA<sub>0</sub> (chaîne alpha) permet de calculer le pourcentage de HbA<sub>2</sub> dans l'hémolysât total en se référant à une droite d'étalonnage obtenue à partir des peptides issus de la digestion enzymatique des standards de référence (calibrants). La stratégie globale est schématisée dans la Figure 3.

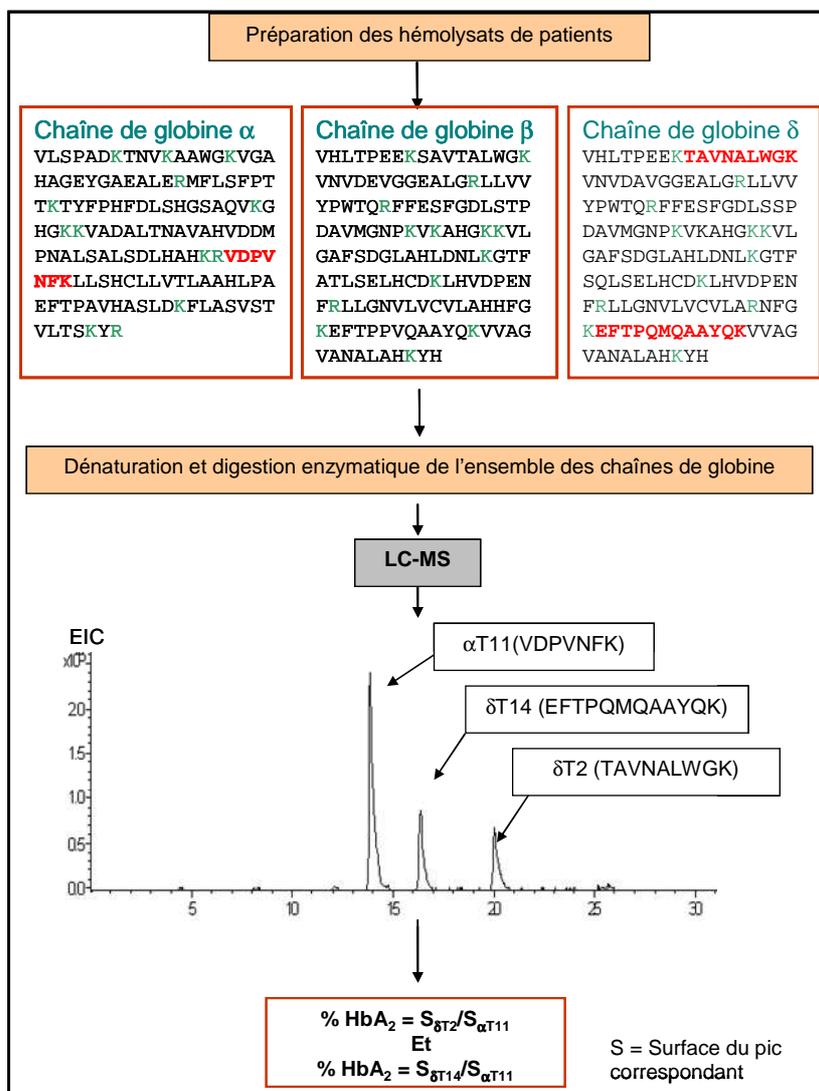


Figure 3. Stratégie globale de la méthode de dosage de HbA<sub>2</sub> dans le sang par LC-MS

Les optimisations pour le développement de cette stratégie ont porté sur plusieurs points :

- Le choix de l'enzyme pour générer des peptides spécifiques de chaque chaîne ( $\alpha$  et  $\delta$ );
- Le protocole permettant la digestion complète des chaînes de globines;
- Le choix des peptides utilisés pour la quantification;
- La séparation chromatographique des peptides de digestion;
- La quantification fine par spectrométrie de masse.

## 2.3. Développement méthodologique

Les détails des protocoles de préparation d'échantillon et d'analyse sont décrits dans le mode opératoire normalisé (SOP) situé en annexe 5.

### 2.3.1. Choix de l'endoprotéase

Notre stratégie est basée sur la quantification de peptides spécifiques de la chaîne delta et de la chaîne alpha par LC-MS. Il existe cependant une forte homologie de séquence entre les différentes chaînes de globines. L'endoprotéase utilisée doit donc permettre de générer des peptides spécifiques de chaque chaîne. De plus ces peptides doivent être parfaitement séparés chromatographiquement des autres peptides de digestion de l'hémolysât. En effet, la coélution avec d'autres espèces pourrait entraîner une suppression de signal, particulièrement dans le cas des peptides de la chaîne delta qui sont très minoritaires dans l'échantillon (0 à 10%).

Frag.	Res.	Sequence	Mmono
$\alpha$ T1	1-7	VLSPADK	728,41
$\alpha$ T3	12-16	AAWGK	531,28
$\alpha$ T4	17-31	VGAHAGEYGAEALER	1528,73
$\alpha$ T5	32-40	MFLSFPTTK	1070,55
$\alpha$ T6	41-56	TYFPHFDLSHGSAQVK	1832,88
$\alpha$ T11	93-99	VDPVNFK	817,43
$\alpha$ T13	128-139	FLASVSTVLTSK	1251,71
$\beta$ T2	9-17	SAVTALWGK	931,51
$\beta$ T3	18-30	VNVDEVGGEALGR	1313,66
$\beta$ T5	41-59	FFESFGDLSTPDAVMGNPK	2057,94
$\beta$ T10	83-95	GTFATLSELHCDK	1420,67
$\beta$ T12	105-120	LLGNLVLCVLAHFGK	1718,97
$\beta$ T13	121-132	EFTPPVQAAYQK	1377,69
$\delta$ T2	9-17	TAVNALWGK	958,52
$\delta$ T3	18-30	VNVDAVGGEALGR	1255,65
$\delta$ T5	41-59	FFESFGDLSSPDAVMGNPK	2043,92
$\delta$ T10	83-95	GTFSQLSELHCDK	1463,67
$\delta$ T12	105-116	LLGNLVLCVLAR	1268,76
$\delta$ T14	121-132	AFTPMQAAYQK	1440,67

**A** : Peptides spécifiques des chaînes  $\alpha$ ,  $\beta$  et  $\delta$  après digestion à la trypsine

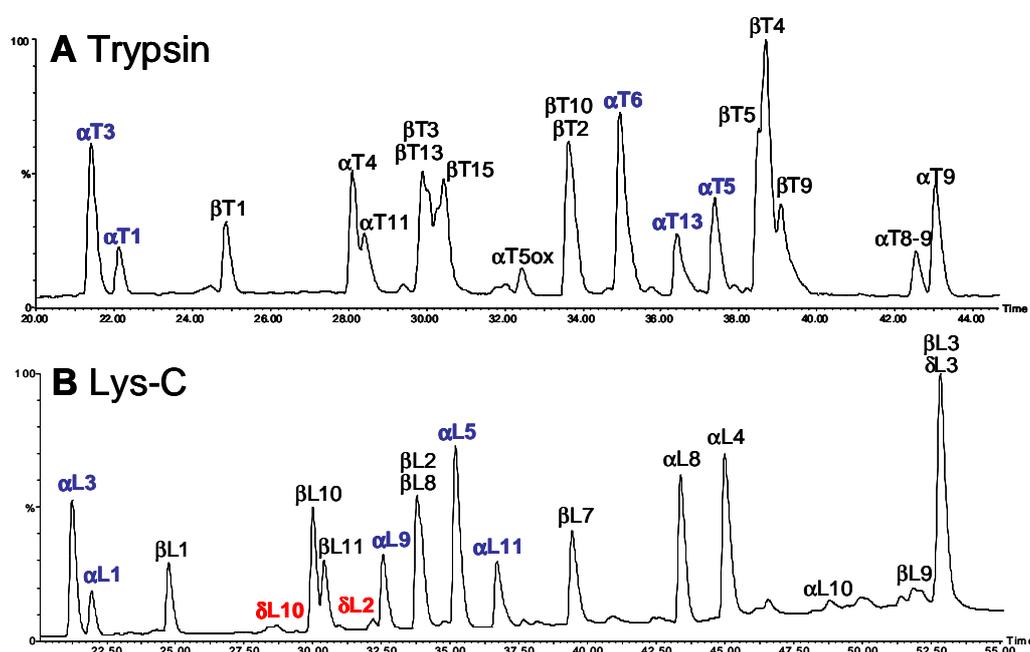
Frag.	Res.	Sequence	Mmono
$\alpha$ L1	1-7	VLSPADK	728.41
$\alpha$ L3	12-16	AAWGK	531.28
$\alpha$ L5	41-56	TYFPHFDLSHGSAQVK	1832.88
$\alpha$ L9	91-99	LRVDPVNFK	1086.62
$\alpha$ L11	128-139	FLASVSTVLTSK	1251.71
$\beta$ L2	9-17	SAVTALWGK	931.51
$\beta$ L8	83-95	GTFATLSELHCDK	1420.67
$\beta$ L10	121-132	EFTPPVQAAYQK	1377.69
$\delta$ L2	9-17	TAVNALWGK	958.52
$\delta$ L8	83-95	GTFSQLSELHCDK	1463.67
$\delta$ L10	121-132	EFTPMQAAYQK	1440.67

**B** : Peptides spécifiques des chaînes  $\alpha$ ,  $\beta$  et  $\delta$  après digestion à la Lys C

**Tableau 1.** Peptides spécifiques des chaînes de globines  $\alpha$ ,  $\beta$  et  $\delta$  après digestion *in silico* avec de la trypsine (A) et de la Lys C (B)

Deux enzymes ont été testées, la trypsine et la Lys C. Le Tableau 1 présente les peptides spécifiques de chaque chaîne de globine obtenus par digestion *in silico* avec les deux endoprotéases.

Un hémolysat total est digéré avec chacune des enzymes puis les digestats sont analysés par LC-MS (Figure 4). Le profil chromatographique obtenu pour la digestion à la Lys C (Figure 4B) montre clairement une meilleure résolution des pics que celle obtenue avec la trypsine (Figure 4A) dans les mêmes conditions d'analyse. Ceci peut s'expliquer par le fait que la digestion trypsique génère un plus grand nombre de peptides (Tableau 1). De plus, deux pics correspondant à des peptides spécifiques de la chaîne delta ( $\delta$ L10 et  $\delta$ L2) sont observés dans le digestat Lys C (Figure 4B) contrairement au profil obtenu avec la digestion à la trypsine.



**Figure 4.** LC-MS d'un digestat d'hémolysat total à la trypsine (A) et à la Lys C (B) avec une colonne Vydac, C18, 2 x 250mm, 5 $\mu$ m et en utilisant les mêmes conditions chromatographiques

Ces résultats semblent indiquer que la Lys C est plus adaptée que la trypsine pour générer des peptides spécifiques des 2 chaînes alpha et delta séparés par chromatographie liquide. Cependant, le but de cette étude était de développer une méthode de référence, transposable sur différents spectromètres de masse, robuste, peu coûteuse et utilisable sur du long terme. La Lys C a été considérée comme trop "exotique", ne remplissant pas les critères de coût et de longévité des réactifs utilisés. Pour ces raisons, la trypsine a donc été choisie pour cette étude.

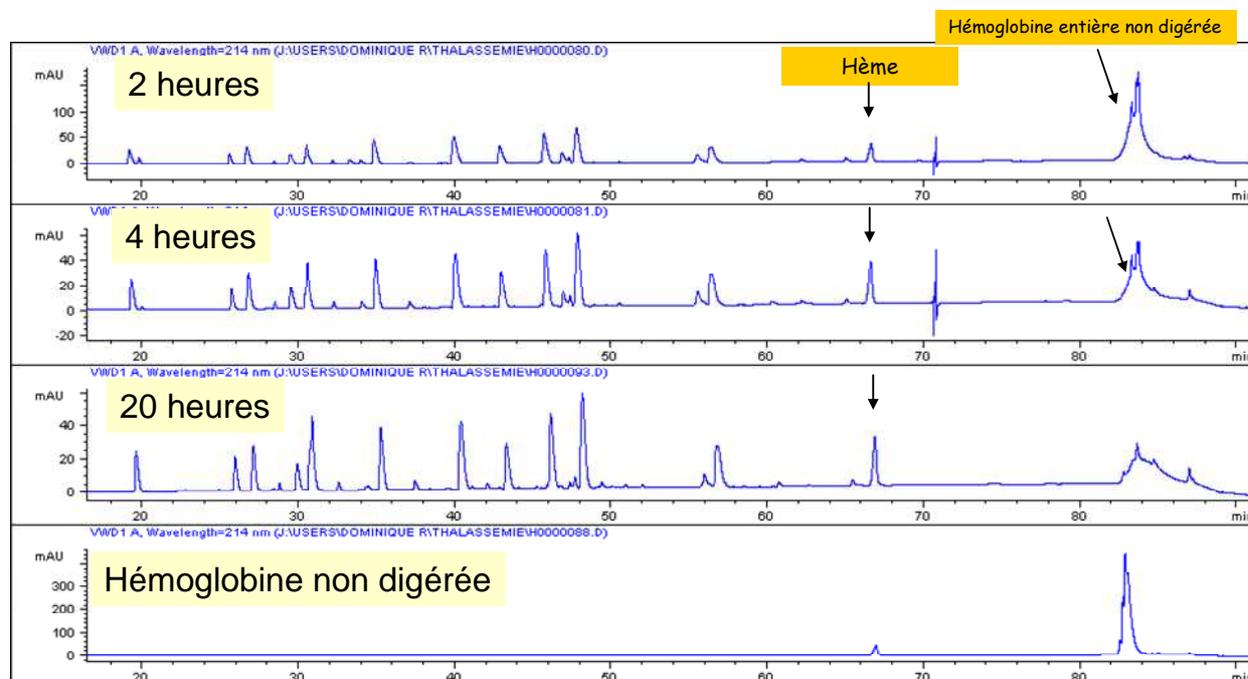
### 2.3.2. Protocole de digestion

Cette méthode de quantification de l'hémoglobine A2 est basée sur le clivage reproductible des chaînes de globine présentes en mélange dans l'hémolysat de départ. Il est donc primordial que

la digestion enzymatique soit complète pour que la concentration des peptides spécifiques soit constante sur plusieurs digestions d'un même échantillon.

Les chaînes de globines ne possèdent pas de pont disulfures, il n'est donc pas nécessaire de les réduire et de les alkyler. Néanmoins pour une meilleure accessibilité des sites de clivages et pour que la digestion soit reproductible, les protéines doivent être parfaitement dénaturées c'est pourquoi nous avons utilisé un milieu organique avec du TFE (TriFluoroEthanol). Les optimisations ont ensuite porté sur la durée et la température de dénaturation, le ratio enzyme/substrat et le temps de digestion. Le chromatogramme obtenu après 20 heures de digestion montre que le pic d'hémoglobine disparaît totalement (Figure 5).

Les conditions de digestion optimisées sont donc les suivantes : 1 heure de dénaturation à température ambiante en milieu organique (TFE), suivie par 20 heures de digestion à 37°C, avec un rapport enzyme/substrat de 1/20.



**Figure 5.** Cinétique de digestion à la trypsine du mélange de chaînes d'hémoglobine HbA<sub>0</sub> et HbA<sub>2</sub> par LC-UV

### 2.3.3. Choix des peptides spécifiques pour la quantification

Le choix des peptides impliqués dans la quantification de HbA<sub>2</sub> est basé sur les critères suivants :

- 2 ou 3 peptides spécifiques de la chaîne delta (HbA<sub>2</sub>);
- 1 peptide spécifique de la chaîne alpha (HbA<sub>0</sub>);
- ne pas être affecté par les mutations les plus courantes;
- avoir une bonne résolution chromatographique;
- être détectable dans une gamme de masse de 250 à 2500 m/z;

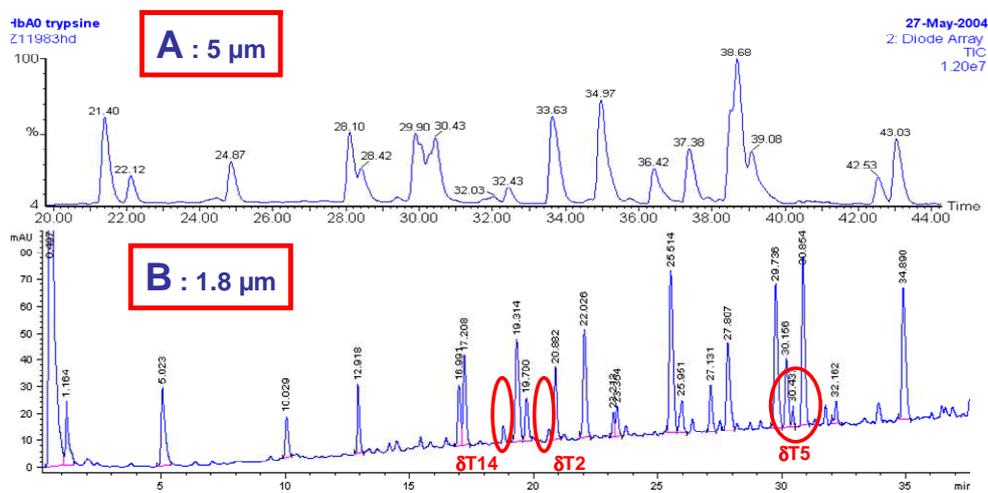
- avoir des facteurs de réponse proches en MS;
- éluer dans une zone d'élution restreinte pour limiter la durée de la séparation chromatographique.

Les peptides répondant à ces critères choisis pour la quantification de HbA<sub>2</sub> sont donc  $\delta T_2$ ,  $\delta T_{14}$  et  $\alpha T_{11}$  (Tableau 1).

### 2.3.4. Séparation chromatographique

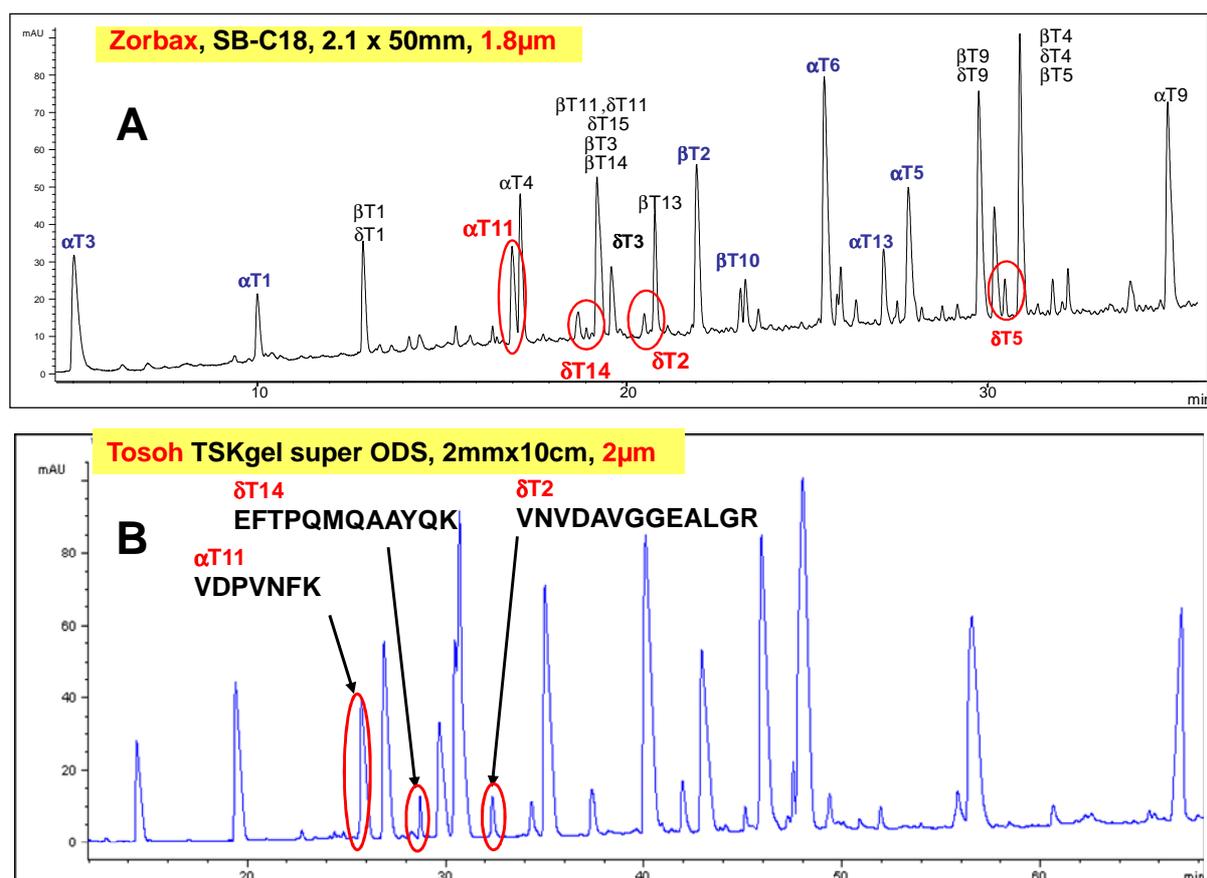
La digestion enzymatique du mélange des différentes chaînes de globine génère un grand nombre de peptides (une quinzaine par chaîne de globine). Les peptides de la chaîne delta sont minoritaires par rapport aux peptides des chaînes alpha et beta. Il est donc primordial de disposer de la meilleure résolution chromatographique possible pour séparer les peptides et éviter les suppressions de signal.

L'optimisation des conditions chromatographiques a porté sur le gradient et le débit mais aussi sur le type de colonne. Tout d'abord, l'utilisation de colonnes ayant des particules de plus petit diamètre (< 2  $\mu m$ ) a permis d'améliorer nettement la résolution et la séparation chromatographique (Figure 6). Ces colonnes sont compatibles avec les systèmes HPLC classiques et ne nécessitent pas l'utilisation d'un équipement UHPLC.



**Figure 6.** Séparation chromatographique (LC-UV) du mélange peptidique (obtenue par digestion enzymatique d'un mélange HbA<sub>0</sub> et HbA<sub>2</sub>) avec des colonnes Vydac, C18, 2 x 250 mm, 5  $\mu m$  (A) et Zorbax, SB-C18, 2.1 x 50 mm, 1.8 $\mu m$  (B)

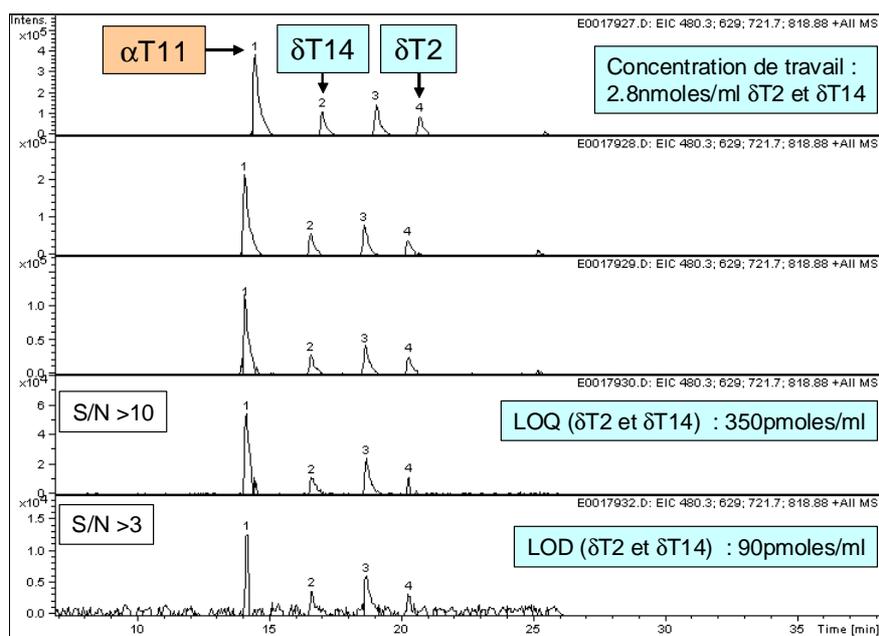
Ensuite, deux colonnes à petites particules ont été comparées (Zorbax vs TSK-Gel) Ces colonnes possèdent une sélectivité différente, et la colonne TSK-Gel (Tosoh) permet d'obtenir une meilleure séparation des peptides sélectionnés pour la quantification de HbA<sub>2</sub> et a donc été choisie pour la suite de l'étude (Figure 7).



**Figure 7.** Séparation chromatographique (LC-UV) du mélange peptidique (obtenue par digestion enzymatique d'un mélange HbA<sub>0</sub> et HbA<sub>2</sub>) avec des colonnes Zorbax, SB-C18, 2.1 x 50 mm, 1.8 µm (A) et Tosoh, TSKgel super ODS, 2 mm x 10 cm, 2 µm (B)

### 2.3.5. Limite de quantification

Afin de déterminer la limite de quantification (LOQ) de cette méthode, nous avons utilisé le standard de référence (calibrant) contenant la quantité de HbA<sub>2</sub> la plus faible, soit 1,5%, ce qui correspond à une concentration en chaîne delta de 2,8 nmoles/ml. Par dilutions successives du digestat de ce calibrant, la LOQ (pour un rapport signal sur bruit de 10) des peptides d'intérêt a pu être déterminée, et est à 350 pmoles/ml (Figure 8). La concentration de travail est donc 8 fois supérieure à la LOQ.



**Concentration de travail: 8 fois > à LOQ**

**Figure 8.** Chromatogrammes EIC (Extracted Ion Chromatogram) des peptides d'intérêt pour des concentrations décroissantes pour déterminer la LOQ (limite de quantification) et la LOD (limite de détection)

### 2.3.6. Etalonnage par LC-MS

Une droite d'étalonnage est établie en analysant les digestats des différents calibrants ayant chacun des concentrations différentes. Chaque point de la droite est déterminé par le rapport de l'aire du pic correspondant au peptide de la chaîne delta ( $\delta T2$  ou  $\delta T14$ ) sur l'aire du pic correspondant au peptide de la chaîne alpha ( $\alpha T11$ ).

Nous avons vérifié qu'il n'y avait aucune interférence de peptides trypsiques ayant à la fois des  $m/z$  et des temps de rétention identiques à ceux de  $\delta T2$  et  $\delta T14$ . De plus, pour des concentrations allant de 0% à 7,5% de HbA<sub>2</sub>, le rapport des aires est parfaitement linéaire (pour  $\delta T2$ ,  $R^2 = 0,9988$ ; pour  $\delta T14$ ,  $R^2 = 0,9969$ ).

## 2.4. Résultats et discussion

Afin d'évaluer la méthode, 20 échantillons (2007-01 à 20) ont été analysés en utilisant des calibrants à 6 concentrations différentes (Tableau 2A). Ces concentrations ont été choisies pour encadrer les pourcentages mesurés dans les échantillons biologiques. De plus, des échantillons "contrôles" avec des concentrations en HbA<sub>2</sub> connues, déterminées par HPLC (Tableau 2B) ont été analysés pour vérifier la justesse de la méthode. Le Tableau 2B indique les valeurs mesurées pour ces contrôles (valeur vraie). Pour que la concentration expérimentale soit considérée comme juste, elle doit être comprise dans la "gamme acceptable".

Deux digestions indépendantes ont été réalisées sur chaque échantillon et chaque digestat est analysé en duplicat, soit quatre mesures par échantillon. Les détails de la séquence d'analyse sont donnés dans le mode opératoire normalisé (Annexe 5, SOP version A2).

<b>A</b> Calibrants		<b>B</b> Echantillons contrôle			
Noms	% HbA <sub>2</sub>	% HbA <sub>2</sub>			
CAL G07-1	0.0	Nom	Valeur Vrai	Gamme Acceptable	Moyenne Expérimentale
CAL G07-2	1.5	2007 C1	2.6	2.4 – 2.8	2,9
CAL G07-3	3.0	2007 C2	4.2	4.0 – 4.5	4,8
CAL G07-4	4.5	2007 C3	6.2	5.8 – 6.6	6,6
CAL G07-5	6.0				
CAL G07-6	7.5				

**Tableau 2.** Description des calibrants (A) et des contrôles (B)

Le Tableau 3 indique les équations des droites d'étalonnage et les coefficients de détermination ( $R^2$ ) pour chaque série de calibrants analysés pour  $\delta T2/\alpha T11$  et  $\delta T14/\alpha T11$ . Ces données permettent de déterminer la linéarité et donc de montrer que les ratios  $\delta T2/\alpha T11$  et  $\delta T14/\alpha T11$  sont proportionnels à leurs concentrations à l'intérieur d'une gamme de concentration donnée (0% à 7,5%).

On obtient des coefficients de détermination ( $R^2$ ) supérieurs à 0,99. Dans cette étude, les ratios  $\delta T2/\alpha T11$  et  $\delta T14/\alpha T11$  sont donc proportionnels aux concentrations en HbA<sub>2</sub>.

$\delta T2/\alpha T11$		$\delta T14/\alpha T11$	
Equation de la droite d'étalonnage	$R^2$	Equation de la droite d'étalonnage	$R^2$
$y = 0,0391x + 0,0083$	0,9959	$y = 0,0522x + 0,0093$	0,9978
$y = 0,0376x + 0,0066$	0,9972	$y = 0,0472x + 0,0071$	0,9987
$y = 0,044x + 0,0004$	0,9992	$y = 0,0495x + 0,0046$	0,9900
$y = 0,0405x - 0,0081$	0,9845	$y = 0,0433x + 0,0062$	0,9973
$y = 0,0383x + 0,0053$	0,9947	$y = 0,0584x - 0,0037$	0,9967
$y = 0,0361x + 0,0013$	0,9977	$y = 0,0502x - 0,0088$	0,9766
$y = 0,0412x - 0,0077$	0,9903	$y = 0,0527x - 0,0082$	0,9914
$y = 0,0351x - 0,0017$	0,9876	$y = 0,0421x - 0,0024$	0,9859
<b>Moyenne</b>	<b>0,9934</b>	<b>Moyenne</b>	<b>0,9918</b>

**Tableau 3.** Equations des droites d'étalonnage et coefficients de détermination pour chaque série de calibrants analysée

Afin de tester la justesse de la méthode, des échantillons contrôles (concentration en HbA<sub>2</sub> connue) ont été analysés. Les résultats sont décrits dans le Tableau 2B ("moyenne expérimentale"). Les moyennes des concentrations obtenues expérimentalement sont légèrement supérieures à la gamme acceptable.

Peptides: $\delta T2/\alpha T11$	Hb A <sub>2</sub> , %									
	Digestion 1				Digestion 2				TOTAL	
	Echantillons	Analyse 1	Analyse 2	moyenne	CV	Analyse 1	Analyse 2	moyenne	CV	moyenne
2007-1	3,10	3,19	3,15	1,5%	3,31	3,42	3,36	1,6%	3,26	3,6%
2007-2	3,13	3,36	3,25	3,6%	2,63	3,04	2,83	7,1%	3,04	8,7%
2007-3	6,52	6,39	6,45	1,0%	5,52	5,52	5,52	0,0%	5,99	7,8%
2007-4	5,84	6,03	5,94	1,6%	5,50	5,49	5,50	0,1%	5,72	4,0%
2007-5	3,36	3,50	3,43	2,1%	2,71	2,94	2,83	4,2%	3,13	10,2%
2007-6	6,20	6,73	6,46	4,1%	6,10	6,40	6,25	2,4%	6,36	3,8%
2007-7	2,77	2,45	2,61	6,2%	2,13	2,40	2,27	5,8%	2,44	9,3%
2007-8	3,45	3,66	3,56	2,9%	3,29	3,28	3,28	0,0%	3,42	4,5%
2007-9	6,01	6,05	6,03	0,3%	5,23	5,51	5,37	2,6%	5,70	6,0%
2007-10	8,05	7,29	7,67	5,0%	7,02	6,96	6,99	0,4%	7,33	5,9%
2007-11	5,35	5,51	5,43	1,5%	4,97	5,27	5,12	2,9%	5,27	3,7%
2007-12	6,97	6,69	6,83	2,1%	6,24	6,91	6,58	5,0%	6,70	4,2%
2007-13	3,46	3,53	3,49	1,1%	3,75	3,60	3,67	2,0%	3,58	3,0%
2007-14	3,29	3,32	3,31	0,3%	3,04	3,15	3,10	1,8%	3,20	3,4%
2007-15	7,58	7,63	7,61	0,3%	7,00	7,25	7,13	1,8%	7,37	3,5%
2007-16	3,47	3,10	3,28	5,6%	3,08	3,25	3,17	2,6%	3,22	4,8%
2007-17	5,95	6,02	5,98	0,6%	5,74	6,05	5,89	2,6%	5,94	2,0%
2007-18	3,31	3,10	3,20	3,3%	3,01	3,35	3,18	5,3%	3,19	4,4%
2007-19	6,43	6,02	6,22	3,3%	6,29	6,69	6,49	3,1%	6,36	3,8%
2007-20	2,90	1,83	2,36	22,7%	3,29	3,36	3,33	1,1%	2,85	21,6%
			<b>moyenne CV</b>	<b>3,5%</b>			<b>moyenne CV</b>	<b>2,6%</b>	<b>moyenne CV</b>	<b>5,9%</b>

**Tableau 4.** Résultats obtenus en utilisant le peptide  $\delta T2$ , en rouge les coefficients de variation supérieurs à 10% et en bleu ceux compris entre 5 et 10%

Peptides: $\delta T14/\alpha T11$	Hb A <sub>2</sub> , %									
	Digestion 1				Digestion 2				TOTAL	
	Echantillons	Analyse 1	Analyse 2	moyenne	CV	Analyse 1	Analyse 2	moyenne	CV	moyenne
2007-1	3,06	3,24	3,15	2,8%	3,19	3,28	3,24	1,3%	3,19	2,6%
2007-2	3,20	3,41	3,31	3,2%	2,73	3,15	2,94	7,3%	3,12	8,0%
2007-3	5,95	6,49	6,22	4,3%	5,39	5,71	5,55	2,9%	5,89	6,8%
2007-4	6,58	6,39	6,49	1,4%	5,43	5,73	5,58	2,7%	6,03	7,8%
2007-5	3,16	3,55	3,36	5,7%	2,87	3,04	2,95	2,8%	3,16	7,9%
2007-6	6,50	6,68	6,59	1,3%	6,06	6,89	6,47	6,4%	6,53	4,7%
2007-7	2,59	2,41	2,50	3,5%	2,30	2,49	2,39	4,1%	2,45	4,4%
2007-8	3,56	3,35	3,46	3,0%	3,35	3,32	3,33	0,6%	3,40	2,8%
2007-9	6,29	5,61	5,95	5,7%	5,32	5,88	5,60	5,0%	5,77	6,2%
2007-10	7,43	7,13	7,28	2,1%	7,00	7,22	7,11	1,5%	7,19	2,2%
2007-11	5,30	5,86	5,58	5,0%	5,15	5,69	5,42	5,0%	5,50	5,2%
2007-12	6,64	7,11	6,88	3,4%	6,46	7,15	6,80	5,1%	6,84	4,3%
2007-13	3,75	3,74	3,75	0,3%	3,83	3,93	3,88	1,3%	3,81	2,0%
2007-14	3,65	3,61	3,63	0,5%	3,22	3,50	3,36	4,1%	3,50	4,8%
2007-15	7,71	8,22	7,97	3,2%	6,50	6,90	6,70	3,0%	7,33	9,2%
2007-16	3,51	3,26	3,39	3,6%	3,17	3,24	3,20	1,1%	3,29	3,9%
2007-17	6,44	6,21	6,33	1,8%	5,74	6,65	6,20	7,3%	6,26	5,4%
2007-18	3,36	3,27	3,31	1,3%	3,22	3,26	3,24	0,5%	3,28	1,5%
2007-19	6,94	6,05	6,49	6,9%	6,34	6,93	6,64	4,4%	6,57	5,9%
2007-20	2,84	3,99	3,41	16,9%	3,20	3,33	3,27	1,9%	3,34	12,4%
			<b>moyenne CV</b>	<b>3,8%</b>			<b>moyenne CV</b>	<b>3,4%</b>	<b>moyenne CV</b>	<b>5,4%</b>

**Tableau 5.** Résultats obtenus en utilisant le peptide  $\delta T14$ , en rouge les coefficients de variation supérieurs à 10% et en bleu ceux compris entre 5 et 10%

Les résultats obtenus sont décrits dans le Tableau 4 pour le peptide  $\delta$ T2 et dans le Tableau 5 pour le peptide  $\delta$ T14. Les colonnes en jaune correspondent au pourcentage de HbA<sub>2</sub> calculé pour chaque analyse. Dans les colonnes "CV" sont indiqués les coefficients de variations calculés pour chacune des digestions puis pour l'ensemble des analyses. Les CV supérieurs à 10% sont indiqués en rouge, ceux compris entre 5 et 10% sont en bleu et ceux inférieurs à 5% sont en noir. L'échantillon 2007-20 est le seul pour lequel on obtient un coefficient de variation supérieur à 10%. De plus, on observe que les CV obtenus avec le peptide  $\delta$ T14 sont plus élevés qu'avec le peptide  $\delta$ T2. Le peptide  $\delta$ T14 possède une méthionine dans sa séquence qui peut subir une oxydation partielle, ce qui pourrait expliquer le fait que des CV plus élevés soient obtenus avec ce peptide. De plus, les coefficients de variation au sein d'une même digestion sont globalement meilleurs qu'entre 2 digestions d'un même échantillon. Ces résultats montrent donc que la reproductibilité de la digestion enzymatique est un point critique. Néanmoins, la majorité des coefficients de variation calculés sont inférieurs ou proches de 5%.

## 2.5. Conclusion et perspectives

Le but de ce travail était de développer une méthode d'analyse permettant de garantir le contenu de HbA<sub>2</sub> par rapport à HbA<sub>0</sub> avec un coefficient de variation inférieur à 5% et avec un intervalle de confiance de  $\pm 0,3$  %, pour des solutions allant de 1 à 8 %. Ces critères de quantification sont bien plus stricts que ce qui est habituellement demandé dans le dosage de protéines.

Nous avons développé une série d'approches basées sur une digestion du mélange protéique suivi d'une quantification par LC-MS des peptides spécifiques de chacune des hémoglobines à doser.

Nous sommes parvenus à obtenir des CV proches de 5%, mais la méthode doit encore être améliorée. Nous allons concentrer nos efforts sur : la reproductibilité de la digestion, l'amélioration de la sélectivité, l'utilisation de peptides synthétiques alpha T11 et delta T2 marqués au C13 comme standards internes, l'utilisation de LC-MRM.

Cette première phase du projet a donné lieu à un mode opératoire normalisé (SOP) (annexe 5) qui sera modifié par la suite en fonction des dernières améliorations réalisées. La méthode sera ensuite évaluée par des tests inter-laboratoires sur des séries d'échantillons avant d'être validée par l'IFCC.

Les retombées de ce travail de quantification de l'hémoglobine A2 se situent au niveau du savoir faire pour l'analyse complète des digestats d'hémoglobines avec la possibilité d'identifier les mutations les plus courantes par LC-MS/MS.



# **ANNEXE 5: mode opératoire normalisé (SOP)**



**Candidate Reference Measurement Procedure for HbA<sub>2</sub>  
based on Peptide Mapping**

**STANDARD OPERATING PROCEDURE**

Version No. 2.2

Valid from: June 8, 2007

## Index

### **Part 1 Overview**

- 1.1 Principle
- 1.2 Materials and equipment
  - 1.2.1 Patient samples
  - 1.2.2 Primary calibrators
  - 1.2.3 Chemicals
  - 1.2.4 Equipment
  - 1.2.5 Analytical instrumentation

### **Part 2 Sample preparation**

- 2.1 Patient samples
- 2.2 Primary calibrators
- 2.3 Enzymatic digestion
  - 2.3.1 Solvents and reagents
  - 2.3.2 Digestion procedures

### **Part 3 Measurement procedures**

- 3.1 HPLC conditions
- 3.2 MS conditions
- 3.3 Sample analysis
- 3.4 Analysis sequence
- 3.5 Calculation of raw data and calibration

### **Part 4 Glossary**

### **Part 5 Appendix**

- 5.1 Example of Primary Calibrators data Sheet
- 5.2 HPLC-MS chromatograms
- 5.3 Mass spectra
- 5.4 Typical calibration curve
- 5.5 Total hemoglobin determination

## **Part 1**                      **Overview**

### 1.1 Principle

HbA<sub>2</sub> ratio to whole hemoglobin is determined as ratio of a delta chain specific peptide to an alpha chain specific peptide. Peptides are obtained by treating total red blood cell lysate with trypsin. Peptide mixture is analyzed by HPLC-ESI/MS. Calibration is performed against primary calibrators made of mixtures of HbA<sub>2</sub> and HbA<sub>0</sub> primary reference materials.

### 1.2 Materials and equipment

#### 1.2.1 *Patient samples*

Details are described in Part 2.

#### 1.2.2 *Primary calibrators*

Details are described in Appendix 5.1.

#### 1.2.3 *Chemicals*

Acetonitrile, carbon tetrachloride, formic acid, chloridic acid (HCl), Na<sub>2</sub>EDTA, potassium cyanide (KCN), potassium dihydrogen phosphate (KH<sub>2</sub>PO<sub>4</sub>), potassium ferricyanide, sodium chloride (NaCl), sucrose, trifluoroacetic acid (TFA), trifluoroethanol (TFE), triton X 100, TRIS-HCl, deionized water (MilliQ, 18 MOhm).

All chemicals analytical grade.

Trypsin sequencing grade (Promega, Cat. No. V511A)

Standards for ESI-MS calibration and tuning:

- Polypeptide Met-Arg-Phe-Ala (MRFA, Finnigan MAT, Bremen, part No. 40061-60002)
- ESI-MS Tune Mix (Agilent Technologies, part N°G1969-85000).

#### 1.2.4 *Equipment*

- Pipettes calibrated to handle microliter volumes.
- Incubator at 39 °C (±1.0 °C).
- Centrifuge (3000 g at 8 °C).
- Eppendorf centrifuge (12000 rpm at room temperature).
- 500 µL white polypropylene Eppendorf vials.

- Screw capped vials with 200 µL glass or 100 µL polypropylene insert.

### 1.2.5 Analytical instrumentation

#### A) HPLC-ESI-MS

- Analytical HPLC system with 214 nm photometric detector, cell volume 1.0 µL, suitable for microbore columns with flow rates of 0.5 mL/min, slow gradients for water-TFA/acetonitrile-TFA and switching valve between HPLC column and detector. Autosampler, autoinjector (temperature controlled equipment in the range 6-10 °C is preferable) and column oven are necessary. Analytical HPLC column TSKgel super ODS, 2µ; 2 x 100 mm (in Europe: Tosoh Bioscience GmbH, Zettachring 6, 70567 Stuttgart, Germany, Tel +49 711 13257 89; part No. P/N 19542; in USA: Tosoh Bioscience LLC, 156 Keystone Drive, Montgomeryville, PA 18936, USA, Tel +1 800 366 4875).
- Electrospray-Ionisation-Mass spectrometer coupled to microbore HPLC system (HPLC-ESI-MS). Mass spectrometer HCT Bruker (ion trap) and Agilent Technologies HPLC system 1100 series or LTQ ThermoElectron (linear ion trap) and Surveyor LC Pump Plus ThermoElectron.

#### B) Data processing

Chromatography data systems capable of precise automatic and/or manual integration.

## **Part 2**                    **Sample preparation**

### 2.1 Patient samples

Fresh blood samples from patients are collected in EDTA, stored at +4 °C and used within 3 days. Blood (1-2 mL) is centrifuged (1000 g, 10 min, +4 °C) and plasma removed. The erythrocytes are washed 3 times with isotonic saline (9 g/L NaCl) and lysed by addition of distilled water and carbon tetrachloride (1.0 and 0.4 volumes per volume of erythrocytes, respectively). After centrifugation (3000 g, 20 min, +4 °C) the hemoglobin solution forming the supernatant is collected.

Total hemoglobin concentration is measured (as described in Appendix 5.5) and the hemolysates are diluted to 30 g/L total hemoglobin with the following buffer: 20 mmol/L TRIS, 1.0 mmol/L KCN, 0.5 mmol/L Na<sub>2</sub>EDTA, adjusted to pH 7.5 with 1 mol/L HCl. Solid sucrose is added to the hemolysates to give a final 50 g/L concentration.

Hemolysates (100 µL each) are then stored at -80 °C in screw capped vials with 200 µL volume glass insert.

### 2.2 Primary calibrators

Different setpoint HbA<sub>2</sub> between 0.0 % and 7.5 %.

Total Hb concentration is 30 g/L.

Details are described in package insert of each lot (Appendix 5.1).

### 2.3 Enzymatic digestion

#### 2.3.1 *Solvents and reagents*

##### A) TRIS buffer

50 mmol/L TRIS-HCl adjusted to pH 8.5 with 1 mol/L NaOH.

The buffer is dispensed in 1000 µL aliquots in polypropylene Eppendorf vials and stored at -20 °C. Use one aliquot per day, don't reuse the same aliquot for another experiment.

##### B) Enzyme solution

Carefully open the vial with trypsin in order to avoid loss of enzyme.

Reconstitute 100 µg trypsin vial with 135 µL of TRIS buffer. In case of reconstituting more than one vial then the contents have to be pooled. It is advisable to perform the experiment (i.e. the digestion of calibrators and samples with trypsin vials of the same batch).

### 2.3.2 Digestion procedure

- It is advisable to perform two digestions for each sample and calibrator.
- Transfer 5 µL of sample/calibrator to 500 µL white polypropylene Eppendorf vial.
- Add 5 µL of Trifluoroethanol and vortex as soon for 2-5 s. Let the sample at room temperature for 1 hour.
- To the denaturated sample add, in this order, 30 µL of TRIS buffer then 10 µL of trypsin solution. The ratio enzyme/substrate should be 1/20 (w/w).
- Vortex a few seconds and let digest for 2 hours at 39 °C.
- After 2 hours of digestion, vortex again and let continue the digestion overnight (total digestion time should be 20 ± 1 hours).
- To stop the digestion, take out the Eppendorf vial from the incubator and vortex. Add 1 µL of 10 % formic acid (v/v) and vortex again.
- Store at -80 °C until analysis.

## **Part 3**                      **Measurement procedures**

### 3.1 HPLC conditions

Column: Tosoh TSKgel super ODS 2 $\mu$ m, 2 x 100 mm.

Column temperature: 40 °C.

Maximum column pressure: 230 bar.

Column switching valve after the column, in order to send the eluate in the mass spectrometer only for analysis.

Flow rate: 0.55 mL/min.

Eluent A: 0.05 % TFA/water.

Eluent B: 0.05 % TFA/acetonitrile.

Eluent degaser.

Detector (optional): UV 214 nm (1  $\mu$ L cell volume) on line with ESI source.

Sample temperature: room temperature for HCT Bruker, 8 °C for LTQ ThermoElectron.

Injection volume: 5-10  $\mu$ L (see 3.3).

Elution gradient:

Time, min:	0	5	20	30	32	34	35	42
B %:	5	8	16	35	90	90	5	5

Before 13 and after 23 min, the HPLC flow is switched out of the ion source. Expected retention time of peptides are between 14 - 22 min.

### 3.2 MS conditions

Two different instrumentations and settings have been used so far. These are listed separately as follows:

- a) Mass spectrometer: HCT Bruker  
Ionization mode: Electrospray positive  
Capillary potential: 4000 V  
Dry gas flow: 10 L/min  
Dry gas temperature: 250 °C  
Nebulizer: 20 psi  
Capillary exit: 140 V  
Skimmer: 40 V  
MS scan range: 450-1500 m/z  
Scan mode: standard enhanced (8100 m/z per sec)

- b) Mass spectrometer: LTQ ThermoElectron  
Ionisation mode: Electrospray positive  
Capillary potential: 4500 V  
Sheath gas flow: 70 U  
Auxiliary gas flow: 25 U  
Capillary temperature: 250 °C  
Capillary voltage: 46 V  
Tube lens: 105 V  
MS scan range: 450-1500 m/z  
Scan mode: normal full scan

Settings for ESI-MS are only indicative; exact settings depend on actual system performance and have to be optimized with the appropriate standards for ESI-MS calibration and tuning (point 1.2.3). The ion source and analyzer are set in order to monitor the most significant ion (state of charge) for each peptide. Using the described experimental conditions, the ions at  $m/z$  481, 721 and 818 for  $\delta$ T2,  $\delta$ T14 and  $\alpha$ T11 respectively, are selected (Fig. 2).

### 3.3 Sample analysis

- Immediately before analysis the frozen samples/calibrators are thawed at room temperature. Samples and calibrators must be taken from the same set of digest.
- For HCT Bruker analysis: vortex the samples at room temperature and add 200  $\mu$ L of 0.1 % TFA/water to make a 1:5 dilution.
- For LTQ ThermoElectron analysis: vortex the samples at room temperature and add 450  $\mu$ L of 0.05 % TFA/water to make a 1:10 dilution.
- Before the injection, centrifuge each sample 10 min at 12000 rpm (Eppendorf centrifuge).
- Injection volume: 5  $\mu$ L for HCT Bruker analysis or 10  $\mu$ L for LTQ ThermoElectron analysis.
- At least two injections/vial may be needed depending on the stability of ion source.
- Record and report the signals of the peptides  $\delta$ T2,  $\delta$ T14 and  $\alpha$ T11. The detection program has to be adapted to the instrument used.

### 3.4 Analysis sequence

In a typical experiment a maximum of 6 calibrators and 10 samples are analyzed with two digestions and two analyses/digestion for each sample and calibrator, in order to be sure of the instrumental stability.

- Digestion 1
  - Primary calibrator set - first run
  - Patient samples - first run
  - Primary calibrator set - second run
  - Patient samples - second run
- Digestion 2
  - Primary calibrator set - first run
  - Patient samples - first run
  - Primary calibrator set - second run
  - Patient samples - second run

### 3.5 Calculation of raw data and calibration

- The peaks for the peptides  $\delta$ T2,  $\delta$ T14 and  $\alpha$ T11 are carefully integrated by a well-settled data analysis program.
- From the peak areas raw data, the ratio  $\delta$ T2 vs  $\alpha$ T11 and  $\delta$ T14 vs  $\alpha$ T11 are calculated.
- The following notations are used:
  - $S_x$  = signal obtained for [HbA<sub>2</sub>] = area  $\delta$ T2 or  $\delta$ T14
  - $S_y$  = signal obtained for [total Hb] = area  $\alpha$ T11
  - $R_{\text{sign}}$  = ratio of signal  $S_x/S_y$
  - $Z_{\text{conc}}$  = percent HbA<sub>2</sub> in the sample
- For calibration of the method the ratio of the signals ( $R_{\text{sign}} = S_x/S_y$ ) is plotted versus the percent of HbA<sub>2</sub> ( $Z_{\text{conc}}$ ) of the primary calibrators, as assigned in the package insert of the calibrators.
- A linear regression analysis is performed.
- From this calibration function,  $Z_{\text{conc}}$  (HbA<sub>2</sub>, %) for unknown patient samples are calculated.

The HbA<sub>2</sub> concentrations in patient samples are calculated using each individual calibration curve (i.e. the calibration curve of the primary calibrator set - first run is used to give the values of HbA<sub>2</sub> to the patient samples – first run, and so on).

## **Part 4**                      **Glossary**

°C	degree centigrade
μL	microliter
μm	micrometer
ESI	electrospray ionization
HbA <sub>0</sub>	Hemoglobin (α <sub>2</sub> β <sub>2</sub> )
HbA <sub>2</sub>	Hemoglobin (α <sub>2</sub> δ <sub>2</sub> )
HCl	chloridric acid
HPLC	high performance liquid chromatography
KH <sub>2</sub> PO <sub>4</sub>	potassium dihydrogen phosphate
KCN	potassium cyanide
IT	ion trap
NaCl	sodium chloride
Na <sub>2</sub> EDTA	(ethylenedinitrilo)-tetraacetic acid disodium salt
NaOH	sodium hydroxide
nm	nanometer
primary calibrator	mixture of primary standards with specified HbA <sub>2</sub> content
primary standards	solution with certified concentration of purified HbA <sub>2</sub> or HbA <sub>0</sub>
psi	pounds/square inch (1 psi = 0.0689476 bar)
sec	seconds
TFA	trifluoroacetic acid
TFE	trifluoroethanol
TRIS	tris(hydroxymethyl)-aminomethane
UV	ultraviolet photometric detection
TSK gel Super-ODS	Trademark of Tosoh Bioscience

**Part 5                      Appendix**5.1 Example for Primary Calibrator Data Sheet

<b>Primary Calibrators HbA<sub>2</sub></b>
--

<b>Data Sheet lot CAL F05</b>
-------------------------------

General:                      material is for scientific use only within IFCC working group HbA<sub>2</sub> standardization.

Date of production:      June 2005

Raw materials:            Purified HbA<sub>2</sub> (at least 99.3 % pure)  
Purified HbA<sub>0</sub> (at least 98.7 % pure)

Storage buffer:            TRIS                      20 mmol/L  
KCN                        1 mmol/L  
Sucrose                    50 g/L  
pH                            7.5

Content each vial:        total Hb: 2.1 mg  
total volume: 70 µL  
total Hb concentration: 30 g/L

Storage conditions:      -80 °C

Shipment:                 dry ice

Identification code	HbA <sub>2</sub> %	Hb concentration g/L	Total Hb mg	Volume µL
2005-5	0.0	30	2.1	70
2005-6	1.5	30	2.1	70
2005-7	3.0	30	2.1	70
2005-8	4.5	30	2.1	70
2005-9	6.0	30	2.1	70
2005-10	7.5	30	2.1	70

Safety precautions: solution contain Potassium Cyanide. Material is of human origin.

## 5.2 HPLC-MS chromatograms

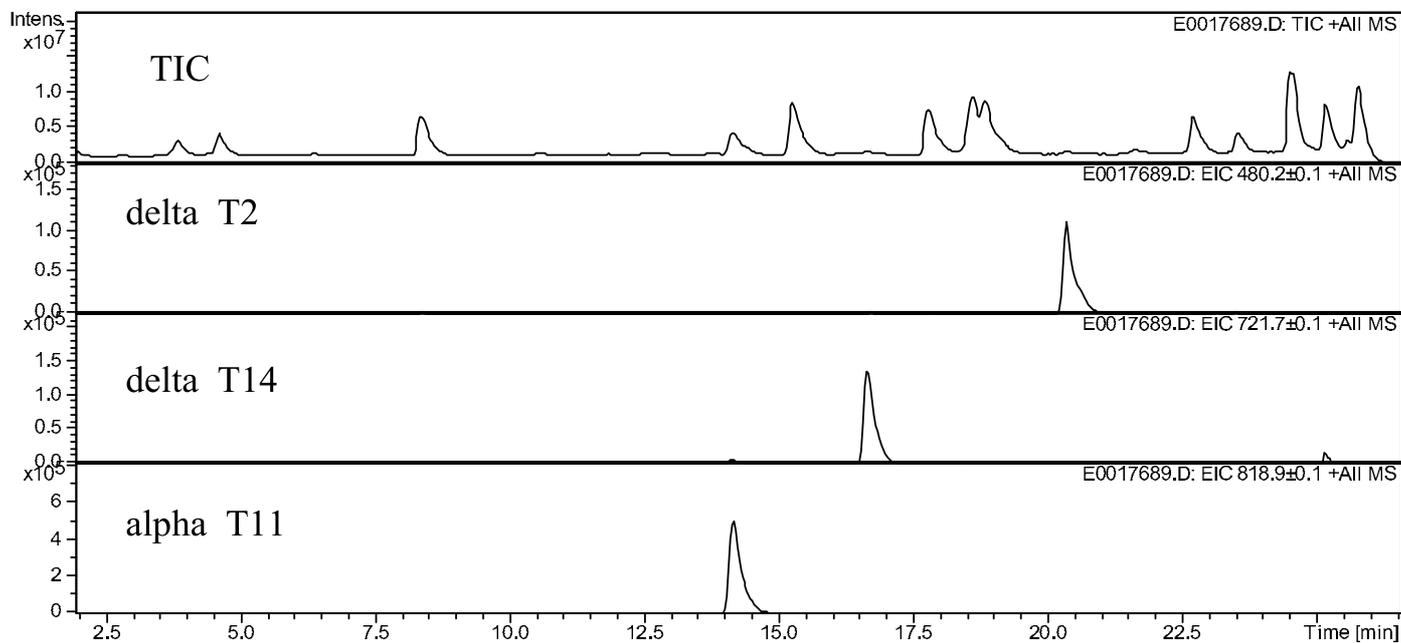
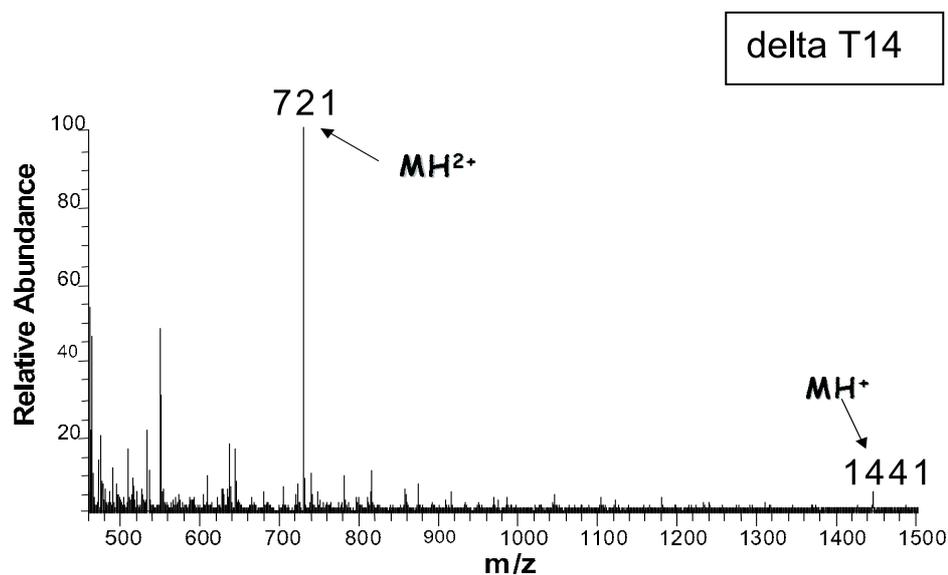
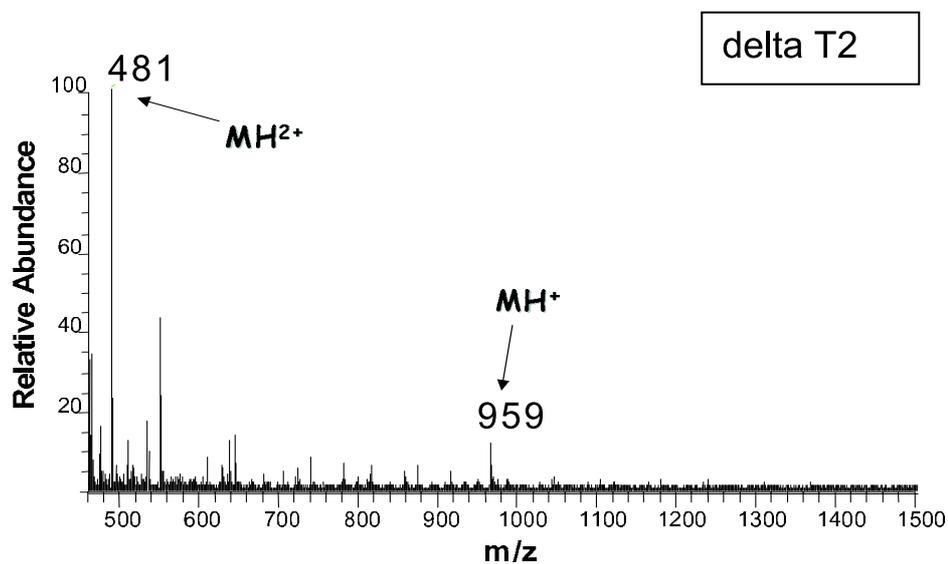


Fig. 1. Example of HPLC-MS chromatograms, for total ion current (TIC) and selected ion-monitoring of HbA<sub>2</sub> tryptic peptides ( $\delta$ T2 and  $\delta$ T14,  $\alpha$ T11).

### 5.3 Mass spectra



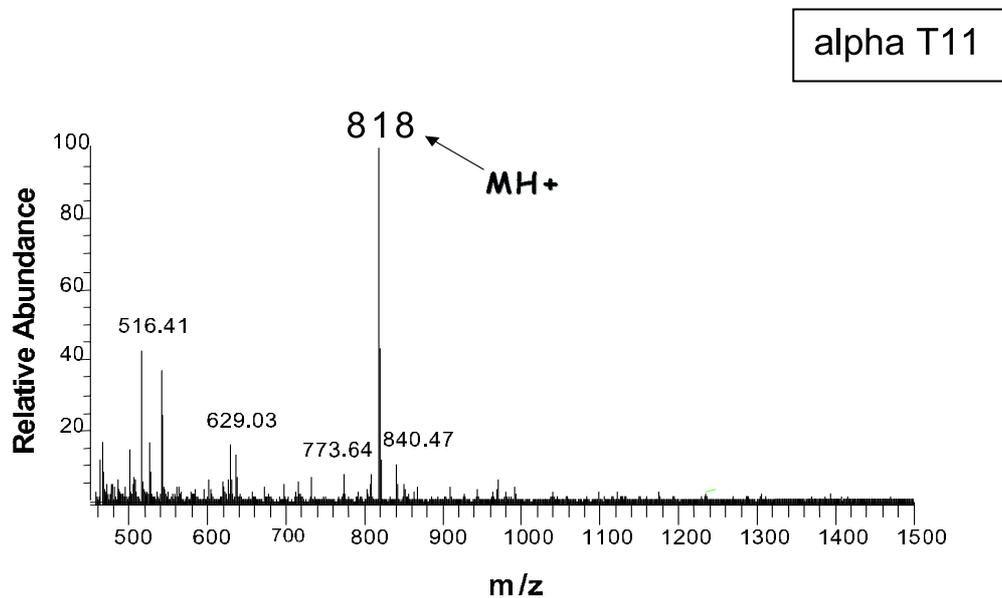


Fig. 2. Electrospray mass spectra of  $\delta$ T2,  $\delta$ T14 and  $\alpha$ T11 peptides.

5.4 Typical calibration curves

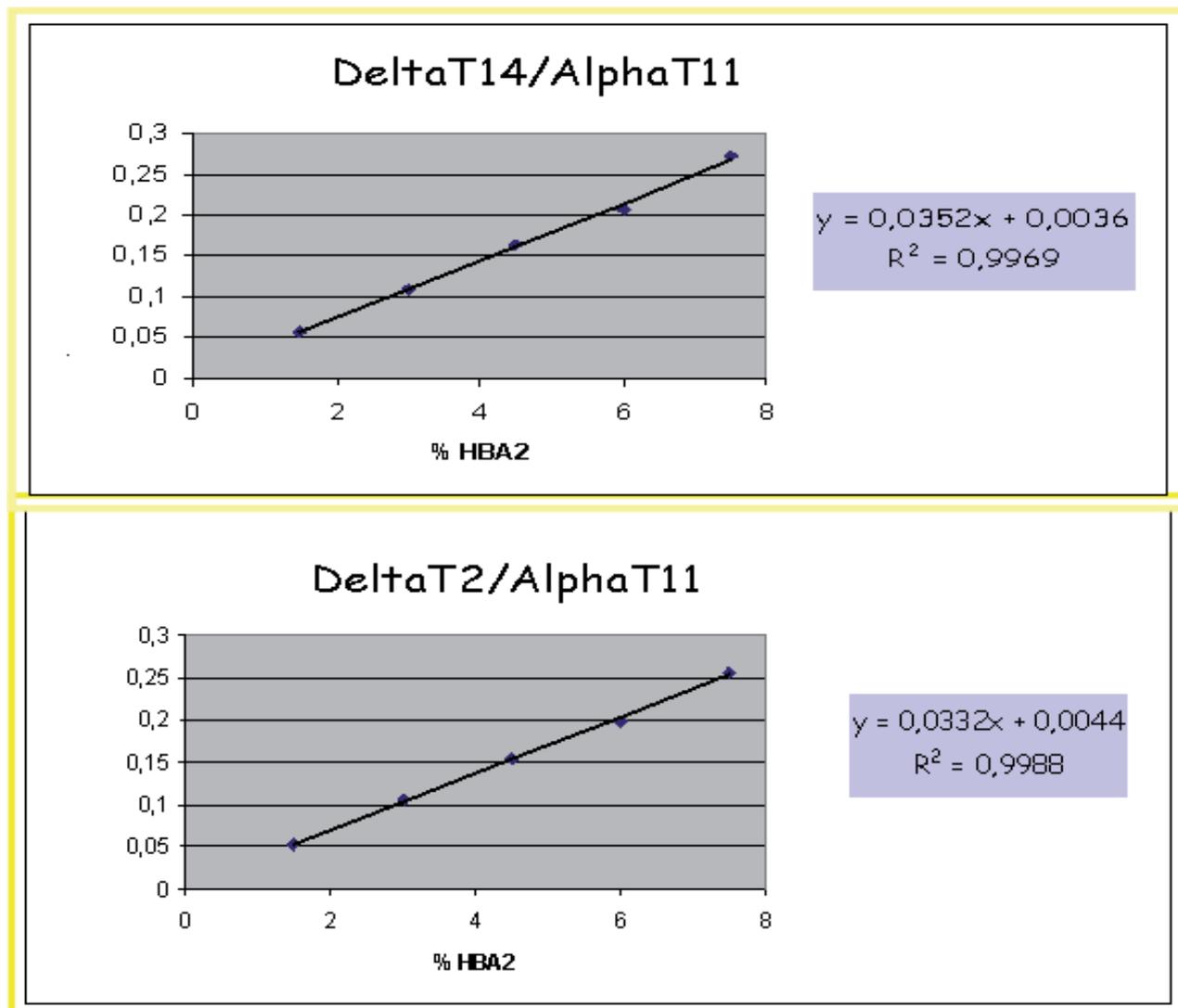


Fig. 3. Typical calibration curves for  $\delta T_{14}/\alpha T_{11}$  and  $\delta T_2/\alpha T_{11}$ .

## 5.5 Total hemoglobin determination

The purpose of this SOP is a precise determination of hemoglobin concentration in the hemolysates from patient samples.

### 5.5.1 *Test principle*

Total hemoglobin concentration is determined by the Drabkin method. The Fe(II) of hemoglobin is oxidized to Fe(III) by ferricyanide and then converted into stable cyanmethemoglobin by addition of KCN. The absorbance of cyanmethemoglobin is measured at 540 nm.

### 5.5.2 *Equipment*

Spectrophotometer at 540 nm.  
Balance in 3 digits after the g-decimal.  
Settable pipettes suitable to handle volumes of 20 and 1000 µL.  
Disposable reagent tubes of 10 mL.

### 5.5.3 *Solutions*

Drabkin reagent: 0.60 mmol/L potassium ferricyanide, 0,77 mmol/L KCN, 1.0 mmol/L KH<sub>2</sub>PO<sub>4</sub>, 1.0 g/L Triton X100, pH 7.2.  
Store in an amber glass bottle. Stable for 1 years at +4 °C.

### 5.5.4 *Measurement procedure*

- Perform the measurements in triplicate
- Pipette 2 mL of Drabkin reagent in 10 mL test tubes.
- Add 20 µL of hemolysates.
- Mix carefully.
- Let stand for 10-30 minutes at room temperature.
- Measure absorbance at 540 nm against water.

### 5.5.6 *Calculation of total hemoglobin*

$$\text{Total hemoglobin, g/L} = (A \times 101)/0.686$$

A = absorbance at 540 nm

101 = dilution factor

0.686 = extinction coefficient of Hb at 540 nm expressed in g/L

### 3. Caractérisation de mutations de l'hémoglobine

Ce projet a été réalisé dans le cadre d'une collaboration avec le Dr Emmanuel Bissé de l'Université de Fribourg (Allemagne).

#### 3.1. Contexte de l'étude

La mise en évidence d'une mutation d'hémoglobine est un élément capital pour le diagnostic d'une thalassémie. Les mutations les plus courantes peuvent être identifiées par des techniques d'analyse, peu coûteuses et performantes, disponibles dans les laboratoires d'analyse biologique. Malheureusement, l'identification de mutations peut être difficile (il existe plus de 1000 mutations décrites) et dans le cas de mutations nouvelles ces techniques d'analyse classiques sont inopérantes. L'objectif de l'étude présentée ici est de tirer partie des méthodes d'analyse protéomique classique (digestion enzymatique, LC-MS et LC-MS/MS) pour caractériser de nouveaux variants de l'hémoglobine en les adaptant.

L'analyse par spectrométrie de masse des chaînes de globines entières permet de mesurer la différence de masse entre les globines normales et la forme mutante et permet ainsi d'identifier les variants [Dilip K. Rai et al. 2002, D. K. Rai et al. 2004, Kleinert et al. 2008]. Mais dans la plupart des cas, une seconde approche analytique est nécessaire [Caruso et al. 2005]. Des méthodes de criblage des mutations par spectrométrie de masse existent mais sont limitées à l'analyse d'échantillons de sang qui contiennent des variants d'hémoglobine connus [Daniel et al. 2005, Basilico et al. 2007, Kleinert et al. 2008, Zanella-Cleon et al. 2009]. Caruso et al [Caruso et al. 2005] montrent que différentes stratégies analytiques sont nécessaires pour caractériser certaines hémoglobines anormales liées à la grande variété de protéines mutées et de situations cliniques.

Pour caractériser ces nouvelles mutations, nous avons développé une stratégie qui combine une étape de purification de l'hémoglobine mutée et des outils utilisés pour l'analyse protéomique. En effet, l'analyse des échantillons d'hémoglobine par les méthodes standards d'hématologie nous fournissent des hypothèses sur la nature des mutations qui peuvent se révéler fausses par la suite. De plus, les mutations peuvent être partielles (la chaîne de globine mutée est donc présente en faible quantité dans l'hémolysat) ou sont parfois portées par des chaînes minoritaires comme la globine delta. Dans ces cas, les expériences de LC-MS sur un digestat total d'hémolysat ne permettent pas de détecter le peptide qui porte la mutation.

Nous avons donc développé une stratégie qui consiste à isoler l'hémoglobine mutée, par chromatographie échangeuse d'ions ou par RP-HPLC. Cette fraction est ensuite analysée par LC-MS pour mesurer la masse moléculaire de la chaîne mutée. La différence de masse entre la chaîne mutée et la chaîne normale donne un nombre limité de possibilités de changement d'acides aminés

gouvernés par le changement d'une seule base dans le codon nucléotidique (Tableau 6). Cette première étape permet de cibler les acides aminés qui pourraient induire cette différence de masse.

		Chiffre des unités									
		0	1	2	3	4	5	6	7	8	9
Chiffre des dizaines	0	Leu->Ile Gln->Lys	Lys->Glu Ile->Asn Gln->Glu Asn->Asp		Lys->Met	Pro->Thr					Gln->His
	10	Ser->Pro		Thr->Ile	Thr->Asn	Gly->Ala Val->Leu Val->Ile Asp->Glu Ser->Thr Asn->Lys	Leu->Gln Ile->Lys	Val->Asp Ala->Ser Phe->Tyr Pro->Leu Ser->Cys		Ile->Met Leu->Met	His->Arg
	20			22 Da = Asp->His	Asn->His	Leu->His	Met->Arg	His->Tyr Ala->Pro Ser->Leu Ser->Ile	Ser->Asn Thr->Lys	Lys->Arg Gln->Arg Ala->Val	
	30	Val->Glu Gly->Ser Ala->Thr Arg->Trp Thr->Met	Pro->Gln	Val->Met		Leu->Phe Ile->Phe					
	40	Pro->His		Gly->Val	Leu->Arg Ile->Arg	Ala->Asp Cys->Phe		Gly->Cys		Val->Phe Asp->Tyr	Asn->Tyr
	50				Cys->Arg		Thr->Arg			Gly->Asp Ala->Glu	Pro->Arg
	60	Cys->Tyr Ser->Phe									Ser->Arg
	70			Gly->Glu	Leu->Trp			Ser->Tyr			
	80				Cys->Trp						
	90										Gly->Arg Ser->Trp
	120										Gly->Trp

**Tableau 6.** Différences de masse induites par le changement d'une base dans le codon nucléotidique (seuls sont présentées les mutations autorisées par le code génétique). Par exemple, le changement de Asp en His augmente la masse de la chaîne de globine de 22 Da et le changement inverse diminue la masse de 22 Da. D'après : HbVar: a Database of Human Hemoglobin Variants and Thalassemias. [http://globin.cse.psu.edu/hbvar/mass\\_chart.html](http://globin.cse.psu.edu/hbvar/mass_chart.html)

Après avoir localisés les sites de mutation potentiels sur les peptides, la protéine purifiée est soumise à une digestion enzymatique puis le digestat est analysé par LC-MS et LC-MS/MS. Les peptides identifiés permettent alors de caractériser la mutation.

## 3.2. Résultats

Ce travail a permis d'identifier de nouvelles mutations, qui n'avaient encore jamais été décrites, chez des patients malades pour lesquels les résultats d'analyse d'hématologie classique ne permettaient pas d'établir un diagnostic sûr.

### 3.2.1. Caractérisation de la mutation "Noah Mehmet Oeztuerk"

L'une des mutations qui a pu être caractérisée par cette étude, et que nous avons nommé "Noah Mehmet Oeztuerk", est portée par la chaîne delta et correspond au remplacement d'une histidine par une tyrosine en position 143. Ce travail a donné lieu à une publication située en annexe 6.

Pour cet échantillon, l'analyse par LC-MS sur phase inverse de l'hémolysât total n'a pas permis de détecter la chaîne delta mutée. En effet, dans cet échantillon, la proportion de HbA<sub>2</sub> était faible (~ 1,2%) et la forme mutée représentait environ 50% de HbA<sub>2</sub> total soit environ 0,6% de l'hémolysat total. Pour la caractériser, le variant de HbA<sub>2</sub> a donc été isolé par chromatographie sur support DEA-cellulose.

L'analyse par LC-MS de cet échantillon purifié a permis de mettre en évidence une différence de + 26 Da entre les deux chaînes delta mutée et non mutée, ce qui peut correspondre à quatre mutations : His→Tyr, Ala→Pro, Ser→Leu et Ser→Ile. Au cours de cette étude, les analyses ADN effectuées par nos collaborateurs ont permis de réduire le champ des possibilités à la mutation : His→Tyr en position 143.

Après digestion trypsique, l'échantillon a été analysé par nanoLC-MS et nanoLC-MS/MS. Le peptide portant la mutation a ainsi été identifié et caractérisé grâce à l'interprétation de son spectre de fragmentation par séquençage *de novo*. L'ensemble des informations relatives à la préparation d'échantillon et aux paramètres instrumentaux sont disponibles dans la partie matériel et méthode de la publication en annexe 5.

Nous avons ainsi caractérisé un nouveau variant de la chaîne delta de l'hémoglobine pour lequel, l'histidine en position 143 est remplacée par une tyrosine. Bien que cette mutation semble accroître l'affinité de l'hémoglobine pour l'oxygène, étant donné le faible taux de HbA<sub>2</sub> dans le sang, les mutations aux niveaux de la chaîne delta ont généralement peu d'influence. Cependant, la présence d'une mutation entraîne une diminution du taux de HbA<sub>2</sub> normal et peut donc entraîner une erreur lors du diagnostic de la beta-thalassémie, qui est basée sur le dosage de HbA<sub>2</sub> (cf. § 2.1).

### 3.2.2. Caractérisation d'une nouvelle mutation sur la chaîne beta 93Cys→Ser

Dans cette deuxième étude, la mutation se situe au niveau de la chaîne beta. Contrairement au variant décrit précédemment (Noah Mehmet Oeztuerk) qui se situait sur la chaîne delta, une LC-MS en phase inverse de l'hémolysat total a permis de séparer les deux formes de chaîne beta, mutée et sauvage et ainsi de collecter la forme mutée. La différence de masse observée entre les deux formes était de - 16 Da ce qui peut correspondre à cinq mutations : Asp→Val, Ser→Ala, Tyr→Phe, Leu→Pro et Cys→Ser.

La chaîne beta mutée purifiée a été digérée avec de la trypsine et analysée par microLC-MS et microLC-MS/MS. L'interprétation des données expérimentales MS/MS a permis d'identifier la globine beta avec un pourcentage de couverture de 73%, à l'aide du moteur de recherche Mascot.

Les résultats sont présentés dans le Tableau 7. Tous les peptides surlignés en gris ont été identifiés par microLC-MS/MS du digestat de la chaîne mutée. Les deux seuls peptides contenant des sites potentiels de mutation qui n'ont pas été identifiés sont :  $\beta$ T10 et  $\beta$ T12. La mutation devait donc se situer sur l'un de ces deux peptides.

Peptides tryptiques de la chaîne beta					
nom	position	(M + H) <sup>+</sup>	(M + 2H) <sup>2+</sup>	(M + 3H) <sup>3+</sup>	séquence
$\beta$ T1	1-8	952,51	476,75	318,17	VHL <b>L</b> TPEEK
$\beta$ T2	9-17	932,52	466,76	311,51	<b>S</b> AVTAL <b>L</b> WGK
$\beta$ T3	18-30	1314,66	657,83	438,89	V <b>N</b> VDEVGGE <b>A</b> LGR
$\beta$ T4	31-40	1274,73	637,86	425,58	<b>L</b> L <b>V</b> V <b>Y</b> PWTQR
$\beta$ T5	41-59	2058,95	1029,97	686,98	FF <b>S</b> FGD <b>L</b> STP <b>D</b> AVMGNPK
$\beta$ T6	60-61	246,18	123,59	82,73	VK
$\beta$ T7	62-65	412,23	206,62	138,08	AHGK
$\beta$ T8	66-66	147,11	74,06	49,70	K
$\beta$ T9	67-82	1669,89	835,45	557,30	V <b>L</b> GAF <b>S</b> D <b>G</b> LAH <b>L</b> D <b>N</b> LK
$\beta$ T10	83-95	1421,67	711,34	474,56	GTFAT <b>L</b> SE <b>L</b> H <b>C</b> DK
$\beta$ T11	96-104	1126,56	563,78	376,19	<b>L</b> H <b>V</b> D <b>P</b> ENFR
$\beta$ T12	105-120	1719,97	860,49	573,99	<b>L</b> LGN <b>V</b> L <b>V</b> C <b>V</b> LAH <b>H</b> FGK
$\beta$ T13	121-132	1378,70	689,85	460,23	EFTPPVQA <b>A</b> YQK
$\beta$ T14	133-144	1149,67	575,34	383,89	VVAGVAN <b>A</b> L <b>A</b> HK
$\beta$ T15	145-146	319,14	160,07	107,05	<b>Y</b> H

Peptides identifiés lors de l'analyse par nanoLC-MS/MS de la chaîne mutée  
**en rouge** Sites de mutation potentiels

Tableau 7. Digestion tryptique *in silico* de la chaîne beta

Les courants d'ions extraits correspondant aux peptides théoriques  $\beta$ T10' et  $\beta$ T12' (Tableau 8) montrent un pic intense à 24 min pour l'ion  $(M+2H)^{2+}$  à  $m/z$  703,3 qui correspond à la masse de  $\beta$ T10'. Cette information a permis de localiser la zone d'élution de ce peptide et de rechercher ainsi son spectre de fragmentation MS/MS (Figure 9).

Peptides tryptiques de la chaîne beta potentiellement mutés					
nom	position	(M + H) <sup>+</sup>	(M + 2H) <sup>2+</sup>	(M + 3H) <sup>3+</sup>	séquence
$\beta$ T10	83-95	1421,67	711,34	474,56	GTFATLSELHCDK
$\beta$ T10'	83-95	1405,67	703,34	469,22	GTFATPSELHCDK
					GTFATLSEPHCDK
					GTFATLSELHSDK
					GTFATLSELHCVK
$\beta$ T12	105-120	1719,97	860,49	573,99	LLGNVLVCVLAHHFGK
$\beta$ T12'	105-120	1703,97	852,49	568,66	PLGNVLVCVLAHHFGK
					LPGNVLVCVLAHHFGK
					LLGNVPVCVLAHHFGK
					LLGNVLVSVLAHHFGK
					LLGNVLVCVPAHHFGK

en rouge Sites de mutation potentiels

en bleu Site muté

Tableau 8. Séquences des peptides tryptiques de la chaîne beta potentiellement mutés

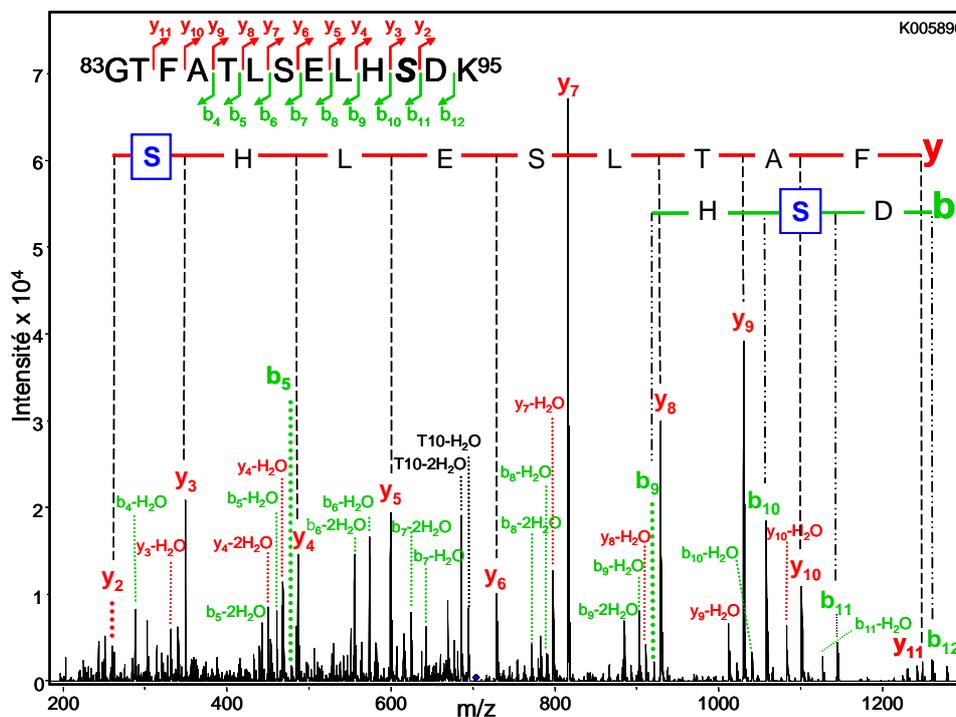


Figure 9. Spectre MS/MS de l'ion 703,3 qui correspond au peptide  $\beta$ T10' : <sup>83</sup>GTFATLSELHSDK<sup>95</sup>

L'interprétation *de novo* de ce spectre a permis d'identifier, le tag FATLSELHSD qui est inclus dans le peptide  $^{83}\text{GTFATLSELHSDK}^{95}$  soit  $\beta\text{T10}$  muté au niveau de la cystéine en position 93 qui est remplacé par une sérine (Figure 9).

Nous avons donc identifié une nouvelle mutation située sur la chaîne beta de HbA<sub>0</sub> en position 93 et qui correspond au remplacement d'une cystéine par une sérine. Une étude fonctionnelle est en cours afin d'établir l'impact de cette mutation au niveau physiologique. Une publication est en cours de rédaction.

### **3.3. Conclusion et perspectives**

Ce travail illustre la puissance des expériences de LC-MS/MS et de l'interprétation automatique des données pour la caractérisation de mutants d'hémoglobine humaine.

Cette méthodologie pourrait être appliquée de manière systématique pour la détection, chez les nouveaux nés, des mutations les plus courantes. Pour cela, une mesure de masse des chaînes entières de globine alpha et beta permettrait de déterminer la présence ou non d'une mutation. Après digestion enzymatique, une analyse LC-MS/MS suivie d'une interprétation automatisée à l'aide des logiciels utilisés en protéomique permettrait d'identifier les mutations les plus courantes.

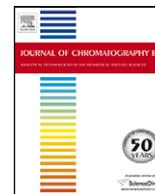
## **ANNEXE 6: Publication**

**Hemoglobin Noah Mehmet Oeztuerk, ( $\alpha 2\delta 2143$  (H21)His  $\rightarrow$  Tyr: A novel  $\delta$ -chain variant in the 2,3-DPG binding site**

Emmanuel Bisse, Christine Schaeffer, Agnès Hovasse, Sabine Preisler-Adams, Thomas Epting, Manfred Baumstark, Alain Van Dorselaer, Jürgen Horst Horst, Heinrich Wieland Wieland

Journal of Chromatography B, 2008, 871(1), 55-59.





## Haemoglobin Noah Mehmet Oeztuerk ( $\alpha_2 \delta_2 143$ (H21)His $\rightarrow$ Tyr: A novel $\delta$ -chain variant in the 2,3-DPG binding site

Emmanuel Bissé<sup>a,\*</sup>, Christine Schaeffer<sup>b</sup>, Agnès Hovasse<sup>b</sup>, Sabine Preisler-Adams<sup>c</sup>, Thomas Epting<sup>a</sup>, Manfred Baumstark<sup>d</sup>, Alain Van Dorselaer<sup>b</sup>, Jürgen Horst<sup>c</sup>, Heinrich Wieland<sup>a</sup>

<sup>a</sup> Department of Clinical Chemistry, University Medical Center, Hugstetterstrasse 55, D-79106 Freiburg, Germany

<sup>b</sup> Laboratoire de Spectrométrie de Masse Bio-Organique UMR 7178, CNRS/ULP: Institut Pluridisciplinaire Hubert Curien, Université Louis Pasteur, 67087 Strasbourg, France

<sup>c</sup> Institut für Humangenetik Westfälische Wilhelms-Universität, D-48149 Münster, Germany

<sup>d</sup> Department of Prevention, Rehabilitation, and Sports Medicine, University Hospital, Freiburg, Germany

### ARTICLE INFO

#### Article history:

Received 31 December 2007

Accepted 22 June 2008

Available online 9 July 2008

#### Keywords:

Hb  $\delta$ -variant

Nano-LC-MS

HPLC

Oxygen affinity

### ABSTRACT

A new  $\delta$ -chain variant,  $\delta 143$  (H21) His  $\rightarrow$  Tyr or Hb Noah Mehmet Oeztuerk, was discovered during the investigation of the cause of hemolytic anaemia in a 6-month-old infant of Turkish descent. It was detected by Cation exchange high-performance liquid chromatography (CE-HPLC) using PolyCAT A column.  $P_{50}$  was  $20.6 \pm 0.60$  mmHg and  $29.3 \pm 0.40$  mmHg for the carrier and the wild-type, respectively. This suggests an increase in oxygen affinity. On routine CE-HPLC Hb A<sub>2</sub> was low (1.2%) and the variant was not detected. An extended family study revealed that the variant was not associated with the anaemia or with any other clinical abnormality.

© 2008 Elsevier B.V. All rights reserved.

### 1. Introduction

Haemoglobin A<sub>2</sub> or  $\alpha_2 \delta_2$  is a naturally occurring haemoglobin which has electrophoretically less negative charge than the adult haemoglobin Hb A. The primary sequence of the  $\delta$ -chain differs from the  $\beta$ -chain of Hb A in 10 out of 146 residues [1]. Haemoglobin A<sub>2</sub> (Hb A<sub>2</sub>) is expressed at low concentration and its concentration in red cells from normal adults is about 2–3%. In patients with  $\beta$ -thalassemia Hb A<sub>2</sub> levels are often increased (>3.5%). The role of Hb A<sub>2</sub> in the human normal red cells is not yet known. But, its increase is used as a criterion for diagnosis of  $\beta$ -thalassemia. The methods commonly used to determine Hb A<sub>2</sub> include elution following cellulose acetate electrophoresis, DE-52 micro-chromatography and Cation exchange high-performance liquid chromatography (CE-HPLC). CE-HPLC is emerging as the method of choice for quantification of Hb A<sub>2</sub> and Hb F.  $\delta$ -globin gene mutations, can, however, mask the presence of  $\beta$ -thalassemia when the diagnosis is based only on the determination of Hb A<sub>2</sub>. This depends on whether the variant co-chromatographs with Hb A<sub>2</sub> or not. In this report, we have characterized a new  $\delta$ -chain variant in eight members of a

Turkish family with a spurious Hb A<sub>2</sub> level of 1.2% as measured by a routine, 7.5 min, CE-HPLC analyzer (Tosoh HLC-723 GHb VII).

### 2. Materials and methods

#### 2.1. Blood samples

Venous blood was collected in tubes containing disodium salt of ethylenediaminetetraacetic acid (EDTA) as anticoagulant. Samples from 7 family members, living in Turkey were shipped by air to Freiburg, Germany. Informed consent was obtained.

#### 2.2. Haematology and haemoglobin analyses

Haematological evaluations were carried out on routine equipment. The presence of a possible abnormal Hb variant was assessed by electrophoretic and chromatographic procedures routinely used in our laboratory [2]. Haemoglobin components were separated and quantified by high-resolution cation exchange HPLC using PolyCAT A column [2].

Isolation of normal and abnormal Hb A<sub>2</sub> required a preparative chromatography. It was performed on diethylaminoethyl-cellulose as described before [3]. A column (30 mm  $\times$  250 mm) of DEAE-cellulose equilibrated with 0.2 M glycine –0.01% KCN buffer was used. Hemolysate containing 200 mg haemoglobin is dialyzed against water overnight at 4 °C and applied to the column. The

\* Corresponding author. Tel.: +49 761 270 3627; fax: +49 761 270 3444.

E-mail address: [emmanuel.bisse@uniklinik-freiburg.de](mailto:emmanuel.bisse@uniklinik-freiburg.de) (E. Bissé).

wild-type Hb A<sub>2</sub> and the variant  $\delta$ -chain were eluted with a linear gradient of NaCl going from 0.0 to 0.2 M in the equilibrating buffer. The purity was checked by PolyCAT A HPLC.

### 2.3. Mass spectrometry

#### 2.3.1. Material and reagents

All chemicals and reagents were of analytical grade and purchased from Sigma–Aldrich (Steinheim, Germany). Tris–HCl was from Euromedex (Souffelweyersheim, France). Porcine trypsin was from Promega (Madison, WI, USA). Distilled water produced from a Milli-Q Water System™ (Millipore, Guyancourt, France) was used for the preparation of the buffer and solvents.

#### 2.3.2. Mass spectrometry analysis

Mutated haemoglobin (5  $\mu$ l) fraction isolated by DEACellulose chromatography was denatured with 5  $\mu$ l of 2,2,2-trifluoroethanol. Then 30  $\mu$ l of Tris–HCl (50 mM, pH 8.5) buffer and porcine trypsin in the ratio 1/20 (w/w) were added. The sample was left at 39 °C for twenty hours. The digestion was stopped with 1  $\mu$ l of 10% formic acid. Before LC–MS or LC–MS/MS analysis, the solution was diluted (1:50) with 0.1% trifluoroacetic acid in water. One microliter of this solution was finally injected by nano-liquid chromatography–mass spectrometry (nano-LC/MS) and nano-LC–MS/MS. The analyses were performed using an Agilent 1100 series HPLC–Chip/MS system (Agilent Technologies, Palo Alto, USA) coupled to an HCT Plus ion trap (Bruker Daltonics, Bremen, Germany). The column used was a chip C18 SB-ZORBAX, 300 Å (75  $\mu$ m  $\times$  43 mm, 5  $\mu$ m). The solvent system consisted of 2% acetonitrile, 0.1% formic acid in water (solvent A) and 2% water, 0.1% formic acid in acetonitrile (solvent B). Elution was performed at a flow rate of 300 nl/min with a 8–40% gradient (solvent B) over the first 7 min followed by a 70% stage (solvent B) over 3 min before the reconditioning of the column at 92% of solvent A.

The voltage applied to the capillary cap was optimized to –1750 V. For tandem MS experiments, the system was operated with automatic switching between MS and MS/MS modes. The three most abundant peptides, preferring doubly charged ions, were selected on each MS spectrum for further isolation and fragmentation. The MS/MS scanning was performed in the ultrascan resolution mode at a scan rate of 26,000  $m/z$  per second. A total of 6 scans were averaged to obtain a MS/MS spectrum. The complete system was fully controlled by ChemStation (Agilent Technologies) and EsquireControl (Bruker Daltonics) softwares.

Mass data collected during nano-LC–MS/MS analysis were processed, converted into .mgf files, and interpreted using a local Mascot (Matrix Science, London, U.K.) server. Searches were performed against the National Center for Biotechnology Information nonredundant (NCBI) database without any taxonomic, molecular weight, or isoelectric point restrictions. Searches were performed with a tolerance on mass measurements of 0.25 Da in both MS and MS/MS modes. For peptides not corresponding to expected masses for normal delta haemoglobin, the corresponding MS/MS spectrum was submitted to manual de novo sequencing.

### 2.4. DNA analysis

Genomic DNA was extracted from peripheral blood using the QIAamp DNA Mini Kit (Qiagen) according to the manufacturer's protocol.

The three exons of the  $\delta$ -globin gene were amplified by PCR using the oligonucleotides C, F, H and I, as published by De Angioletti et al. [4]. The amplified products were sequenced directly with the Dye Primer Cycle Sequencing Ready Reaction Kit (PE Applied Biosystems) according to the manufacturer's recommendations

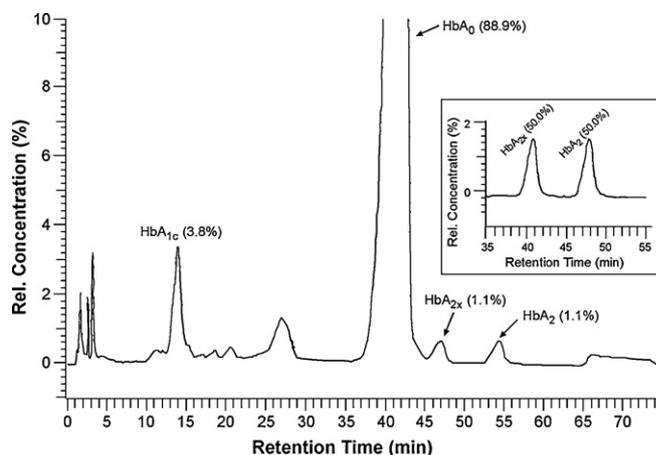


Fig. 1. CE-HPLC separation of the Hb components in the whole red cell lysate from a carrier of Hb Noah Mehmet Oeztuerk and in the Hb fraction isolated by DEACellulose chromatography (inset).

(oligonucleotides as for amplification). Sequencing reactions were analyzed on an ABI 3730 automated sequencer.

### 2.5. Oxygen affinity

The oxygen binding was assessed on the isolated Hb A2 fraction containing 2.5 g/dl haemoglobin. The determination of the oximetric parameters was performed on a ABL720 blood gas analyzer (Radiometer Copenhagen).  $P_{50}$ , the partial oxygen pressure for a given blood sample at which the haemoglobin is half-saturated, was calculated using the oxygen status algorithm, version 3, of Siggaard-Andersen and Siggaard-Andersen [5]. Catalase (20 mg/l) was added to the effluent fractions to limit methhaemoglobin formation.

## 3. Results

### 3.1. Family

The propositus, a 6-month-old infant of Turkish descent was admitted to the hospital for an investigation of the cause of the anaemia. He had a significant haemolytic anaemia with hypochromia and anisocytosis and was transfusion-dependent. The Hb level was 6.8 g/dl, mean corpuscular volume (MCV) 97.6 fl, mean corpuscular Hb (MCH) 27.3 pg, mean corpuscular Hb concentration (MCHC) 28.0 g/dl, and packed cell volume (PCV) 0.244 l/l. There is the suspicion that the propositus has a Diamond-Blackfan syndrome, a red cell aplasia in which the bone marrow is of normal cellularity but in which there is a relative absence of erythroid precursors.

He was found to have a spurious low level of Hb A2. Thirteen members of his family were evaluated; eight including her father showed inconspicuous low Hb A2 concentration (1.0–1.2%) but were not anaemic. Hb ranged from 13.0 to 15.8 g/dl.

### 3.2. Haemoglobin analyses

Electrophoresis on agarose gel and DE-52 micro-chromatography failed to reveal the presence of an abnormal haemoglobin, the level of Hb A<sub>2</sub> was  $1.47 \pm 0.22\%$  and  $2.10 \pm 0.17\%$ , respectively. The samples from the carriers run on a routine, 7.5 min, CE-HPLC (Tosoh HLC-723 GHb VII) showed a mean level of Hb A<sub>2</sub> of  $1.23 \pm 0.11\%$  and the variant was not detected. CE-HPLC with PolyCAT A column (Fig. 1) demonstrated the presence of minor abnormal peak of 1.0% with a retention time of 48 min eluting

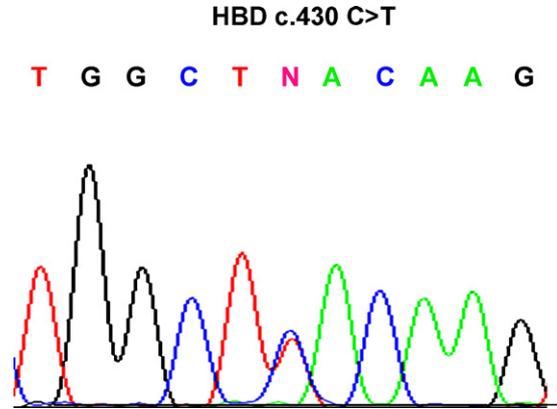
between Hb A0 and Hb A2. Hb F (2.3%) was detectable for the propositus. Lysate from affected subjects showed the normal HbA<sub>2</sub> ( $1.08 \pm 0.05$ ) and the abnormal component ( $1.06 \pm 0.04\%$ ) in equal proportions. The Hb variant was masked after blood transfusion for the propositus, but the isolated total Hb A2 fraction comprised approximately an equal proportion of the normal and abnormal components after chromatography on PolyCAT A HPLC (Fig. 1 inset). The purity of Hb A<sub>2</sub> isolated from a control was 99%. Due to the low amount of Hb variant LC–MS experiments performed on the tryptic digest of the whole red cell lysate did not permit the detection of the variant peptide. Therefore, abnormal HbA<sub>2</sub> was purified by DEA-cellulose chromatography to perform further mass spectrometry experiments.

### 3.3. DNA analysis

Analysis of the DNA sequence of the amplified  $\delta$ -globin gene fragment from the propositus and his father revealed the presence of single CAC  $\rightarrow$  TAC base change at nucleotide 430 codon 143 in DNA that leads to a replacement of histidine by tyrosine (Fig. 2).

### 3.4. Mass spectrometry

The mutant HbA<sub>2</sub> isolated from DEA-cellulose chromatography was submitted to trypsin digestion and the mixture of peptides was analyzed by nano-LC–MS, in order to have the very best chromatographic time resolution to detect any shoulder in the chromatogram given by the ion current (Fig. 3). Then a nano-LC–MS/MS experiment was performed in exactly the same chromatographic conditions allowing the identification of peptides. Results of identifications from molecular masses and fragmentations are summarized on Table 1. The chromatographic peak at 6 min displays a mass of 1174.6 Da which does not correspond to any expected mass for a tryptic peptide from HbA<sub>2</sub>. Its fragmentation spectrum shows it has a sequence identical to that of normal peptide delta T15 with a mutation of His 143 to Tyr. The fragments showing the loss of Tyr 143 display a  $\Delta M$  of 163.01 Da

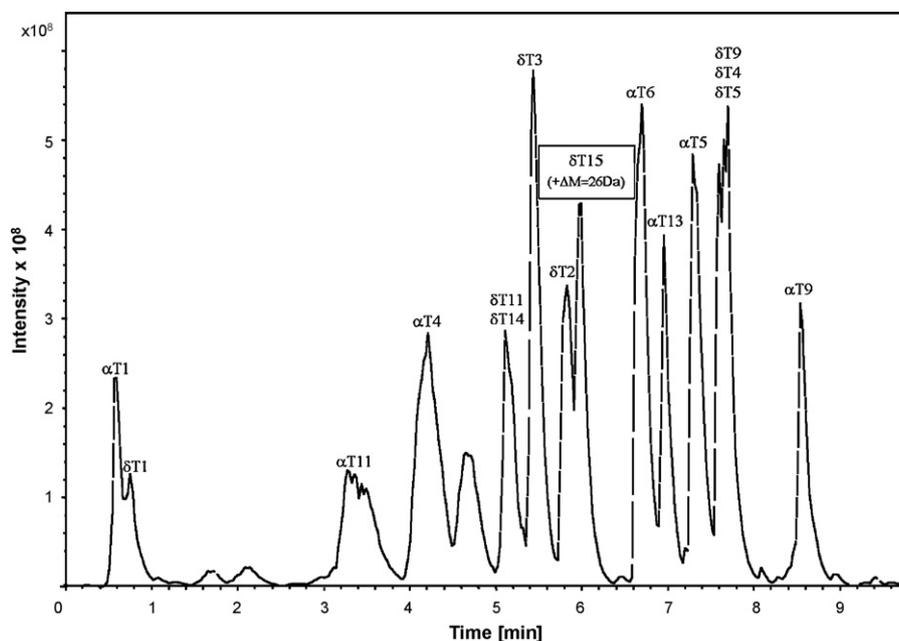


**Fig. 2.** DNA sequencing of the  $\delta$ -globin gene. Portion of sequence printout showing location of a transitional mutation that results in replacement of arginine by tyrosine in codon 143, heterozygosity for c.430 C > T.

for a theoretical  $\Delta M$  of 163.06 Da. This mutation deduced from the MS/MS spectrum (Fig. 4) matches the difference in molecular mass observed (measured: 1175.6 Da; calculated: 1149.7;  $\Delta M = 25.9$  Da).

### 3.5. Oxygen affinity

The data represent the mean  $\pm$  standard deviation (S.D.) from isolated Hb A<sub>2</sub> from 3 carriers and 3 control persons. On the isolated Hb A<sub>2</sub> fractions from the carriers the mean oximetric readings were: pH  $7.56 \pm 0.02$ ; COHb =  $0.58 \pm 0.20\%$ ; Hb =  $2.23 \pm 0.06$  g/dl; methhaemoglobin =  $1.1 \pm 0.15\%$ ; sO<sub>2</sub> (oxygen saturation) =  $92.3 \pm 0.52\%$ ; pO<sub>2</sub> =  $50.70 \pm 2.9$  mmHg; pCO<sub>2</sub> =  $3.8 \pm 0.02$  mmHg. They were for the control persons as follows: pH  $7.30 \pm 0.04$ ; COHb =  $0.40 \pm 0.14\%$ ; Hb =  $3.2 \pm 0.25$  g/dl; methhaemoglobin =  $0.9 \pm 0.10\%$ ; sO<sub>2</sub> (oxygen saturation) =  $92.0 \pm 0.45\%$ ; pO<sub>2</sub> =  $69.40 \pm 2.5$  mmHg; pCO<sub>2</sub> =  $3.8 \pm 0.02$  mmHg. The mean calculated P<sub>50</sub> was lower in the Hb A<sub>2</sub> fraction from the carrier than in the Hb A2 fraction from the control persons ( $20.6 \pm 0.60$  mmHg versus  $29.3 \pm 0.40$  mmHg).

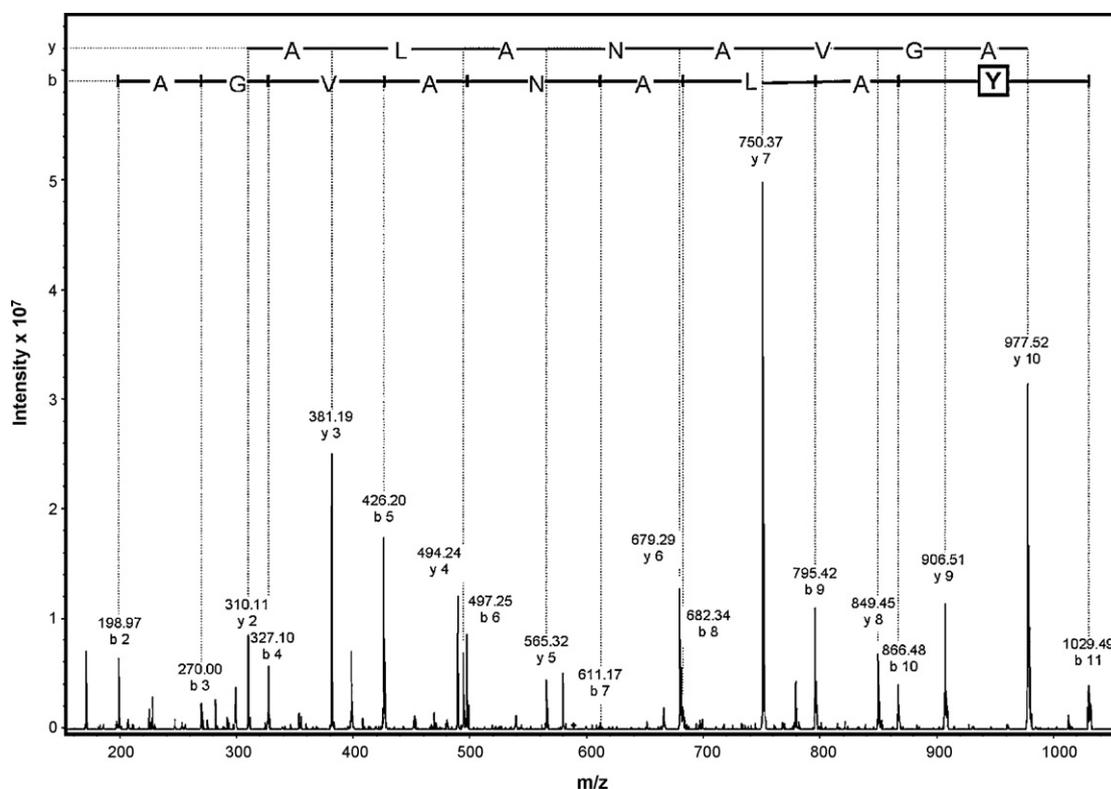


**Fig. 3.** Base peak chromatogram (BPC) obtained from nano-LC–MS analysis of tryptic HbA<sub>2</sub> mutant. The peak at 6 min exhibited a mass of 1148.7 Da which does not correspond to any tryptic peptide expected for normal HbA<sub>2</sub>.

**Table 1**  
Peptides identified by nano-LC–MS and nano-LC–MS/MS analyses of tryptic digest from isolated mutant HbA<sub>2</sub>

Fragment	Peptide sequence	Theoretical mass (M + H <sup>+</sup> )	Experimental mass (M + H <sup>+</sup> )
αT1	V <sub>1</sub> LSPADK <sub>7</sub>	729.41	729.22
δT1	V <sub>1</sub> HLPPEEK <sub>8</sub>	952.50	952.41
αT11	V <sub>93</sub> DPVNFK <sub>99</sub>	818.43	818.38
αT4	V <sub>17</sub> GAHAGEYGAEALER <sub>31</sub>	1529.73	1529.60
δT11	L <sub>96</sub> HVDPENFR <sub>104</sub>	1126.56	1126.42
δT14	E <sub>121</sub> FTPQMQAAYQK <sub>132</sub>	1441.67	1441.52
δT3	V <sub>18</sub> NVDAVGGEGALGR <sub>30</sub>	1256.65	1256.54
δT2	T <sub>9</sub> AVNALWVK <sub>17</sub>	959.52	959.49
δT15	δT15 with mass excess of 26 Da:	1149.67	1175.56
	V <sub>133</sub> VAGVANALA(Y)K <sub>144</sub>	1175.67 (ΔM = 26.00)	(ΔM = 25.89)
αT13	F <sub>128</sub> LASVSTVLTSK <sub>139</sub>	1252.71	1252.58
αT5	M <sub>32</sub> FLSFPTTK <sub>40</sub>	1071.55	1071.42
δT9	V <sub>67</sub> LGAFSDGLAHLNLK <sub>82</sub>	1669.88	1669.72
δT4	L <sub>31</sub> LVVYPWTQR <sub>40</sub>	1274.72	1274.56
δT5	F <sub>41</sub> FESFGDLSSPDVAVMGPNK <sub>59</sub>	2044.92	2044.60
αT9	V <sub>62</sub> ADALNAVAHVDDMPNALSALSDLHAHK <sub>90</sub>	2996.48	2996.38

Sequence coverage from MS/MS data for this mutant δ-globin chain was 68%.



**Fig. 4.** MS/MS fragment ion spectrum obtained from the 2<sup>+</sup> molecular ion at *m/z* 588.4 from the 12 amino acids mutated tryptic peptide δT15 (V<sub>133</sub>VAGVANALAYK<sub>144</sub>).

On account of this significant difference, it is concluded that the variant has markedly increased oxygen affinity.

#### 4. Discussion

Nearly 80 δ-globin variants have been characterized so far [6]. In this study we describe the characterization of a novel variant in which His is replaced by Tyr at or δ143 (H21). The first δ-variant with the substitution at 143 (H21) corresponds to that found in Hb A<sub>2</sub> Abruzzo or δ143 (H21) His → Arg, which was recently described in Italian patients [7] and is homologous to that of the β-chain, Hb Abruzzo [8–10]. Because Hb A<sub>2</sub> represents only a minor portion of the total haemoglobin, δ-variants are of little clinical importance. However, in β-thalassemia, Hb A<sub>2</sub> levels are often increased

and this characteristic increase is used as criterion for diagnosis of β-thalassemia. Hb Noah Mehmet Oeztuerk did not appear to give rise to appreciable changes in the haematological picture. Indeed all carriers displayed normal values except the propositus who was anaemic and required transfusion. However, we found no association between his anaemia and the δ143 His → Tyr mutation. The presence of this mutation results in decreased expression of Hb A<sub>2</sub> wild-type, and therefore, could lead to a misdiagnosis of β-thalassemia based on the level of Hb A<sub>2</sub>. Electrophoresis, DE-52 micro-chromatography, and CE-HPLC are routinely used for quantifying Hb A<sub>2</sub>. In both electrophoresis and DE-52 micro-chromatography the variant peak and the normal peak co-eluted as a single peak. Densitometric evaluation of Hb A<sub>2</sub> after electrophoresis was not an accurate and reliable method for quantifying Hb

A<sub>2</sub> as the sum of wild-type and the variant in samples from the carriers of the  $\delta 143$  His  $\rightarrow$  Tyr mutation. Indeed the Hb A<sub>2</sub> level was still low. Therefore, the detection of  $\beta$ -thalassemia trait may be compromised using this method. However, on DE-52 microchromatography the measurement of Hb A<sub>2</sub> as the sum of two  $\delta$ -peaks was accurate. This method could, therefore, be used for the diagnosis of  $\beta$ -thalassemia in carriers of the  $\delta 143$  His  $\rightarrow$  Tyr mutation. On routine CE-HPLC (e.g., Tosoh HLC-723 GHb VII) short  $\beta$ -thalassemia programs no abnormal peak was detected but Hb A<sub>2</sub> was still low, and therefore, the diagnosis of  $\beta$ -thalassemia based on the Hb A<sub>2</sub> level only will be compromised. Two peaks are detected by CE-HPLC using PolyCAT A column with extended gradient. In this case attention should be paid to the additional peak when the diagnosis of  $\beta$ -thalassemia is based only on the determination of Hb A<sub>2</sub>. However, the delta variant also induces a risk of missing  $\beta$ -thalassemia carrier ship when a diagnosis is based only on the retention time of the Hb A<sub>2</sub> peak. In general mutations which cause a low Hb A<sub>2</sub> could lead to misdiagnosis when inherited together with a  $\beta$ -thalassemia.

The replacement of His by Tyr at  $\delta 143$  (H21) has an equivalent on the  $\beta$ -globin gene, named Hb Old Dominion/Burton-upon-Trent [11]. This  $\beta$ -variant is associated with a slight increase in oxygen affinity but without hematologic effect. We did not carry out the functional studies on isolated Hb Noah Mehmet Oeztuerk, but molecular modelling of the Hb A<sub>2</sub> molecule (data not shown) with the  $\delta 143$  His  $\rightarrow$  Tyr mutation shows that the hydrogen bond between  $\delta_1 143$ Tyr and  $\delta_2 139$ Asn, which is present in the template 1S14 is conserved in Hb Noah Mehmet Oeztuerk. This hydrogen bond has been reported to stabilize the R conformation of the molecule, and therefore, increases its oxygen affinity [12]. Histidine at  $\beta 143$  (H21) is involved in the binding of diphospho-glycerate

(DPG) and five  $\beta$  Hb variants at this position have been reported having an increased oxygen affinity, due to reduced DPG binding [13]. Considering the close structural similarity of the  $\beta$ - and  $\delta$ -globins, it is expected that Hb Noah Mehmet Oeztuerk will have increased oxygen affinity. Indeed in the Hb A<sub>2</sub> fractions containing 50% of the variant the oxygen affinity was increased with a P<sub>50</sub> of  $20.6 \pm 0.60$  mmHg compared with  $29.3 \pm 0.40$  mmHg for the wild-type Hb A<sub>2</sub>. This significant difference suggests that Hb Noah Mehmet Oeztuerk has markedly increased oxygen affinity.

## References

- [1] H.F. Bunn, B.G. Forget, Hemoglobin: Molecular, Genetic and Clinical Aspects, W. B. Saunders, Philadelphia, 1986.
- [2] E. Bissé, H. Wieland, J. Chromatogr. 575 (1992) 223.
- [3] T.H.J. Huisman, in: T.H.J. Huisman (Ed.), Methods in Hematology, Churchill Livingstone, Edinburgh, 1986.
- [4] M. De Angioletti, G. Lacerra, C. Gaudiano, G. Mastrodonardo, L. Pagano, L. Mas-trullo, S. Masciandro, C. Carestia, Hum. Mut. 20 (2002) 358.
- [5] M. Siggaard-Andersen, O. Siggaard-Andersen, Acta Anaesthesiol. Scand. Suppl. 107 (1995) 13.
- [6] Globin Gene Server, Hbvar: a database of human haemoglobin variants and thalasse-mias. <http://globin.cse.psu.edu/cg-bin/hbvar/counter>, accessed November 2007.
- [7] M.J. Bouva, C.L. Hartevelde, P. van Delft, P.C. Giordano, Haematologica 91 (2006) 129.
- [8] L. Tentori, M. Carta Sorcini, C. Brecella, Clin. Chim. Acta 38 (1972) 258.
- [9] C. Bonaventura, J. Bonaventura, G. Amiconi, L. Tentori, M. Brunori, E. Antonini, J. Biol. Chem. 250 (1975) 6273.
- [10] A. Mosca, R. Paleari, F.M. Rubino, L. Zecca, G. De Bellis, S. Debernadi, F. Baudo, D. Cappellini, G. Fiorell, Hemoglobin 17 (1993) 261.
- [11] G.E. Elder, T.R. Lappin, A.B. Horne, V.F. Fairbanks, R.T. Jones, P.C. Winter, B.N. Green, J.D. Hoyer, T.M. Reynolds, D.T. Shih, D.J. McCormick, K.S. Kubik, B.J. Mad-den, C.G. Head, D. Harvey, N.B. Roberts, Mayo Clin. Proc. 73 (1998) 321.
- [12] M.F. Perutz, Nature New Biol. 243 (1973) 180.
- [13] T.H.J. Huisman, M.F.H. Carver, G.D. Efremov (Eds.), A Syllabus of Human Hemoglobin Variants, Sickle Cell Anemia Foundation, Augusta, GA, 1998.



## 4. Quantification de HbA<sub>2</sub> et détection des mutations les plus courantes tels que Hb C, E, S et A2' par une approche protéomique

Ce travail a été réalisé dans le cadre d'une collaboration avec le Dr Emmanuel Bissé de l'Université de Fribourg (Allemagne).

### 4.1. Contexte de l'étude

La méthode basée sur la spectrométrie de masse, développée pour des échantillons certifiés contenant un pourcentage précis de HbA<sub>2</sub> (§ 2), pourrait aussi être utilisée directement pour analyser le sang des patients. Bien sûr, ceci n'est pas fait puisque des méthodes biochimiques classiques, bien plus simples et bon marché, permettent de le faire dans les laboratoires d'analyse biologique. Il existe cependant des cas, où la méthode développée pour les standards de référence serait très utile. En effet, certains cas de pathologies pour lesquelles les techniques d'analyse classique des laboratoires de biologie (HPLC-SCX et électrophorèse) génèrent des chromatogrammes aberrants. Des déplacements de pics sont observés et il est alors impossible de donner le taux exact de HbA<sub>2</sub> [Higgins et al. 2009]. C'est dans ces cas que nous proposons de réaliser le dosage de HbA<sub>2</sub> en utilisant la méthode décrite dans le § 2. Les données de LC-MS permettront d'identifier les peptides normaux, et les peptides anormaux (masse moléculaire non attendue) seront alors identifiables par MS/MS. Ces données de MS/MS seront alors exploitées avec tout l'arsenal de l'analyse protéomique que nous avons largement décrit dans les chapitres 1 et 2. Cette approche permettrait non seulement d'obtenir un dosage de HbA<sub>2</sub> utile pour le diagnostic de la thalassémie mais les mutations éventuelles pourraient également être détectées dans la même expérience.

L'objectif de cette étude est de donc déterminer, à partir d'une seule méthode, le taux de HbA<sub>2</sub>, tout en caractérisant les variants les plus courants, répertoriés dans les banques, et même les mutations nouvelles en utilisant l'approche *de novo* en mode automatique.

Pour démontrer la faisabilité de cette approche, nous avons donc analysé une série d'échantillons contenant des variants connus en réalisant une analyse LC-MS permettant de quantifier HbA<sub>2</sub> suivie d'une LC-MS/MS complémentaire pour identifier les variants les plus courants (HbS, HbC, HbE et HbA<sub>2</sub>). Les résultats ont ensuite été comparés à ceux obtenus par des techniques classiques d'HPLC et d'électrophorèse. Bien sûr, il ne s'agit là que d'une étape préliminaire et de nombreuses mises au point seront nécessaires avant l'automatisation complète de ce type d'analyse. Ces mises au point pourront s'appuyer sur d'autres travaux de screening de mutations par spectrométrie de masse. Par exemple, Basilico et al [Basilico et al. 2007] ont développé une méthode de recherche des variants connus de l'hémoglobine par microLC-MS/MS des peptides issus d'une digestion trypsique et

recherche dans une banque de peptides mutés. Une autre approche décrit l'utilisation en mode MRM (multiple reaction monitoring) ciblé sur les transitions spécifiques des variants [Daniel et al. 2005].

## 4.2. Résultats

Les mutations en acides aminés correspondants aux différents variants recherchés sont présentées dans le Tableau 9.

Variant	Globine	Peptide non muté	Masse du peptide sauvage (M+H) <sup>+</sup>	Peptide muté	Masse du peptide muté (M+H) <sup>+</sup>
HbS	β	<sup>1</sup> VHLTP <b>EE</b> K <sup>8</sup>	952,5	<sup>1</sup> VHLTP <b>VE</b> K <sup>8</sup>	922,5
HbC	β	<sup>1</sup> VHLTP <b>EE</b> K <sup>8</sup>	952,5	<sup>1</sup> VHLTP <b>K</b> <sup>6</sup>	694,4
HbE	β	<sup>18</sup> VNVDEVGG <b>E</b> ALGR <sup>30</sup>	1314,6	<sup>18</sup> VNVDEVGG <b>K</b> <sup>26</sup>	916,5
HbA <sub>2</sub> '	δ	<sup>9</sup> TAVNALW <b>G</b> K <sup>17</sup>	959,5	<sup>9</sup> TAVNALW <b>R</b> <sup>16</sup>	930,5

**Tableau 9.** Les variants les plus courants de l'hémoglobine et les peptides tryptiques qui portent la mutation

Une première analyse par LC-MS de l'échantillon permet de doser HbA<sub>2</sub> en utilisant la méthode décrite précédemment (§ 2). Dans le cas des échantillons contenant le variant HbA<sub>2</sub>', la quantification doit être effectuée en utilisant le peptide δT14 car le peptide δT2 est muté.

Afin de déterminer la présence d'un des quatre variants présentés dans le Tableau 9, les ions correspondant aux peptides mutés sont recherchés dans l'analyse LC-MS. Lorsqu'un échantillon présente un ou plusieurs de ces ions, une analyse LC-MS/MS est réalisée afin de séquencer le peptide d'intérêt et ainsi de caractériser la mutation.

Plusieurs échantillons ont été soumis à ces expériences afin d'évaluer la méthode. Pour cela, nous avons comparé nos résultats à ceux obtenus par des méthodes classiques d'HPLC échangeuse d'ions (SCX) et d'électrophorèse capillaire (CE). Les résultats sont présentés dans le Tableau 10.

N°échantillon	% HbA <sub>2</sub>			Variant
	HPLC-SCX	CE	LC-MS	Diagnostic MS
1	1,8	1,8	1,7	Normal
2	2,2	2,4	2,1	Normal
3	2,0	2,1	2,1	Normal
4	2,1	2,4	2,1	Normal
5	2,0	2,2	2,3	Normal
6	2,0	2,3	2,2	Normal
7	1,8	2,2	2,0	Normal
8	2,1	2,2	2,0	Normal
9	2,0	2,6	2,4	Normal
10	2,0	2,0	2,1	Normal
11	3,3	2,8	3,2	HbS
12	1,8	1,8	2,4	HbS + HbA <sub>2</sub>
13	3,7	non mesurable	3,1	HbE
14	3,9	3,3	<b>3,6</b>	β-Thal
15	5,0	5,5	<b>5,2</b>	β-Thal
16	4,8	non mesurable	<b>6,2</b>	β-Thal + HbE
17	5,0	4,4	<b>6,8</b>	β-Thal
18	3,1	3,0	<b>3,8</b>	β-Thal + HbS
19	5,5	4,7	<b>4,8</b>	β-Thal + HbS
20	0,9	non mesurable	1,1	HbC

**Tableau 10.** Résultats obtenus en combinant la méthode de dosage de HbA<sub>2</sub> par LC-MS et la recherche et la caractérisation des variants les plus courants de l'hémoglobine par LC-MS/MS; comparaison avec les résultats obtenus par les méthodes classiques de quantification de HbA<sub>2</sub> : HPLC-SCX et électrophorèse capillaire

Le dosage de HbA<sub>2</sub> est basé sur la quantification de deux peptides issus de la digestion trypsique de la chaîne de globine delta ( $\delta T2$  et  $\delta T14$ ). Les études réalisées dans la première partie de ce chapitre ont montré que l'on obtenait des coefficients de variation plus faible en utilisant le peptide  $\delta T2$ . Ce peptide a donc été utilisé pour les dosages présentés dans le Tableau 10, à l'exception de l'échantillon 12 pour lequel le peptide  $\delta T2$  porte une mutation. Plusieurs échantillons présentent un taux de HbA<sub>2</sub> supérieur à 3,3% qui est donc diagnostique de porteur de beta-thalassémie.

De plus, la méthode a permis d'identifier les différents variants recherchés et de les caractériser sans ambiguïté. Ainsi, par exemple pour l'échantillon 12, nous avons mis en évidence deux mutations, une portée par la chaîne beta (HbS) et l'autre par la chaîne delta (HbA<sub>2</sub>).

Un autre cas intéressant est l'échantillon 16 dont les résultats montrent un taux élevé de HbA<sub>2</sub> indicateur de  $\beta$ -thalassémie et la présence d'une mutation sur la chaîne beta (HbE). En effet, l'hémoglobine E est parfois désignée comme étant une « hémoglobinopathie thalassémique ». En fait, l'association des 2 mutations (HbE/ $\beta$ -thalassémie) est l'une des plus importantes causes de thalassémie cliniquement grave de par le monde.

Cette approche pourrait donc être particulièrement intéressante pour le diagnostic des patients, porteurs simultanément de  $\beta$ -thalassémie et d'une mutation de la chaîne beta (les plus courantes sont HbS, HbE et HbC), ce qui peut provoquer l'apparition d'une maladie significative sur le plan clinique.

### **4.3. Conclusion et perspectives**

Dans ce travail exploratoire, nous avons combiné la méthode développée pour la quantification de HbA<sub>2</sub> avec une LC-MS/MS. Cette approche pourrait permettre non seulement de mesurer le taux de HbA<sub>2</sub> dans les cas où des mutations rendraient les dosages non réalisables par les méthodes biochimiques classiques, mais également permettre d'identifier d'éventuelles mutations. Pour un des hémolysats, nous avons même pu identifier deux mutations simultanées chez le même patient.

Il est probable que les développements instrumentaux permettront de réaliser de telles expériences en routine sur des instruments robustes et fiables. En effet, les dernières évolutions en spectrométrie de masse (automatisation de la MS/MS, interprétation automatisée des données, nouvelle génération de triple quadripôle qui permet de cibler des centaines de transition en MRM) nous font penser qu'il serait possible de mettre sur pied une méthode d'analyse, pouvant en moins d'une demi-heure, diagnostiquer soit la présence des mutations les plus courantes (par une approche MRM ciblée), soit une approche plus générale, un peu plus difficile à mettre au point, basée sur la LC-MS/MS qui soit capable de détecter également de nouvelles mutations. De nombreux hôpitaux sont maintenant équipés de LC-MS/MS de type triple quadripôle, ce qui rendrait possible la réalisation systématique de telles analyses pour tous les nouveaux nés.

## 5. Conclusion

Au cours de ce travail, les méthodes de chromatographie liquide couplée à la spectrométrie de masse ont permis de quantifier l'hémoglobine A<sub>2</sub> et de caractériser différentes formes d'hémoglobine mutée.

Nous avons développé une méthode basée sur une digestion enzymatique suivie d'une analyse LC-MS pour la quantification de HbA<sub>2</sub> avec un coefficient de variation proche de 5%, qui devra être encore améliorée grâce à plusieurs optimisations pour être inférieur à 5% et d'améliorer la précision pour de grandes séries d'échantillons.

Nous avons également mis en place une stratégie pour caractériser de nouvelles mutations de l'hémoglobine, permettant ainsi de lever les ambiguïtés qui pouvaient être soulevées par les méthodes classiquement utilisées. Deux nouvelles mutations ont ainsi été identifiées.

Enfin, nous avons montré sur une série d'échantillons de patients sains et atteints de différentes mutations, qu'il était possible à la fois de quantifier HbA<sub>2</sub> et de caractériser deux mutations d'hémoglobine chez le même patient.

Nous pensons que les progrès instrumentaux rendront ce type d'analyse routinière dans les hôpitaux pour des coûts très raisonnables.



# **BIBLIOGRAPHIE**



- Alexander, D. L., Mital, J., Ward, G. E., Bradley, P. et Boothroyd, J. C. (2005). "Identification of the moving junction complex of *Toxoplasma gondii*: a collaboration between distinct secretory organelles." *PLoS Pathog* 1(2): e17.
- Balgley, B. M., Laudeman, T., Yang, L., Song, T. et Lee, C. S. (2007). "Comparative Evaluation of Tandem MS Search Algorithms Using a Target-Decoy Search Strategy." *Molecular & Cellular Proteomics* 6(9): 1599-1608.
- Basilico, F., Di Silvestre, D., Sedini, S., Petretto, A., Levreri, I., Melioli, G., Farina, C., Mori, F. et Mauri, P. L. (2007). "New approach for rapid detection of known hemoglobin variants using LC-MS/MS combined with a peptide database." *J Mass Spectrom* 42(3): 288-292.
- Bednarczyk, A. (2008). "Nouvelles méthodologies en protéomique pour une caractérisation fine des protéines." Université de Strasbourg.
- Brissonnet, F. et Maujean, A. (1993). "Characterization of Foaming Proteins in a Champagne Base Wine." *Am. J. Enol. Vitic.* 44(3): 297-301.
- Carapito, C. (2006). "Vers une meilleure utilisation des données de spectrométrie de masse en analyse protéomique." Université Louis Pasteur de Strasbourg.
- Carrascal, M., Gay, M., Ovelleiro, D., Casas, V., Gelpi, E. et Abian, J. (2010). "Characterization of the human plasma phosphoproteome using linear ion trap mass spectrometry and multiple search engines." *J Proteome Res* 9(2): 876-884.
- Carruthers, V. et Boothroyd, J. C. (2007). "Pulling together: an integrated model of *Toxoplasma* cell invasion." *Curr Opin Microbiol* 10(1): 83-89.
- Caruso, D., Crestani, M., Mitro, N., Da Riva, L., Mozzi, R., Sarpau, S., Merlotti, C. et Franzini, C. (2005). "High pressure liquid chromatography and electrospray ionization mass spectrometry are advantageously integrated into a two-levels approach to detection and identification of haemoglobin variants." *Clin Lab Haematol* 27(2): 111-119.
- Cilindre, C., Castro, A. J., Clément, C., Jeandet, P. et Marchal, R. (2007). "Influence of *Botrytis cinerea* infection on Champagne wine proteins (characterized by two-dimensional electrophoresis/immunodetection) and wine foaming properties." *Food Chemistry* 103(1): 139-149.
- Cilindre, C., Jegou, S., Hovasse, A., Schaeffer, C., Castro, A. J., Clement, C., Van Dorsselaer, A., Jeandet, P. et Marchal, R. (2008). "Proteomic approach to identify champagne wine proteins as modified by *Botrytis cinerea* infection." *J Proteome Res* 7(3): 1199-1208.
- Clarke, G. M. et Higgins, T. N. (2000). "Laboratory investigation of hemoglobinopathies and thalassemias: review and update." *Clin Chem* 46(8 Pt 2): 1284-1290.
- Daher, W. et Soldati-Favre, D. (2009). "Mechanisms controlling glideosome function in *apicomplexans*." *Current Opinion in Microbiology* 12(4): 408-414.
- Dambrouck, T. (2004). "Contribution à l'étude des protéines des mouûts et des vins de Champagne: Détermination de leur origine; Etude d'une glycoprotéine majeure, l'invertase de raisin." Université de Reims.
- Daniel, Y. A., Turner, C., Haynes, R. M., Hunt, B. J. et Dalton, R. N. (2005). "Rapid and specific detection of clinically significant haemoglobinopathies using electrospray mass spectrometry-mass spectrometry." *Br J Haematol* 130(4): 635-643.
- Deutsch, E. W., Lam, H. et Aebersold, R. (2008). "Data analysis and bioinformatics tools for tandem mass spectrometry in proteomics." *Physiol Genomics* 33(1): 18-25.

Dubey, J. P., Lindsay, D. S. et Speer, C. A. (1998). "Structures of *Toxoplasma gondii* tachyzoites, bradyzoites, and sporozoites and biology and development of tissue cysts." *Clin Microbiol Rev* 11(2): 267-299.

Dzierszinski, F., Mortuaire, M., Dendouga, N., Popescu, O. et Tomavo, S. (2001). "Differential expression of two plant-like enolases with distinct enzymatic and antigenic properties during stage conversion of the protozoan parasite *Toxoplasma gondii*." *Journal of Molecular Biology* 309(5): 1017-1027.

Elias, J. E. et Gygi, S. P. (2007). "Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry." *Nat Meth* 4(3): 207-214.

Ferreira, R. B., Piçarra-Pereira, M. A., Monteiro, S., Loureiro, V. B. et Teixeira, A. R. (2001). "The wine proteins." *Trends in Food Science & Technology* 12(7): 230-239.

Gajria, B., Bahl, A., Brestelli, J., Dommer, J., Fischer, S., Gao, X., Heiges, M., Iodice, J., Kissinger, J. C., Mackey, A. J., Pinney, D. F., Roos, D. S., Stoeckert, C. J., Jr., Wang, H. et Brunk, B. P. (2007). "ToxoDB: an integrated *Toxoplasma gondii* database resource." *Nucl. Acids Res.*: gkm981.

Gallien, S. (2009). "Nouvelles méthodologies protéomiques d'aide à l'annotation des génomes et à la validation des séquences protéiques." Université de Strasbourg.

Gaskins, E., Gilk, S., DeVore, N., Mann, T., Ward, G. et Beckers, C. (2004). "Identification of the membrane receptor of a class XIV myosin in *Toxoplasma gondii*." *The Journal of Cell Biology* 165(3): 383-393.

Gissot, M., Kim, K., Schaap, D. et Ajioka, J. W. (2009). "New eukaryotic systematics: a phylogenetic perspective of developmental gene expression in the Apicomplexa." *Int J Parasitol* 39(2): 145-151.

Gonzalez, J., Takao, T., Hori, H., Besada, V., Rodriguez, R., Padron, G. et Shimonishi, Y. (1992). "A method for determination of N-glycosylation sites in glycoproteins by collision-induced dissociation analysis in fast atom bombardment mass spectrometry: Identification of the positions of carbohydrate-linked asparagine in recombinant [alpha]-amylase by treatment with peptide-N-glycosidase F in 18O-labeled water." *Analytical Biochemistry* 205(1): 151-158.

Handman, E., Goding, J. W. et Remington, J. S. (1980). "Detection and characterization of membrane antigens of *Toxoplasma gondii*." *J Immunol* 124(6): 2578-2583.

Higgins, T. N., Khajuria, A. et Mack, M. (2009). "Quantification of HbA2 in Patients With and Without Beta-Thalassemia and in the Presence of HbS, HbC, HbE, and HbD Punjab Hemoglobin Variants." *American Journal of Clinical Pathology* 131(3): 357-362.

Jones, A., R., Siepen, J., A., Hubbard, S., J. et Paton, N., W. (2009). "Improving sensitivity in proteome studies by analysis of false discovery rates for multiple search engines." *PROTEOMICS* 9(5): 1220-1229.

Joutovsky, A., Hadzi-Nesic, J. et Nardi, M. A. (2004). "HPLC Retention Time as a Diagnostic Tool for Hemoglobin Variants and Hemoglobinopathies: A Study of 60000 Samples in a Clinical Diagnostic Laboratory" *Clinical Chemistry* 50: 1736-1747.

Kaji, H., Saito, H., Yamauchi, Y., Shinkawa, T., Taoka, M., Hirabayashi, J., Kasai, K., Takahashi, N. et Isobe, T. (2003). "Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify N-linked glycoproteins." *Nat Biotechnol* 21(6): 667-672.

Kapp, E., A., Schütz, F., Connolly, L., M., Chakel, J., A., Meza, J., E., Miller, C., A., Fenyo, D., Eng, J., K., Adkins, J., N., Omenn, G., S. et Simpson, R., J. (2005). "An evaluation, comparison, and accurate benchmarking of several publicly available MS/MS search algorithms: Sensitivity and specificity analysis." *PROTEOMICS* 5(13): 3475-3490.

- Keeley, A. et Soldati, D. (2004). "The glideosome: a molecular machine powering motility and host-cell invasion by *Apicomplexa*." *Trends Cell Biol* 14(10): 528-532.
- Kibe, M. K., Coppin, A., Dendouga, N., Oria, G., Meurice, E., Mortuaire, M., Madec, E. et Tomavo, S. (2005). "Transcriptional regulation of two stage-specifically expressed genes in the protozoan parasite *Toxoplasma gondii*." *Nucleic Acids Res* 33(5): 1722-1736.
- Kim, K. et Weiss, L. M. (2004). "*Toxoplasma gondii*: the model apicomplexan." *International Journal for Parasitology* 34(3): 423-432.
- Kleinert, P., Schmid, M., Zurbriggen, K., Speer, O., Schmutz, M., Roschitzki, B., Durka, S. S., Leopold, U., Kuster, T., Heizmann, C. W., Frischknecht, H. et Troxler, H. (2008). "Mass Spectrometry: A Tool for Enhanced Detection of Hemoglobin Variants." *Clin Chem* 54(1): 69-76.
- Küster, B. et Mann, M. (1999). "18O-Labeling of N-Glycosylation Sites To Improve the Identification of Gel-Separated Glycoproteins Using Peptide Mass Mapping and Database Searching." *Analytical Chemistry* 71(7): 1431-1440.
- Kuzuhara, T. et Horikoshi, M. (2004). "A nuclear FK506-binding protein is a histone chaperone regulating rDNA silencing." *Nat Struct Mol Biol* 11(3): 275-283.
- Kwon, S. W. (2004). "Profiling of Soluble Proteins in Wine by Nano-High-Performance Liquid Chromatography/Tandem Mass Spectrometry." *Journal of Agricultural and Food Chemistry* 52(24): 7258-7263.
- Li, L., Brunk, B. P., Kissinger, J. C., Pape, D., Tang, K., Cole, R. H., Martin, J., Wylie, T., Dante, M., Fogarty, S. J., Howe, D. K., Liberator, P., Diaz, C., Anderson, J., White, M., Jerome, M. E., Johnson, E. A., Radke, J. A., Stoeckert, C. J., Jr., Waterston, R. H., Clifton, S. W., Roos, D. S. et Sibley, L. D. (2003). "Gene discovery in the apicomplexa as revealed by EST sequencing and assembly of a comparative gene database." *Genome Res* 13(3): 443-454.
- Marchal, R., Seguin, V. et Maujean, A. (1997). "Quantification of Interferences in the Direct Measurement of Proteins in Wines From the Champagne Region Using the Bradford Method." *Am. J. Enol. Vitic.* 48(3): 303-309.
- Marchal, R., Tabary, I., Valade, M., Moncomble, D., Viaux, L., Robillard, B. et Jeandet, P. (2001). "Effects of *Botrytis cinerea* infection on Champagne wine foaming properties." *Journal of the Science of Food and Agriculture* 81(14): 1371-1378.
- Nesvizhskii, A. I., Vitek, O. et Aebersold, R. (2007). "Analysis and validation of proteomic data generated by tandem mass spectrometry." *Nat Meth* 4(10): 787-797.
- Odenthal-Schnittler, M., Tomavo, S., Becker, D., Dubremetz, J. F. et Schwarz, R. T. (1993). "Evidence for N-linked glycosylation in *Toxoplasma gondii*." *Biochem J* 291 ( Pt 3): 713-721.
- Okuda, T., Fukui, M., Takayanagi, T. et Yokotsuka, K. (2006). "Characterization of Major Stable Proteins in Chardonnay Wine." *Food Science and Technology Research* 12(2): 131-136.
- Rai, D. K., Green, B. N., Landin, B., Alvelius, G. et Griffiths, W. J. (2004). "Accurate mass measurement and tandem mass spectrometry of intact globin chains identify the low proportion variant hemoglobin Lepore-Boston-Washington from the blood of a heterozygote." *Journal of Mass Spectrometry* 39(3): 289-294.
- Rai, D. K., Landin, B., Alvelius, G. et Griffiths, W. J. (2002). "Electrospray Tandem Mass Spectrometry of Intact  $\beta^2$ -Chain Hemoglobin Variants." *Analytical Chemistry* 74(9): 2097-2102.
- Robinson, N. E. et Robinson, A. B. (2001). "Deamidation of human proteins." *Proc Natl Acad Sci U S A* 98(22): 12409-12413.

Searle, B. C., Turner, M. et Nesvizhskii, A. I. (2008). "Improving Sensitivity by Probabilistically Combining Results from Multiple MS/MS Search Methodologies." *Journal of Proteome Research* 7(1): 245-253.

Segu, Z. M., Hussein, A., Novotny, M. V. et Mechref, Y. (2010). "Assigning N-Glycosylation Sites of Glycoproteins using LC/MSMS in Conjunction with Endo-M/Exoglycosidase." *J Proteome Res.*

Serjeant, G. et Serjeant, B. (2001). "Sickle Cell Disease." 3rd edn. Oxford University Press: Oxford.

Stephens, A. D. (2005). "Haemoglobinopathies : Screening and diagnosis." *BIOMEDICAL SCIENTIST* February 2005.

Tenter, A. M., Heckerth, A. R. et Weiss, L. M. (2000). "Toxoplasma gondii: from animals to humans." *Int J Parasitol* 30(12-13): 1217-1258.

Wagner-Rousset, E., Bednarczyk, A., Bussat, M. C., Colas, O., Corvaia, N., Schaeffer, C., Van Dorselaer, A. et Beck, A. (2008). "The way forward, enhanced characterization of therapeutic antibody glycosylation: Comparison of three level mass spectrometry-based strategies." *J Chromatogr B Analyt Technol Biomed Life Sci* 872(1-2): 23-37.

Wajcman, H., Prehu, C., Bardakjian-Michau, J., Prome, D., Riou, J., Godart, C., Mathis, M., Hurtrel, D. et Galacteros, F. (2001). "Abnormal hemoglobins: laboratory methods." *Hemoglobin* 25(2): 169-181.

Waters, E. J., Hayasaka, Y., Tattersall, D. B., Adams, K. S. et Williams, P. J. (1998). "Sequence Analysis of Grape (*Vitis vinifera*) Berry Chitinases That Cause Haze Formation in Wines." *Journal of Agricultural and Food Chemistry* 46(12): 4950-4957.

Weatherall, D. et Clegg, J. (2001). "The thalassaemia syndromes." 4th edn. Blackwell Scientific: Oxford.

Wild, B. J. et Bain, B. J. (2004). "Detection and quantitation of normal and variant haemoglobins: an analytical review." *Ann Clin Biochem* 41: 355-369

Wild, B. J., Green, B. N., Cooper, E. K., Lalloz, M. R., Erten, S., Stephens, A. D. et Layton, D. M. (2001). "Rapid identification of hemoglobin variants by electrospray ionization mass spectrometry." *Blood Cells Mol Dis* 27(3): 691-704.

Zanella-Cleon, I., Joly, P., Becchi, M. et Francina, A. (2009). "Phenotype determination of hemoglobinopathies by mass spectrometry." *Clin Biochem* 42(18): 1807-1817.

Zhou, X. W., Kafsack, B. F., Cole, R. N., Beckett, P., Shen, R. F. et Carruthers, V. B. (2005). "The opportunistic pathogen *Toxoplasma gondii* deploys a diverse legion of invasion and survival proteins." *J Biol Chem* 280(40): 34233-34244.

# **CONCLUSION GENERALE**





## CONCLUSION GENERALE

L'objectif de mon travail de thèse était de développer des méthodologies d'analyse protéomique qui permettent une caractérisation très fine des N-glycosylations et de leur micro-hétérogénéité tout en permettant l'identification d'un maximum de protéines dans des mélanges très complexes, y compris d'éventuelles isoformes. De plus, ces méthodologies doivent être applicables particulièrement dans les cas où les échantillons ne sont disponibles qu'en très petites quantités et ne pouvant être fournis qu'une seule fois par les biologistes pour diverses raisons (préparation des extraits très longues, sources pathologiques rares, etc...). Ce dernier point est essentiel puisque, lorsque l'unique échantillon disponible est à peine suffisant pour pouvoir réaliser une seule expérience de LC-MS/MS, il faut disposer de paramètres instrumentaux qui donnent les meilleures chances d'obtenir des données utilisables. C'est là un aspect essentiel de notre travail puisque l'objectif de notre laboratoire est de réaliser des analyses protéomiques dans des collaborations où les échantillons ne pourront être que difficilement renouvelés et/ou qui représentent parfois des mois de travail de la part des biologistes.

Après avoir rappelé dans notre introduction bibliographique les principales méthodes couramment utilisées pour la caractérisation des protéines glycosylées, nous avons montré les directions dans lesquelles des progrès devaient être faits pour améliorer les méthodologies protéomiques et en particulier la description des micro-hétérogénéités de N-glycosylations dans le cas d'échantillons en faibles quantités et/ou uniques.

Nos résultats ont été présentés en deux parties. La première partie (Chapitres 1 et 2) présente les différents développements instrumentaux et méthodologiques réalisés principalement pour permettre une meilleure description des N-glycosylations. Dans la seconde partie (Chapitres 3, 4 et 5) nous avons ensuite décrit les résultats obtenus dans le cadre de trois collaborations avec des biologistes en utilisant les méthodologies et stratégies mises en place dans la première partie.

Dans la partie consacrée au développement de méthodes pour maîtriser les performances du couplage nanoLC-MS/MS (Chapitre 1), aussi bien au niveau chromatographique que du spectromètre de masse, nous avons mis au point une série de méthodes qui permettent non seulement de vérifier les performances de l'ensemble de l'instrument avant chaque série d'analyses, mais également d'aider au diagnostic en cas de panne du système. Ces tests sont basés sur l'analyse nanoLC-MS et MS/MS d'un mélange de peptides issus de la digestion trypsique de la BSA qui permettent de mesurer différents paramètres pour vérifier et optimiser l'état du système. Chacun de ces tests permet de vérifier un point précis du système et ainsi de s'assurer des performances de l'instrument avant chaque série d'analyses.

Dans le chapitre 2 de la partie développement méthodologique, nous avons focalisé notre attention sur la caractérisation de la micro-hétérogénéité des protéines N-glycosylées. La stratégie utilisée est basée sur l'analyse de glycopeptides et l'utilisation des ions diagnostiques issus de la fragmentation spécifique des N-glycanes pour obtenir la distribution des différentes glycoformes portées par un même peptide. Nous avons montré que le rôle central que peuvent jouer les ions diagnostiques pour l'étude des glycopeptides repose sur les possibilités instrumentales. En effet, selon les spectromètres de masse utilisés, les ions diagnostiques ne sont détectables avec une intensité suffisante que s'ils sont générés soit par CID dans l'interface, soit par CID dans la cellule de collision. Par conséquent, selon le cas, il faudra utiliser différemment les possibilités de fragmentation de l'instrument. La meilleure approche consisterait à réaliser une expérience de couplage en mode alterné où un scan serait réalisé à haute énergie (soit dans l'interface, soit dans la cellule de collision) et le second serait réalisé en mode MS/MS. On disposerait ainsi simultanément du chromatogramme des ions diagnostiques de la présence de sucres, et de la possibilité d'obtenir les spectres MS/MS correspondant à leur temps d'élution. En fait, cette approche s'est révélée difficile à mettre en pratique puisque les instruments actuels ne permettent pas de réaliser des balayages assez rapides pour que les données LC-MS (ions diagnostiques) et LC-MS/MS (fragments pour l'identification des peptides et des glycopeptides) soient assez intenses et d'assez bonne qualité. En attendant une prochaine génération de spectromètres de masse à balayage plus rapide, nous avons donc dissocié les deux opérations en réalisant des injections consécutives en LC-MS puis en LC-MS/MS. A côté de ces verrous technologiques, qui concernent les capacités de balayage des spectromètres de masse, il reste un verrou au niveau de l'interprétation automatisée des données MS/MS pour en déduire la structure des glycanes et la séquence de la chaîne peptidique.

Les méthodologies développées dans la première partie de notre travail se sont révélées utiles lors de 3 collaborations avec des biologistes.

Dans le chapitre 3, consacré à une étude protéomique réalisée sur le parasite *Toxoplasma gondii*, nous avons étudié les protéines impliquées dans l'invasion de la cellule hôte par le parasite et nous avons prouvé que des glycoprotéines avaient un rôle important dans ce mécanisme. L'une de ces glycoprotéines, GAP50, a ensuite été caractérisée plus en détail. En appliquant notre méthodologie, nous avons identifié les trois sites de glycosylations de la protéine, ainsi que les 4 glycoformes présentes au niveau de chacun de ces sites. En parallèle, une étude protéomique a été menée pour rechercher les facteurs de transcription impliqués dans l'interconversion de la forme virulente tachyzoïte en forme dormante bradyzoïte du parasite. Parmi les 39 facteurs de transcription identifiés, la protéine que nous avons nommé TgNF3, a pu être étudiée de manière plus approfondie et nous avons pu montrer son rôle dans la virulence du parasite et sa capacité à s'enkyster.

Dans le chapitre 4, nous avons réalisé la première étude protéomique des vins de Champagne en effectuant une analyse différentielle par gel d'électrophorèse bidimensionnel entre un vin issu de raisins contaminés par le champignon *Botrytis cinerea* et un vin sain. Nous avons ainsi mis en évidence de nombreuses protéines différentiellement exprimées en fonction de la contamination. L'une d'entre elles, l'invertase vacuolaire de raisin, qui possède un rôle majeur dans les propriétés

moussantes du vin, a ensuite été caractérisée plus finement. Cette glycoprotéine possède 12 sites potentiels de glycosylation. Les différentes expériences ont permis d'identifier 5 de ces sites de glycosylation et de caractériser 11 glycoformes différentes sur l'un de ces sites.

Enfin, dans le chapitre 5, nous avons développé une méthode de dosage de la chaîne delta de l'hémoglobine permettant de quantifier HbA<sub>2</sub> avec un coefficient de variation proche de 5% pour des concentrations de HbA<sub>2</sub> comprises entre 0 et 8%. Cette méthode de référence doit permettre de certifier des échantillons standards pour étalonner les instruments couramment utilisés pour quantifier l'hémoglobine HbA<sub>2</sub> dans les laboratoires d'hématologie. Nous avons également développé une méthodologie permettant de caractériser de nouveaux variants lorsque les méthodes utilisées classiquement sont inefficaces. Deux nouvelles mutations de l'hémoglobine, dont une mutation mineure sur la chaîne delta, ont ainsi été identifiées. Enfin, nous avons élargi la méthode de dosage de HbA<sub>2</sub> à l'identification de variants les plus courants de l'hémoglobine en ajoutant une étape de LC-MS/MS.

Notre travail de thèse illustre donc la nécessité de faire encore progresser les méthodes d'analyse protéomique. Celles-ci peuvent encore, à notre avis, être considérablement développées, tant du point de vue de la sensibilité des instruments, que de leur amélioration pour les modes de balayage. De plus, l'interprétation bioinformatique des données de spectrométrie de masse devrait pouvoir faire progresser considérablement la caractérisation des glycopeptides, et donc permettre de déterminer les profils de micro-hétérogénéité de façon plus routinières et ainsi faire progresser nos connaissances sur la relation structure-activité biologique des glycoprotéines.



# **PARTIE EXPERIMENTALE**





# PARTIE EXPERIMENTALE

## 1. Partie 1 des résultats

### 1.1. Chapitre 1

#### 1.1.1. Optimisation des conditions chromatographiques en utilisant le méthanol comme solvant organique sur un système nanoLC-Chip/MS

Afin de vérifier les performances du système nanoLC-MS/MS sur un système Agilent 1200 Series HPLC-Chip/MS (Agilent Technologies, Palo Alto, USA) couplé à un spectromètre de masse HCT Ultra ion trap (Bruker Daltonics, Brême, Allemagne), 50 fmol d'un digestat tryptique de BSA sont analysés par nanoLC-MS et nanoLC-MS/MS.

Les voies A de la pompe nano et de la pompe de chargement délivrent un mélange composé de 98% H<sub>2</sub>O, 2% MeOH et 0,1% HCOOH. La voie B de la pompe nano délivre un mélange composé de 98% MeOH, 2% H<sub>2</sub>O, et 0.1% HCOOH.

Le mélange peptidique est chargé sur la pré-colonne d'enrichissement Zorbax 300SB-C18 (volume 40 nL, particules de 5 µm, porosité 300 Å) à 3,75 µL/min. Après dessalage, les peptides sont élués à un débit de 300 nL/min au travers de la colonne analytique Zorbax 300SB-C18.

Pour les puces microfluidiques contenant une colonne Zorbax 300SB-C18 (75 µm de diamètre interne x 150 mm de longueur, particules de 5 µm, porosité 300 Å), le gradient utilisé est le suivant : 10 à 30 % de B en 5 min puis 30 à 60% de B en 15 min, suivi d'un palier de 5 min à 85% de B et d'un rééquilibrage de la colonne à 10% de B pendant 15 min.

Pour les puces microfluidiques contenant une colonne Zorbax 300SB-C18 (75 µm de diamètre interne x 43 mm de longueur, particules de 5 µm, porosité 300 Å), le gradient utilisé est le suivant : 10 à 30 % de B en 1,5 min puis 30 à 60% de B en 7,5 min, suivi d'un palier de 5 min à 85% de B et d'un rééquilibrage de la colonne à 10% de B pendant 5 min.

Le spectromètre de masse HCT Ultra ion trap est étalonné avec des composés de référence (Tuning Mix ES (Agilent Technologies)). Les paramètres généraux du spectromètre de masse sont les suivants : tension capillaire, -1750V; gaz de séchage, 4 L/min; température de séchage, 325°C. Le système opère avec un passage automatique entre les modes MS et MS/MS. L'analyse MS est réalisée dans le mode de résolution "standard enhanced" à une vitesse de balayage de 8100 m/z par seconde avec un contrôle de la charge d'ions à 100000 dans un temps de remplissage maximale de 200 ms. Un total de 4 spectres MS est moyenné pour obtenir les spectres MS finaux. Les 3 ions les plus intenses, préférentiellement doublement ou triplement chargés sont sélectionnés dans chaque spectre MS pour être isolés et fragmentés. L'analyse MS/MS est réalisée dans le mode de résolution "ultrascan" à une vitesse de balayage de 26000 m/z par seconde avec un contrôle de la charge d'ions à 300000. Un total de 3 spectres MS/MS est moyenné pour obtenir les spectres MS/MS finaux. L'ensemble du système est complètement contrôlé par les programmes ChemStation Rev. B.01.03

(Agilent Technologies) et EsquireControl 6.1 Build 78 (Bruker Daltonics). Les données brutes collectées pendant les analyses nanoLC-MS/MS sont traitées et converties avec le programme DataAnalysis 3.4 Build 169 en fichiers « peak list » \*.mgf.

### **1.1.2. Optimisation des conditions chromatographiques en utilisant l'acétonitrile comme solvant organique sur un système nanoLC-Chip/MS**

Afin de vérifier les performances du système nanoLC-MS/MS sur un système Agilent 1100 Series HPLC-Chip/MS (Agilent Technologies, Palo Alto, USA) couplé à un spectromètre de masse HCT Plus ion trap (Bruker Daltonics, Brême, Allemagne), 50 fmol d'un digestat trypsique de BSA sont analysés en nanoLC-MS et nanoLC-MS/MS.

Les voies A de la pompe nano et de la pompe de chargement délivrent un mélange composé de 98% H<sub>2</sub>O, 2% AcN et 0,1% HCOOH. La voie B de la pompe nano délivre un mélange composé de 98% AcN, 2% H<sub>2</sub>O, et 0.1% HCOOH.

Le mélange peptidique est chargé sur la précolonne d'enrichissement Zorbax 300SB-C18 (volume 40 nL, particules de 5 µm, porosité 300 Å) à 3,75 µL/min. Après dessalage, les peptides sont élués à un débit de 300 nL/min au travers de la colonne analytique Zorbax 300SB-C18.

Pour les puces microfluidiques contenant une colonne Zorbax 300SB-C18 (75 µm de diamètre interne x 150 mm de longueur, particules de 5 µm, porosité 300 Å), le gradient utilisé est le suivant : 8 à 40 % de B en 30 min, suivi d'un palier de 3 min à 70% de B et d'un rééquilibrage de la colonne à 8% de B pendant 9 min.

Pour les puces microfluidiques contenant une colonne Zorbax 300SB-C18 (75 µm de diamètre interne x 43 mm de longueur, particules de 5 µm, porosité 300 Å), le gradient utilisé est le suivant : 8 à 40 % de B en 7 min, suivi d'un palier de 3 min à 70% de B et d'un rééquilibrage de la colonne à 8% de B pendant 5 min.

Le spectromètre de masse HCT Plus ion trap est étalonné avec des composés de référence (Tuning Mix ES (Agilent Technologies)). Les paramètres généraux du spectromètre de masse sont les suivants : tension capillaire, -1750V; gaz de séchage, 4 L/min; température de séchage, 325°C. Le système opère avec un passage automatique entre les modes MS et MS/MS. L'analyse MS est réalisée dans le mode de résolution "standard enhanced" à une vitesse de balayage de 8100 m/z par seconde avec un contrôle de la charge d'ions à 100000 dans un temps de remplissage maximale de 200 ms. Un total de 4 spectres MS est moyenné pour obtenir les spectres MS finaux. Les 3 ions les plus intenses, préférentiellement doublement ou triplement chargés sont sélectionnés dans chaque spectre MS pour être isolés et fragmentés. L'analyse MS/MS est réalisée dans le mode de résolution "ultrascan" à une vitesse de balayage de 26000 m/z par seconde avec un contrôle de la charge d'ions à 300000. Un total de 3 spectres MS/MS est moyenné pour obtenir les spectres MS/MS finaux. L'ensemble du système est complètement contrôlé par les programmes ChemStation Rev. B.01.03 (Agilent Technologies) et EsquireControl 5.3 Build 11 (Bruker Daltonics). Les données brutes collectées pendant les analyses nanoLC-MS/MS sont traitées et converties avec le programme DataAnalysis 3.3 Build 146 en fichiers « peak list » \*.mgf.

## 1.2. Chapitre 2

Les expériences de nanoLC-MS et MS/MS ont été réalisées sur un système nanoACQUITY Ultra-Performance-LC (UPLC, Waters, Milford, USA) couplé à un spectromètre de masse de type Q-TOF maXis (Bruker Daltonics, Brême, Allemagne) équipé avec une source nanospray. L'ensemble du système est contrôlé par le logiciel Hystar 3.2 (Bruker Daltonics).

Les échantillons sont chargés sur une pré-colonne Symmetry C18 (20 x 0,18 mm, particules de 5  $\mu\text{m}$ , Waters). Les peptides sont séparés sur une colonne ACQUITY UPLC® BEH130 C18 (75  $\mu\text{m}$  x 200 mm, particules de 1,7  $\mu\text{m}$ , Waters). Les solvants sont composés pour la voie A de 0,1% d'acide formique dans de l'eau et pour la voie B de 0,1% d'acide formique dans de l'acétonitrile. Le chargement de la pré-colonne s'effectue pendant 1 min à 15  $\mu\text{L}/\text{min}$  avec 99% de solvant A et 1% de solvant B. L'élution des peptides s'effectue, à 50°C, à un débit de 450 nL/min, avec un gradient allant de 1 à 35% de solvant B en 21 min suivi d'un saut à 90% de B et un retour aux conditions initiales.

Le spectromètre de masse Q-TOF utilise les paramètres suivant : la température de la source est réglée à 200°C, le gaz de séchage à un débit de 4L/h et la tension sur le nanospray est de 4 kV. L'étalonnage du TOF est réalisé en utilisant le mélange Tuning Mix ES-TOF (Agilent Technologies) dans le mode positif sur la gamme de balayage  $m/z$  50 à 2200.

Une première analyse nanoLC-MS est réalisée pour la détection des glycopeptides. Pour cela, les ions oxoniums sont générés en utilisant un mode de scan alterné dans la cellule de collision en appliquant séquentiellement une haute et une basse énergie (10/28 eV). L'expérience est réalisée dans la gamme de balayage  $m/z$  50 à 2200 avec un temps d'accumulation de 1s par spectre.

Une deuxième analyse nanoLC-MS/MS est réalisée pour caractériser les glycopeptides. Le système fonctionne en réalisant un passage automatique de la MS à la MS/MS. La gamme de balayage est  $m/z$  50 à 2200. En MS la durée d'accumulation est de 0,2 s. En MS/MS la durée d'accumulation est pondérée en fonction de l'intensité du parent soit 0,2 à 1,4 s pour des intensités allant de 24000 au à 400 au. Les 3 ions les plus intenses (intensité minimale 400 au), en favorisant les ions 2, 3 et 4 fois chargés, sont sélectionnés pour chaque spectre MS pour être isolés et fragmentés par CID en utilisant 2 énergies déterminées en fonction de la masse et de la charge de l'ion. La fragmentation est réalisée en utilisant l'argon comme gaz de collision. Chaque ion parent est sélectionné une fois puis exclu durant toute la durée du pic chromatographique.

L'interprétation des données est ensuite réalisée manuellement. Les spectres MS/MS des glycopeptides sont recherchés au niveau des zones d'élution des ions diagnostiques. Les spectres contenant des ions diagnostiques correspondent à la fragmentation de glycopeptides. Les fragments majoritaires permettent de déterminer la structure du glycane et la masse du peptide non glycosylé. Afin d'identifier le site de glycosylation, la glycoprotéine est digérée *in silico* pour prédire les séquences des glycopeptides potentiels. Les masses théoriques des peptides contenant un site de glycosylation sont comparées aux masses mesurées dans le spectre MS/MS. La masse de l'ion correspondant au peptide non glycosylé n'est pas toujours présent sur le spectre MS/MS mais peut être calculée à partir de la masse de l'ion correspondant au peptide portant un GlcNAc qui est souvent le fragment majoritaire. Les résultats sont validés en tenant compte de l'interprétation des trois types

d'ions fragments classiquement générés par les glycopeptides (perte de neutre de sucre, fragments internes du peptide et fragments internes du glycane).

## **2. Partie 2 des résultats**

### **2.1. Chapitre 3**

Le détail des expériences et des protocoles analytiques pour les études protéomiques du glideosome et des facteurs de transcription de *Toxoplasma gondii* est décrit dans les publications des résultats.

La méthode utilisée pour la caractérisation des glycosylations de la protéine GAP50 est identique à celle décrite dans le Chapitre 2.

### **2.2. Chapitre 4**

Le détail des expériences et des protocoles analytiques est décrit dans la publication des résultats.

### **2.3. Chapitre 5**

Le détail des expériences et des protocoles analytiques est décrit dans la publication des résultats.



Institut Pluridisciplinaire Hubert Curien

Service de reprographie  
Michel Schlée

23, rue du Loess  
67037 Strasbourg Cedex 2