



HAL
open science

Un ensemble d'outils protéomiques pour la caractérisation de protéines d'organismes très divers : plantes, champignons et parasites

Tchilabalo Dilezitoko Alayi

► **To cite this version:**

Tchilabalo Dilezitoko Alayi. Un ensemble d'outils protéomiques pour la caractérisation de protéines d'organismes très divers : plantes, champignons et parasites. Autre. Université de Strasbourg, 2013. Français. NNT : 2013STRAF018 . tel-00956806

HAL Id: tel-00956806

<https://theses.hal.science/tel-00956806v1>

Submitted on 7 Mar 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



UNIVERSITÉ DE STRASBOURG

EDSC
Ecole Doctorale des
Sciences Chimiques

ÉCOLE DOCTORALE DES SCIENCES CHIMIQUES

UMR7178-IPHC



THÈSE

Présentée par

Tchilabalo Dilezitoko ALAYI

Pour l'obtention de grade de

Docteur de l'Université de Strasbourg

Discipline / Spécialité : **Chimie Analytique**

**UN ENSEMBLE D'OUTILS PROTÉOMIQUES POUR LA
CARACTÉRISATION DE PROTÉINES D'ORGANISMES TRÈS
DIVERS : PLANTES, CHAMPIGNONS ET PARASITES**

Soutenue le : 28 Mai 2013 devant la commission d'examen

THÈSE dirigée par :

Dr. Alain VAN DORSELAER

Directeur de Recherche, CNRS, Université de Strasbourg

RAPPORTEURS :

Dr. Hubert SCHALLER

Pr. Isabelle VILLENA

Dr. Philippe BULET

Directeur de Recherche, CNRS, Université de Strasbourg

Professeur, Université de Reims Champagne-Ardenne

Directeur de Recherche, CNRS, Université de Grenoble

INVITÉ :

Dr. Christine SCHAEFFER-REISS

Ingénieur de Recherche, CNRS, Université de Strasbourg

A mes parents

A mes frères

A mes amis

« Ce n'est pas parce que les choses sont difficiles que nous n'osons pas, mais c'est parce que nous n'osons pas qu'elles sont difficiles »

Sénèque (4 avant JC - 65 près JC)

REMERCIEMENTS

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

Si ce travail se présente tel quel, c'est grâce à la contribution, la collaboration et à l'implication de certaines personnes à qui il convient d'adresser mes plus sincères remerciements.

Mes remerciements s'adressent tout d'abord au Dr. Alain Van Dorsselaer, Directeur de recherche au CNRS, pour m'avoir fait confiance, accueilli, soutenu, encouragé et donné les moyens nécessaires pour l'accomplissement de ce travail de thèse dans un cadre de travail idéal. Mes remerciements vont également au Dr. Christine Schaeffer-Reiss pour son encadrement au quotidien malgré son emploi de temps très chargé, pour avoir cru en mes idées quand il fallait et m'avoir impliqué dans des projets scientifiques aussi passionnants.

Je remercie également, l'Université de Strasbourg pour m'avoir offert cette belle opportunité de thèse et pour m'avoir supporté financement lors de la réalisation de ce travail de thèse.

Je remercie vivement le Dr. Philippe Bulet qui a accepté de présider mon jury de thèse ainsi que le Dr. Isabelle Villena et le Dr. Hubert Schaller pour avoir accepté d'évaluer mon travail de thèse.

Je remercie également l'ensemble de nos collaborateurs à savoir le Dr. Stanislas Tomavo de l'Institut Pasteur de Lille (France), le Dr. Florence Fontaine et le Dr. Sandrine Jégou de l'Université Champagne-Ardenne à Reims (France) et le Prof. Emmanuel Bisse de l'Université de Freiburg (Allemagne) et à tous les membres de leur équipe.

Je voudrais également dire un grand merci à tous ceux avec qui j'ai partagé ces quelques années dans la termitière (LSMBO):

Aux ouvriers du bureau des précieux (Diego Bertaccini, Sarah Lennon, Jean Michel Saliou, Anna Ziemianin, Magalie Rompais et Marine Plumel) et du côté du bureau des pousseurs (Stéphanie Petiot, Nicolas Barthelemy, Daniel Ayoub, Thierry Wasselin, François Debaene), je vous dis merci à tous pour l'ambiance, les discussions et les moments passés ensemble. Aux anciens du bureau des précieux (Guillaume Bechade, Latitia Fouillen), les instants passés ensemble étaient brefs mais instances, je garde de beaux souvenirs. Je remercie également mes compagnons de tous les jours au moment de la rédaction de ce manuscrit de thèse, Amandine Bœuf, Gilles Schnell vous étiez formidables.

Je voudrais également remercier celle qui n'a jamais eu de camp, la conciliante Dr. Agnès Hovasse, pour m'avoir appris et transmis beaucoup de son savoir faire en protéomique, j'espère que tu trouveras ici satisfaction du produit final, je te remercie infiniment.

Aux ouvriers qualifiés et aux soldats, Hélène Diemer (Mme ISO), Daniel Thiersé (Mme Beau Gel), Dr. Sarah Cianferani (Mme Non-covalent), Dr. Laurence Sabatier (Mme. Précision), Alison Opsomer (Mme SRM), Kevin Jeanpert (M. le commis), Dr. Jean-Marc Strub (M. MALDI), Dr. Fabrice Bertille (M. Biologie), Fabrice Varrier (M. Réseau) et Dr. François Delalande (M. Pratique), Alexandre Burel (M. Script et Macro), Alexandre Walter (M. Proline), Patrick Guterl (M. Toubib de PC), René Winckler (M. Sous-Sol), je ne vous remercierai jamais assez pour votre contribution à ce travail.

Mes sincères remerciements vont également à la grande sœur rêvée de tout thésard, Dr. Christine Carapito pour l'accueil à l'aéroport, les démarches de mon installation à Strasbourg, pour avoir été toujours là et pour les discussions fructueuses ayant contribué à la réalisation de ce travail de thèse. Il n'y a pas de mot assez fort pour exprimer ma profonde gratitude.

Aux nouveaux arrivés dans la termitière (LSMBO), Elisa Ruhland, Drouin Nicolas, Guillaume Terral, Johann Stojko, le temps passe vite, profitez, posez des questions, vous tenez désormais le drapeau du LSMBO.

Un grand merci au Dr. Souad El Hajjaji responsable pédagogique du master Sciences Analytiques à la Faculté des sciences de Rabat (Maroc) pour m'avoir soutenu dans ma démarche de recherche de laboratoire d'accueil et de sujet de thèse depuis le Maroc.

Un grand merci à mon entourage et mes plus proches amis, même si l'éloignement géographique limite les contacts, c'est toujours un immense plaisir de vous retrouver et de vous savoir toujours nombreux à mes côtés.

A ma très Chère Mère patrie le TOGO qui m'a offert cette belle opportunité, à cette terre de nos aïeux que j'ai l'honneur de représenter depuis 2003 à l'extérieur, je lui dédie ce travail.

Enfin, que ceux sans qui ce travail n'aurait pu être mené à bout, et qui souhaitent demeurer dans l'ombre au milieu des étoiles et des planètes, trouvent ici l'expression de ma profonde gratitude.

PLAN GÉNÉRAL

REMERCIEMENTS.....	v
PLAN GÉNÉRAL.....	ix
PLAN DÉTAILLÉ.....	xi
LISTE DES PRINCIPALES ABRÉVIATIONS.....	xv
LISTE DES ACIDES AMINÉS LES PLUS COURANTS.....	xvii
INTRODUCTION GÉNÉRALE.....	1
PARTIE I : INTRODUCTION BIBLIOGRAPHIQUE.....	5
Chapitre 1 : Les fondamentaux en analyses protéomiques par spectrométrie de masse.....	7
Introduction.....	8
I. La spectrométrie de masse : une technique de choix pour la protéomique.....	10
II. Techniques de séparation et de purification protéiques.....	24
III. Stratégie d'identification des protéines par spectrométrie de masse : nanoLC-MS/MS.....	31
Conclusion.....	42
Chapitre 2 : Modifications post-raductionnelles : N-glycosylation.....	51
Introduction.....	52
I. La glycosylation : une PTM courante.....	53
II. Difficultés d'analyse des glycoprotéines : cas des N-glycosylations.....	55
III. Les outils d'interprétation des données LC-MS/MS des glycopeptides et oligosaccharides.....	68
Conclusion.....	70
PARTIE II : RÉSULTATS.....	77
Chapitre 1 : Etude du parasite <i>Toxoplasma gondii</i> responsable de la toxoplasmose.....	79
I. Contexte de cette étude.....	80
II. Identification des protéines impliquées dans l'intrusion du parasite chez l'hôte.....	84
III. Identification des facteurs de transcription qui gouvernent le mécanisme de différenciation de <i>T. gondii</i> de la forme virulente en forme dormante.....	89
IV. Etude protéomique des partenaires de la sortiline (TgSORTLR).....	95
Conclusion et perspectives.....	101
Chapitre 2 : Etudes protéomique et génomique des changements induits par la maladie de l'esca sur le <i>Vitis vinifera</i> L.cv. Chardonnay.....	107
I. Contexte de cette étude.....	108
II. Etude des tiges herbacées de vignes infectées par l'esca.....	111
III. Etude des troncs de vignes infectés par l'esca.....	119
Conclusion et perspectives.....	139
Chapitre 3 : Etude et identification de nouveaux variants d'hémoglobine humaine.....	143
I. Contexte de cette étude.....	144
II. Identification de nouveaux variant.....	148
Conclusion et perspectives.....	155
Chapitre 4 : Caractérisation des glycosylations de l'invertase vacuolaire de raisin.....	163
I. Contexte de cette étude.....	164
II. Identification et caractérisation de la glycosylation de l'invertase vacuolaire de raisin.....	167
Conclusion et perspectives.....	177
CONCLUSION GENERALE.....	181
PARTIE EXPÉRIMENTALE.....	185
LISTE DES PUBLICATIONS ET COMMUNICATIONS.....	191

PLAN DÉTAILLÉ

REMERCIEMENTS	v
PLAN GÉNÉRAL	ix
PLAN DÉTAILLÉ.....	xi
LISTE DES PRINCIPALES ABRÉVIATIONS.....	xv
LISTE DES ACIDES AMINÉS LES PLUS COURANTS	xvii
INTRODUCTION GÉNÉRALE.....	1
PARTIE I : INTRODUCTION BIBLIOGRAPHIQUE.....	5
Chapitre 1 : Les fondamentaux en analyses protéomiques par spectrométrie de masse	7
Introduction	8
I. La spectrométrie de masse : une technique de choix pour la protéomique	10
1. Les sources d'ionisation : Electrospray (ESI).....	12
2. L'interface.....	14
a. Les guides d'ions multipôlaires.....	14
b. Les guides d'ions de type SRIG " <i>Stacked ring ion guide</i> " et IF " <i>Ion funnels</i> "	15
3. Les analyseurs : trappe ionique, quadrupôle, TOF " <i>Time of Flight</i> "	15
a. Analyseurs à une dimension MS : quadripôle et le temps de vol (TOF).....	16
i. Quadrupôle (Q)	16
ii. Le temps de vol (TOF).....	16
b. Analyseur à plusieurs dimensions MS : la trappe ionique.....	18
4. Détecteurs	19
5. Description des spectromètres utilisés.....	20
a. La trappe ionique : amaZon-ETD (Bruker, Daltonics).....	20
b. Les instruments de type Q-TOF: maXis et micrOTOF-Q (Bruker, Daltonics) et Q-TOF Synapt HDMS (Waters, Corporation).....	21
i. Le maXis et micrOTOF-Q (Bruker, Daltonics)	21
ii. Le Synapt HDMS (Waters, Corporation).....	22
6. De l'ion à la séquence protéique : la fragmentation peptidique.....	22
a. La fragmentation peptide : dissociation induite par collision (CID).....	23
Nomenclature de fragmentation des peptides	23
b. Modèle de fragmentation : proton mobile.....	23
II. Techniques de séparation et de purification protéiques	24
1. Séparation des protéines par électrophorèse sur gel : gel 1D et gel 2D.....	24
a. Electrophorèse sur gel monodimensionnelle (gel 1D)	25
b. Electrophorèse sur gel bidimensionnelle (gel 2D)	25
2. Séparation des protéines et peptides par chromatographie liquide.....	26
a. Séparation et purification des protéines par chromatographie liquide : LC.....	26
i. Séparation des protéines par chromatographie de phase inverse (RPLC).....	26
ii. Séparation et purification des protéines par chromatographie d'affinité	26
Les méthodes de co-immunopurification	27
Les méthodes de co-précipitation par affinité	27
Les limites des techniques de co-purification par affinité.....	27
b. Séparation des peptides par nano-chromatographie liquide : nanoLC.....	28
Chromatographie liquide multidimensionnelle.....	30
III. Stratégie d'identification des protéines par spectrométrie de masse : nanoLC-MS/MS	31
1. Acquisition de spectres de fragmentation peptidique (LC-MS/MS)	31
a. Mode d'acquisition données-dépendant : DDA (Data Dependant or Directed Acquisition)	31
b. Mode d'acquisition données-indépendant: DIA (Data-Independent-Acquisition).....	32
2. Interprétation de données MS/MS	32
a. Approche par empreinte de fragmentation peptidique : Peptide Fragment Fingerprinting (PFF)..	32
Limites de la recherche conventionnelle par les moteurs de recherche.....	33

b.	Approche par séquençage <i>de novo</i> et identification inter-espèces par recherche de similarités de séquences (homologie de séquence)	35
i.	Séquençage <i>de novo</i> automatisé des spectres MS/MS	35
	Approche par pseudo-PFF (Exhaustive search approche).....	35
	Reconstruction itérative d'une séquence de peptide (Spectrum graph approach)	36
ii.	Identification inter-espèces par recherche de similitude (homologie de séquences)	37
c.	Les banques de séquences protéiques.....	38
3.	Evaluation des résultats	39
	Conclusion	42
	Références bibliographiques du chapitre 1	43
	Chapitre 2 : Modifications post-raductionnelles : N-glycosylation	51
	Introduction	52
I.	La glycosylation : une PTM courante	53
II.	Difficultés d'analyse des glycoprotéines : cas des N-glycosylations	55
1.	Complexité de la N-glycosylation : Micro- et Macro-hétérogénéité.	55
2.	Stratégie d'analyse de glycoprotéines par spectrométrie de masse	56
3.	Techniques d'enrichissement des glycoprotéines et glycopeptides	57
a.	La chromatographie d'affinité.....	58
	□ Les phases de lectines	58
	□ Les phases hydrazides	58
	□ Phase à base d'oxyde de métaux	59
b.	La chromatographie d'interaction hydrophile	59
4.	Analyse de glycoprotéines entières.....	60
5.	Analyse des peptides déglycosylés et glycopeptides	61
a.	Analyse des peptides déglycosylés.....	61
b.	Analyse MS ⁿ des oligosaccharides	63
c.	Analyse de glycopeptides par MS	64
III.	Les outils d'interprétation des données LC-MS/MS des glycopeptides et oligosaccharides	68
1.	Interprétation des données MS et MS/MS des oligosaccharides.	68
2.	Interprétation des données MS/MS de glycopeptides.....	68
	Conclusion	70
	Références bibliographiques du chapitre 2	71
	PARTIE II : RÉSULTATS	77
	Chapitre 1 : Etude du parasite <i>Toxoplasma gondii</i> responsable de la toxoplasmose	79
I.	Contexte de cette étude	80
1.	<i>Toxoplasma gondii</i> et la toxoplasmose	80
2.	Problématique et objectifs de l'étude.....	82
II.	Identification des protéines impliquées dans l'intrusion du parasite chez l'hôte	84
1.	Contexte et objectif de cette étude	84
2.	Stratégie d'analyse protéomique des complexes protéiques de TgGAP50.....	84
a.	Description des échantillons.....	84
b.	Analyses nanoLC-MS/MS.....	84
3.	Résultats publiés	85
	Conclusion.....	86
	Annexe 1: Publication	87
III.	Identification des facteurs de transcription qui gouvernent le mécanisme de différenciation de <i>T. gondii</i> de la forme virulente en forme dormante	89
1.	Contexte et objectif de cette étude	89
2.	Stratégie d'analyse protéomique des complexes protéiques des TgAlba	89
a.	Description des échantillons.....	89
b.	Analyse nanoLC-MS/MS	89
3.	Résultats publiés	90
	Conclusion.....	91
	Annexe 2: Publication	93

IV. Etude protéomique des partenaires de la sortiline (TgSORTLR)	95
1. Contexte et objectif de cette étude	95
2. Stratégie d'analyse protéomique des complexes protéiques de TgSORLR	95
a. Description des échantillons	95
b. Analyse nanoLC-MS/MS	96
3. Résultats publiés	96
Conclusion	98
Annexe 3: Publication	99
Conclusion et perspectives	101
Références bibliographiques du chapitre 1	103
Chapitre 2 : Etudes protéomique et génomique des changements induits par la maladie de l'esca sur le <i>Vitis vinifera</i> L.cv. Chardonnay	107
I. Contexte de cette étude	108
1. La maladie de l'esca	108
2. Problématique et objectifs de l'étude	109
II. Etude des tiges herbacées de vignes infectées par l'esca	111
1. Description des échantillons	111
2. Stratégie d'analyse protéomique des tiges herbacées de vignes	111
a. Stratégie globale de l'analyse protéomique	111
b. Identification des protéines des tiges herbacées de vignes par recherche conventionnelle dans une banque protéique	112
c. Identification des fonctions des protéines hypothétiques par homologie de séquence BLAST	112
Validation des résultats BLAST	113
3. Résultats publiés	113
Conclusion	115
Annexe 4: Publication	117
III. Etude des troncs de vignes infectés par l'esca	119
1. Description des échantillons	119
2. Stratégie d'analyse protéomique des troncs de vignes	119
a. Stratégie globale de l'analyse protéomique	119
b. Identification des protéines des troncs de vigne par recherche conventionnelle dans une banque protéique avec les algorithmes de recherche Mascot et OMSSA	122
c. Identification des protéines des troncs de vigne par séquençage <i>de novo</i> (PepNovo) et recherche MS-BLAST	123
i. Paramétrage de l'algorithme de séquençage <i>de novo</i> PepNovo	125
ii. Recherche d'homologie de séquences par MS-BLAST	125
iii. Validation des résultats de recherche PepNovo/MS-BLAST	126
3. Résultats	129
a. Résultats de la recherche conventionnelle dans la banque protéiques avec les algorithmes de recherche Mascot et OMSSA	129
b. Résultats de la recherche par séquençage <i>de novo</i> PepNovo/MS-BLAST	130
c. Comparaison de résultats de recherche conventionnelle dans la banque de données et recherche par séquençage <i>de novo</i> PepNovo/MS-BLAST	132
4. Application des deux stratégies d'identification de protéines à d'autres variétés de vignes infectées par la maladie de l'esca	133
a. Résultats des troncs de vigne de la variété Mourvèdre	133
b. Résultats des troncs de vigne de la variété Gewurztraminer	134
Conclusion	135
Annexe 5: Article soumis pour publication	137
Conclusion et perspectives	139
Références bibliographiques du chapitre 2	141
Chapitre 3 : Etude et identification de nouveaux variants d'hémoglobine humaine	143
I. Contexte de cette étude	144
1. L'hémoglobine et les hémoglobinopathies	144
2. Problématique et objectifs de l'étude	146

II. Identification de nouveaux variant	148
1. Description des échantillons	148
2. Stratégie d'identification de nouveaux variants	148
3. Principaux résultats	149
a. LC-MS des chaînes de globines intactes	149
b. Identification et caractérisation des variants d'hémoglobine	151
Conclusion	153
Conclusion et perspectives	155
Annexe 6: Tableaux	157
Références bibliographiques du chapitre 3	161
Chapitre 4 : Caractérisation des glycosylations de l'invertase vacuolaire de raisin	163
I. Contexte de cette étude	164
1. Invertase vacuolaire: fonction et rôle	164
2. Problématique et objectifs de l'étude	165
II. Identification et caractérisation de la glycosylation de l'invertase vacuolaire de raisin	167
1. Stratégie d'analyse protéomique	167
a. Description des échantillons	167
b. Stratégie d'analyse par spectrométrie de masse	167
i. Protocole d'analyse A (peptides déglycosylés)	167
ii. Protocole d'analyse B (glycopeptides intacts)	168
2. Principaux résultats	169
a. Identification des sites de glycosylation des peptides déglycosylés	169
b. Identification et caractérisation des glycopeptides	172
Conclusion et perspectives	177
Références bibliographiques	179
CONCLUSION GENERALE	181
PARTIE EXPÉRIMENTALE	185
I. Partie résultats: chapitre 1 et 2	187
II. Partie résultats: chapitre 3	187
1. Séparation et mesure de masse des chaînes de globines entières	187
2. Analyse nanoLC-MS/MS du digeste trypsique de la chaîne de globine mutée	187
III. Partie résultats: chapitre 4	188
1. Digestions enzymatique de l'invertase	188
2. Analyse nanoLC-MS et MS/MS des peptides et glycopeptides	189
LISTE DES PUBLICATIONS ET COMMUNICATIONS	191
Liste des publications	193
Communications orales	193
Communications par affiche	193

LISTE DES PRINCIPALES ABRÉVIATIONS

ACN :	Acétonitrile
ADN :	Acide désoxyribonucléique
ARN :	Acide ribonucléique
BPC :	Base Peak Chromatogram
BLAST :	Basic Local alignment Search Tool
CID :	Collision-Induced Dissociation
Da :	Dalton
DDA :	Data Dependent Acquisition
DIA :	Data Independent Acquisition
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
HPLC :	High Performance Liquid Chromatography
ICC :	Ion Charge Control
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 (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

INTRODUCTION GÉNÉRALE

INTRODUCTION GÉNÉRALE

Grâce aux progrès instrumentaux et technologiques réalisés en spectrométrie de masse et à l'explosion des séquençages génomiques, l'analyse protéomique est devenue un outil d'étude des protéomes très puissant puisqu'elle peut maintenant s'appuyer sur des banques de séquences protéiques prédites à partir de séquences génomiques. L'analyse protéomique consiste à identifier les protéines présentes dans un extrait protéique provenant de tissus, de fluides biologiques ou de culture de cellules, à un moment donné de leur développement et avec un historique donné. Mais des progrès doivent encore être réalisés lorsque, par exemple, on s'intéresse aux modifications post-traductionnelles (PTMs) ou aux différentes isoformes des protéines présentes dans un échantillon, puisque ces informations ne sont pas codées de façon simple dans les génomes. De même, lorsque les échantillons sont particulièrement complexes ou lorsque la quantité de matériel biologique disponible est très faible, les méthodes usuelles ne permettent pas toujours d'obtenir les résultats escomptés. Malgré les nombreux progrès réalisés ces 20 dernières années, l'analyse protéomique se trouve encore confrontée à des défis techniques et méthodologiques. Il apparaît donc nécessaire de développer des méthodologies et des stratégies d'analyse protéomique adaptées au contexte de l'étude, en combinant souvent différents outils de l'analyse protéomique. C'est donc dans ce contexte que se situent mes travaux de thèse.

L'objectif de mes travaux analytiques a porté sur plusieurs aspects. J'ai tout d'abord essayé d'apporter des éléments de caractérisation structurale, et ceci souvent pour des protéines disponibles en très petites quantités. Dans la mesure du possible, j'ai eu pour objectif de passer de la simple identification des protéines à une caractérisation plus complète des séquences protéiques. Cet objectif de détermination totale des séquences en analyse protéomique reste bien sûr lointain, mais durant mes travaux de thèse j'ai tenté de faire progresser les méthodes dans cette direction. Au-delà de la caractérisation des séquences protéiques dans les différents échantillons étudiés, j'ai été amené à identifier et caractériser aussi finement que possible des PTMs qui affectent très souvent certaines protéines que j'ai identifiées et en particulier les N-glycosylations. J'ai principalement utilisé et optimisé les techniques de séparation telles que les gels d'électrophorèse (1D ou 2D), la chromatographie liquide (LC) couplée ou non à la spectrométrie de masse (MS) et les outils bioinformatiques.

Les sujets abordés dans mes travaux de thèse ont été menés en collaboration avec des biologistes, et portent sur des organismes très variés puisqu'ils touchent à la fois les végétaux, l'homme, les parasites et les champignons. J'ai donc été amené à développer des méthodologies d'analyse protéomique sur des échantillons d'origine très différente, ce qui a demandé un ensemble de mises au point adaptées à chaque problématique.

Ce manuscrit de thèse se divise en deux grandes parties comme suit :

PARTIE I, INTRODUCTION BIBLIOGRAPHIQUE qui résume l'état de l'art en analyse protéomique et présentera :

- **Chapitre I, " Les fondamentaux de l'analyse protéomique par spectrométrie de masse"** qui porte sur les outils et les techniques utilisés (1) lors d'une analyse protéomique pour l'identification des protéines : l'instrumentation, les techniques de purification et de séparation de protéines, les outils bioinformatiques, et (2) les outils utilisés dans les cas particuliers d'identification des protéines issues d'organismes dont les génomes ne sont pas séquencés.
- **Chapitre II, "Modifications post-traductionnelles : N-glycosylation"** qui présente les outils et les stratégies analytiques utilisées lors des études de la N-glycosylation tout en soulignant les difficultés analytiques autour de cette modification post-traductionnelle.

Cette partie d'introduction tentera également de mettre en évidence les axes pour lesquels des progrès et des améliorations sont nécessaires pour l'analyse protéomique en général et dans le cas particulier des N-glycosylations.

PARTIE II, RESULTATS qui comporte 4 chapitres décrivant les résultats obtenus lors des 4 collaborations pour lesquelles nous avons développé des stratégies analytiques spécifiques et adaptées à chaque problématique posée. Ces stratégies ont porté sur l'étude de complexes protéiques, de protéines entières, de glycoprotéines, ou encore sur l'analyse de protéines issues d'organismes fongiques non séquencés.

- Le **chapitre 1** porte sur l'"**Etude du parasite *Toxoplasma gondii* responsable de la toxoplasmose**" avec pour objectif final d'apporter des éléments moléculaires pour développer de nouvelles stratégies

thérapeutiques à l'encontre de ce parasite. Ce projet d'étude de complexes protéiques du parasite *T. gondii* concerne 3 aspects fondamentaux de la biologie de ce parasite que sont : (1) la compréhension du mécanisme d'invasion du parasite dans les cellules hôtes, (2) étude des facteurs de transcription permettant la différenciation de la forme dormante en forme virulente et (3) le trafic intracellulaire des protéines. Nous développerons des stratégies d'analyse protéomique adaptées pour l'analyse des protéines en faible quantité afin d'identifier, et si possible caractériser, les protéines partenaires des protéines cibles (TgGAP 50, TgAlba, et TgSORTLR).

- Le **chapitre 2** est consacré aux "**Etudes protéomique et génomique des changements induits par la maladie de l'esca sur la vigne**". Ce projet d'étude des interactions vigne-pathogènes s'est effectué sur 2 organes de la vigne : les tiges herbacées et les troncs de vigne où siègent les champignons responsables de la maladie. Cette étude va nécessiter la mise en place d'une stratégie d'interprétation des données de LC-MS/MS différente de celle conventionnellement utilisée en analyse protéomique. Nous réaliserons une approche basée sur le séquençage *de novo* suivi de MS-BLAST (recherche par homologie de séquence) afin d'identifier les protéines des champignons de génomes non séquencés. Cette étude devrait permettre de mieux comprendre les mécanismes d'invasion de ces champignons et au final de lutter plus efficacement contre cette maladie de la vigne.
- Le **chapitre 3** présente l'"**Etude et identification de nouveaux variants d'hémoglobine humaine**". Ce projet consacré aux variants d'hémoglobine humaine liés aux anomalies génétiques appelées hémoglobinopathies, a conduit à la mise en place d'une stratégie d'analyse protéomique basée sur la LC-MS et LC-MS/MS. Cette approche vise à caractériser des variants d'hémoglobine pour lesquels les méthodes d'analyse biologique utilisées en routine peuvent donner des résultats erronés et donc conduire à des traitements inadaptés. Ces caractérisations seraient également très utiles pour apporter des informations sur les gènes porteurs de ces mutations et ainsi permettre une analyse d'ADN plus ciblée.
- Le **chapitre 4** porte sur la "**Caractérisation des glycosylations de l'invertase vacuolaire de raisin**" et en particulier la N-glycosylation. Cette invertase, possédant 12 sites potentiels de glycosylation et majoritaire dans les vins de Champagne, est une enzyme clé du métabolisme des sucres du raisin. La caractérisation complète de ces sites de glycosylation est essentielle pour mieux comprendre leur impact sur sa stabilité, sa solubilité, sa conformation, son activité biologique, sa résistance potentielle aux protéases et son rôle dans les propriétés moussantes des vins effervescents. Dans ce contexte, nous avons donc réalisé une étude structurale fine des sites de glycosylation et de leur micro-hétérogénéité en utilisant une combinaison de différentes stratégies de spectrométrie de masse basée (1) sur l'utilisation des techniques d'analyse protéomique classiques, et (2) sur la génération d'ions dits "diagnostiques" qui sont spécifiques à la fragmentation des glycopeptides (pour caractériser la micro-hétérogénéité).

Ce travail passe donc par le développement de stratégies d'analyse protéomique par spectrométrie de masse mais aussi par une meilleure interprétation de données MS/MS, qui devront être adaptées à chaque contexte d'étude. Pour des raisons de facilité de lecture, les introductions bibliographiques consacrées aux contextes biologiques et les différentes stratégies analytiques mises en place pour chaque problématique seront présentées directement dans chacun des 4 chapitres de la partie Résultats.

PARTIE I :

INTRODUCTION BIBLIOGRAPHIQUE

Chapitre 1 : Les fondamentaux de l'analyse protéomique par spectrométrie de masse

Chapitre 2 : Modifications post-traductionnelles: N-glycosylation

Chapitre 1 :

Les fondamentaux de l'analyse protéomique par spectrométrie de masse

Ce chapitre consiste en un rappel introductif sur la protéomique et l'état de l'art actuel de l'analyse protéomique par spectrométrie de masse.

Introduction

Le terme protéomique qui découle du terme protéome, a été utilisé pour la première fois en 1994 par Mark Wilkins¹ et très rapidement l'expression a été reprise par la communauté scientifique. Aujourd'hui, la protéomique désigne l'étude de protéome lui-même défini comme l'ensemble des protéines présentes dans une cellule, un tissu, un organe ou un organisme à un moment précis et sous des conditions données. La protéomique est donc une étude dynamique, en ce sens qu'un seul génome c'est-à-dire l'ensemble de gènes d'une cellule ou d'un organisme, peut conduire à différents protéomes en fonction des étapes du cycle cellulaire, de la réponse à différents signaux biologiques, des états physiologiques (Figure 1).

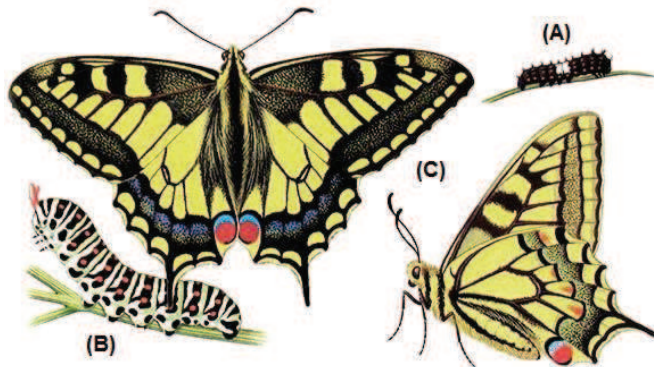


Figure 1. Illustration de la dynamique du protéome chez le papillon *Papilio machaon*. L'expression du génome de *Papilio machaon* diffère suivant le stade de son évolution, c'est-à-dire un même génome conduit à différents protéomes : (A) jeune chenille, (B) chenille adulte, (C) le papillon *Papilio machaon*. Au cours de ces différents stades d'évolution, différentes protéines sont exprimées à partir du même génome ce qui se traduit par différents phénotypes chez *Papilio machaon*².

La protéomique est une science pluridisciplinaire qui s'intéresse aux aspects qualitatifs, quantitatifs, structuraux et fonctionnels des protéines. En pratique, la protéomique s'attache à identifier et caractériser les protéines extraites d'un tissu ou une culture cellulaire ou d'un fluide biologique, leur localisation et leur rôle dans les compartiments cellulaires, leurs éventuelles modifications post-traductionnelles (PTMs) ainsi que leur quantification. Elle permet par exemple de quantifier les variations du taux d'expression d'une protéine en fonction du temps, de l'environnement et l'état d'évolution d'une physiologie ou une pathologie par rapport à un autre. Elle permet aussi d'étudier les interactions protéines – protéines, protéines - ADN ou protéines - ARN.

La spectrométrie de masse est devenue très vite à partir des années 2000, la méthode de choix pour l'analyse des protéines issues d'échantillons complexes. L'utilisation de la spectrométrie de masse en protéomique a connu ce grand essor grâce à l'explosion du nombre de génomes séquencés, aux avancées technologiques réalisées dans plusieurs domaines à l'instar des techniques de purification et de séparation de protéines ou peptides, l'instrumentation et la bioinformatique. L'utilisation de la spectrométrie de masse en protéomique a été surtout possible grâce au développement de méthodes de désorption - ionisation douces (MALDI (Matrix Assisted Laser Desorption Ionisation) et ESI (Electrospray Ionization)) pour l'analyse des macromolécules biologiques dont les auteurs se sont vus décerner le prix Nobel de chimie en 2002 (www.nobelprize.org).

Aujourd'hui, l'analyse protéomique par spectrométrie de masse, a considérablement contribué à la compréhension et au développement de systèmes biologiques, un nouveau "paradigme" des sciences de la vie dans lequel, les processus biologiques sont abordés en termes de réseaux d'interaction moléculaires très dynamiques. Ainsi, la spectrométrie de masse est largement utilisée en protéomique pour la quantification des effets de perturbation du réseau dans le temps et l'espace^{3,4}.

Une analyse protéomique par spectrométrie de masse est composée des principales étapes suivantes : (1) préparation des échantillons, qui consiste à extraire du milieu biologique les protéines ; (2) séparation de protéines ou peptides, qui consiste à dé-complexifier l'échantillon par une séparation des protéines sur gels 1D ou 2D ou par chromatographie liquide (LC) ; (3) l'analyse par spectrométrie de masse (nanoLC-MS/MS), (4) l'interprétation des données d'identification, de caractérisation ou de quantification des protéines (Figure 2).

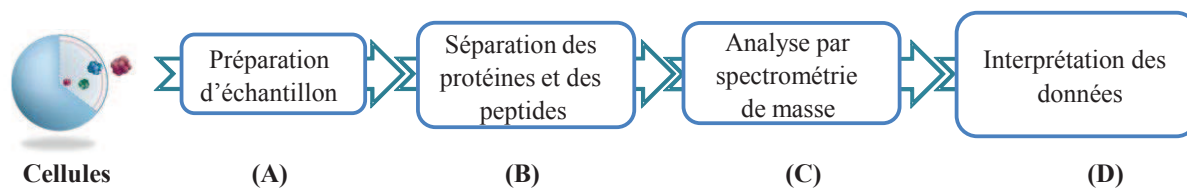


Figure 2. Principales étapes d'une analyse protéomique. Les protéines sont extraites des cellules (A), séparées dans un premier temps par LC ou par gels 1DE ou 2D pour dé-complexifier l'échantillon, les peptides issus de la digestion des protéines sont dans un second temps séparés par LC (B) avant l'analyse par spectrométrie de masse (C). L'étape finale d'une analyse protéomique par spectrométrie de masse est l'interprétation des données de MS ou MS/MS (D). Cette interprétation de données peut se faire aujourd'hui de façon automatisée à l'aide d'outils bioinformatiques.

Nous nous proposons dans ce chapitre de thèse, de faire un point sur l'état de l'art de l'analyse protéomique par spectrométrie de masse en mettant l'accent surtout sur les différents outils et techniques utilisés lors de chacune des étapes des travaux réalisés au cours de cette thèse.

I. La spectrométrie de masse : une technique de choix pour la protéomique

La spectrométrie de masse est née des travaux de Joseph John Thompson (prix Nobel physique 1906) et de Francis William Aston (prix Nobel de chimie 1922). A cette époque, la spectrométrie de masse a été utilisée surtout pour la mesure de masse des atomes et des isotopes. Mais il faut attendre 20 ans d'amélioration du pouvoir de résolution de l'instrument, pour voir le premier spectromètre de masse commercialisé aux Etats Unis en 1942. Très vite le coût et la complexité de l'instrument a été un frein pour son expansion. Les premiers domaines dans lesquels le spectromètre de masse a été utilisé sont: la physique (pour la mesure de masses atomiques), la chimie industrielle (pour le contrôle de qualité de produits) et l'industrie pétrolière (pour la mesure des abondances des hydrocarbures au cours de la production). Ce n'est qu'à partir de 1960 que commence l'intérêt des sciences du vivants et la chimie pour la spectrométrie de masse en tant qu'outils d'aide pour la compréhension de la complexité des molécules fragmentées à l'intérieur de l'instrument, mais aussi le sondage de son utilisation pour d'autres applications. La spectrométrie de masse a été dès lors utilisée pour l'analyse des biomolécules volatiles de petite taille mais pas pour l'analyse de grosses structures moléculaires peu volatiles comme les protéines⁵.

Les années 1980 ont connu une rupture technologique avec la découverte presque simultanée de 2 techniques d'ionisation douces en 1988 qui ont révolutionné l'analyse de macromolécules biologiques : le MALDI (Matrix Assisted Laser Desorption Ionisation) par Koichi Tanaka et l'ionisation électrospray (ESI) par John Fenn (tous deux, lauréats du prix Nobel de chimie en 2002). Ces deux techniques ont rendu possible le transfert de macromolécules intactes de la phase liquide ou solide en phase gazeuse et ont ouvert la voie à la spectrométrie de masse moderne pour l'analyse protéomique⁵⁻⁸.

Aujourd'hui on compte des dizaines de spectromètres de masse disponibles sur le marché dont les constructeurs les plus représentés sont : Agilent Technologies, Applied Biosystems, Bruker Daltonics, PerkinElmer, Shimadzu/Kratos Analytical, Thermo Fisher Scientific, Waters Corporation⁷. Chaque modèle d'instrument est conçu avec un design unique, un système de données, et des performances spécifiques, ce qui entraîne chez chacun de ces instruments des forces ou des faiblesses en fonction des types d'expériences à réaliser.

L'analyse des molécules par spectrométrie de masse se fait en deux temps: dans un première temps, les molécules sont volatilisées et ionisées, et dans un deuxième temps, la mesure des rapports de masses sur charges (m/z) des molécules ionisées est obtenue. Un spectromètre de masse comprend trois parties distinctes : (1) la source où s'effectue la volatilisation et l'ionisation des analytes, (2) l'analyseur où les ions sont séparés selon les rapports m/z et (3) le détecteur où les ions sont détectés sous forme de signaux qui se traduisent en spectres de masse⁹. Le transfert des ions de la source vers le détecteur est assuré par l'effet du champ électrique et/ou magnétique au travers d'un dispositif de focalisation et de guidage d'ions dans une atmosphère où règne un vide poussé (10^{-4} - 10^{-7} mbar). La vitesse et/ou la trajectoire des ions jusqu'au détecteur dépend de leur m/z . Au bout de leur course, les ions frappent le détecteur qui convertit alors l'évènement en signal électrique mesurable (Figure 3), on obtient alors un spectre de masse.

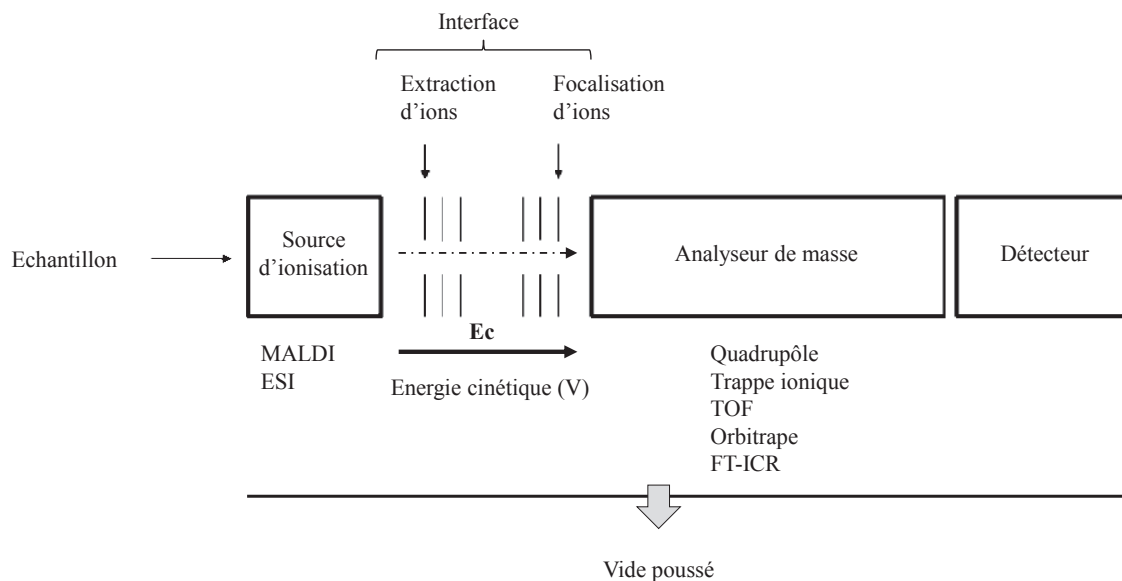


Figure 3. Schéma simplifié d'un spectromètre de masse. Les ions se déplacent de la source jusqu'au détecteur à travers un dispositif d'extraction / focalisation des ions et l'analyseur sous l'effet de l'énergie cinétique de différentes façons suivant les types de source d'ionisation et d'analyseurs utilisés (Figure adaptée de Schaeffer-Reiss⁹).

Pour une analyse protéomique, les principales caractéristiques d'un spectromètre de masse à prendre en considération sont : (1) la résolution de l'analyseur, (2) la précision de mesure de masse, (3) la sensibilité, et (4) la vitesse de balayage¹⁰.

La résolution (R) d'un spectromètre de masse est la capacité à distinguer deux ions de masse voisine : si M_{ex} est la masse expérimentale d'un ion et ΔM_{ex} , la différence de masse expérimentale entre deux ions de masse voisine, alors R est défini par le rapport M_{ex} sur ΔM_{ex} : ($R = M_{ex} / \Delta M_{ex}$) (Figure 4). Pour certains auteurs, la résolution est définie comme le ratio de la masse sur la largeur à mi-hauteur du pic (FWMH, *Full width at Half Maximum*). La résolution est le principal facteur qui va permettre de déterminer le profil isotopique et donc de déterminer l'état de charge des ions¹⁰.

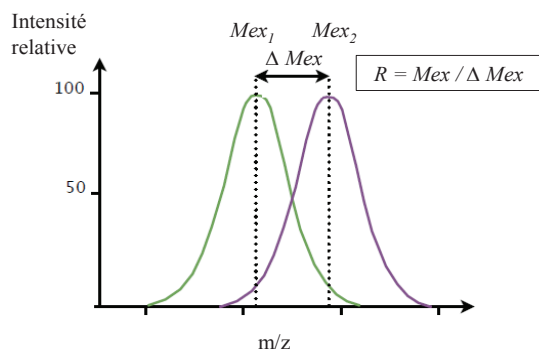


Figure 4. Définition de la résolution (R) d'un spectromètre de masse: R est le rapport entre une masse et la différence entre elle et sa voisine.

La précision de mesure de masse d'un spectromètre de masse quant à elle, décrit la capacité de celui-ci à mesurer une masse expérimentale (M_{ex}) aussi proche que possible de la masse théorique (M_{th}) calculée à partir de la formule brute de chaque élément atomique qui forme cette molécule. Cette précision de masse se mesure en Dalton (Da) ou est ramenée en partie par million (ppm) : $P = \pm 10^6 \times (M_{ex} - M_{th}) / M_{ex}$ (Figure 5). La précision de mesure de masse d'un spectromètre de masse est directement liée à sa résolution.

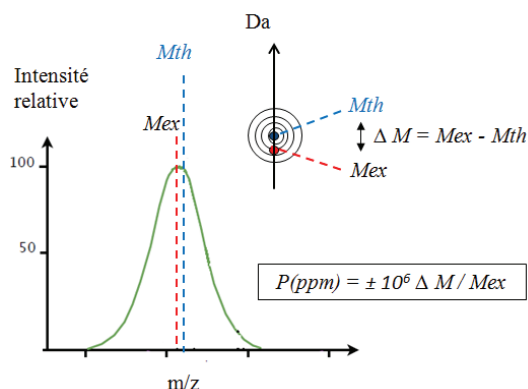


Figure 5. Définition schématisée de la précision de mesure de masse P d'un spectromètre de masse : P exprimé en ppm (partie par million) est le rapport entre la différence de masse expérimentale et théorique ($\Delta M = (Mex - Mth)$) d'un analyte et sa masse expérimentale (Mex) multiplié par 10^6 .

La sensibilité en spectrométrie de masse est définie de deux manières différentes selon que l'on parle d'analyse qualitative ou quantitative. Dans le premier cas, il s'agit de la quantité minimale d'une substance nécessaire pour obtenir un spectre de masse interprétable (rapport ions caractéristiques / ions du bruit de fond). Dans le second cas, il s'agit de la quantité minimale décelable pour un composé mesuré qui donne un signal correspondant à 10 fois le bruit de fond. La sensibilité quantitative est également appelée limite de détection.

Une vitesse de balayage élevée est nécessaire pour l'acquisition d'un grand nombre de spectres des ions qui arrivent en flux continu dans le spectromètre de masse. En outre, plusieurs autres caractéristiques telles que la possibilité d'automatisation et de couplage avec d'autres équipements (HPLC par exemple) pouvant permettre l'analyse protéomique à haut débits sont à prendre en compte en analyse protéomique. De toute évidence, il faut garder à l'esprit qu'il est nécessaire d'avoir un spectromètre de masse de résolution, précision, vitesse de balayage et de sensibilité adaptée pour l'étude protéomique à réaliser¹⁰.

Nous nous limiterons dans les paragraphes suivants à la description des différents éléments majeurs des spectromètres de masse utilisés lors de ces travaux de thèse. Les spectromètres de masse utilisés ont différentes géométries : la trappe ionique (amaZon, Bruker) et les Q-TOF ((maXis et micrOTOF-Q, Bruker) et (Synapt HDMS, Waters)) équipés de source électrospray (ESI).

1. Les sources d'ionisation : Electrospray (ESI)

Les principales sources d'ionisation utilisées en analyse protéomique sont : les sources MALDI et l'Electrospray^{11, 12}. Les deux techniques d'ionisation permettent d'analyser des échantillons à des concentrations de l'ordre du sub-picomolaire. L'ESI s'utilise pour l'analyse des échantillons en phase liquide et donc peut être couplé avec la chromatographie liquide (HPLC).

L'ionisation par électrospray, s'opère par nébulisation à pression atmosphérique des analytes sous forme de nuages d'ions créés sous l'effet d'un champ électrique fort (10^6V.m^{-1}) avec l'assistance pneumatique d'un gaz de nébulisation. En mode d'ionisation positif, la différence de potentiel (ddp) appliquée entre le capillaire métallique d'entrée des analytes et l'entrée du spectromètre de masse va séparer les espèces de charges opposées et repousser les espèces de charges positives multichargées formant le cône de Taylor à l'extrémité du capillaire (Figure 6, (A)) vers le spectromètre de masse sous forme de gouttelettes en plume d'aérosol. Le chauffage de l'enceinte dans laquelle se forment les gouttelettes, va favoriser l'évaporation du solvant. Le volume des gouttelettes chargées se rétrécit jusqu'à la limite de stabilité de Rayleigh. Les forces de répulsions coulombiennes des ions étant en ce moment supérieures à celles de la tension de surface, vont entraîner une explosion coulombienne au fur et à mesure que le solvant s'évapore, jusqu'à ce que l'ion se retrouve complètement libre (Figure 6, (B)).

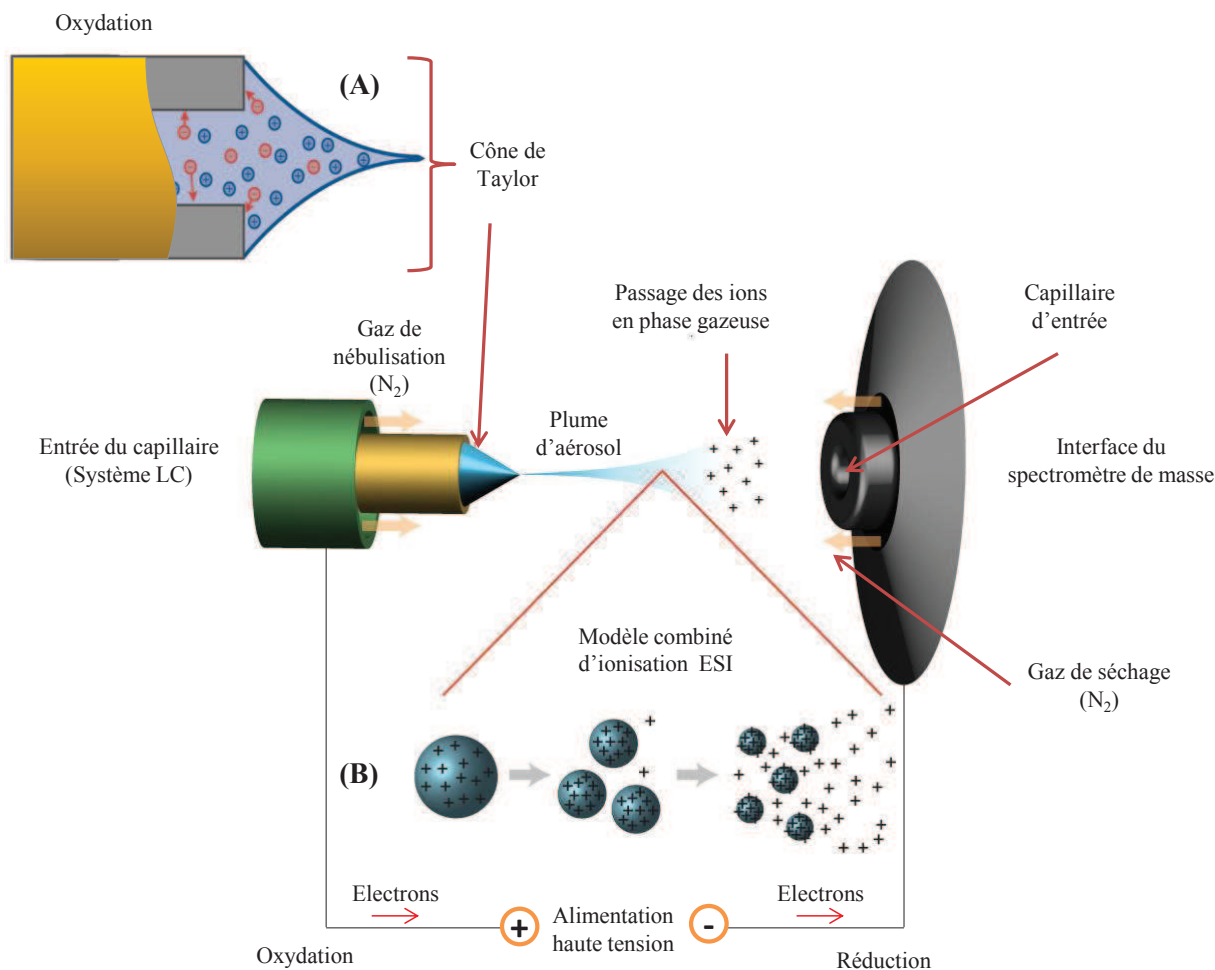


Figure 6. Schéma du principe d'ionisation électrospray en MS en mode d'ionisation positif: production et fission des gouttelettes chargées et transfert des ions en phase gazeuse.

Deux modèles sont admis pour expliquer la production d'ions dé-solvatés en phase gazeuse : le modèle de Dole ou le "**mécanisme de la charge résiduelle**", où les ions sont progressivement désolvatés¹³ et le modèle d'Iribarne et Thomson ou le "**mécanisme de l'évaporation unique**" où les ions sont directement désorbés des gouttelettes formées¹⁴ lors du processus d'ionisation électrospray.

Une source ESI conventionnelle nécessite en général l'utilisation de débits compris entre 1 et 100 $\mu\text{L}/\text{min}$ et l'assistance d'un gaz de nébulisation pour faciliter la formation de gouttelettes chargées à l'extrémité du capillaire (Figure 6). L'amélioration majeure des performances de l'ionisation électrospray est arrivée en 1996, avec le nanospray, qui permet de travailler en ESI avec de faibles débits (300 - 20 nL/min)¹⁵ avec ou sans assistance de gaz de nébulisation. En effet les travaux qui ont été réalisés dans ce contexte ont montré l'influence du débit de solution sur la formation des gouttelettes chargés au cours du processus d'ionisation électrospray. Le faible débit de solvant permet d'assurer la stabilité du spray et la formation de gouttelettes beaucoup plus fines de faible rayon ($R < 10 \text{ nm}$) pour une meilleure efficacité d'ionisation. L'ionisation électrospray étant une technique sensible à la concentration et non à la quantité d'analytes, les sources nanospray permettent non seulement de réduire significativement la consommation d'échantillon, mais également d'avoir un gain de sensibilité d'un facteur 2 à 3¹⁵ et aujourd'hui d'un facteur de près de 10. Tous ces avantages ont très tôt suscité un intérêt particulier du nanospray pour les études protéomiques par MS^{16,17}. On distingue aujourd'hui les systèmes nanospray assistés bénéficiant d'un gaz de nébulisation et les systèmes nanospray non assistés (ex : les puces microfluidiques équipées de nanospray comme l' HPLC ChipTM, Agilent Technologies) qui ont une très bonne sensibilité mais avec une durée de vie limitée (1000 heures en moyen selon notre expérience au laboratoire).

Avec le développement des microtechnologies, de nouveaux dispositifs électrospray sur puces microfluidiques ont été mis au point¹⁸. Les puces microfluidiques ont pu être intégrées à divers systèmes de préparation des échantillons comme la digestion enzymatique ou des systèmes de séparation comme la chromatographie liquide. L'intérêt de ces puces réside non seulement dans l'avantage que représente le

dispositif nanospray que portent ces puces, mais également dans les propriétés intrinsèques de la microfluidique qui sont : la faible consommation d'échantillon, l'amélioration des capacités de séparation et la miniaturisation de plusieurs processus sur un seul système Chip¹⁹. De plus, cette miniaturisation entraîne la réduction des volumes morts et du temps d'analyse. Par exemple l'HPLC-Chip/MS (Agilent, Technologies) est une puce microfluidique qui intègre directement une pré-colonne de concentration et une nanocolonne de séparation à une source nanospray sur un film de polymère. Cette technologie permet ainsi d'éliminer les volumes morts, la dispersion de pics chromatographiques et de donner de bonnes performances chromatographiques (Figure 7). En outre, ces systèmes microfluidiques permettent d'éviter la manipulation délicate des aiguilles nanospray grâce à des dispositifs de gestion intégrés et en entièrement automatisés. A titre d'exemple, l'HPLC-Chip Cube (Agilent Technologies), est un dispositif robotisé qui facilite d'utilisation de la puce microfluidique HPLC-Chip/MS (Agilent Technologies) en terme de réglage de la position de l'aiguille du nanospray et de communication entre la colonne capillaire et les capillaires d'arrivés des échantillons et solvants d'HPLC. Il existe aussi des chips dépourvues de phase stationnaire pour l'infusion.

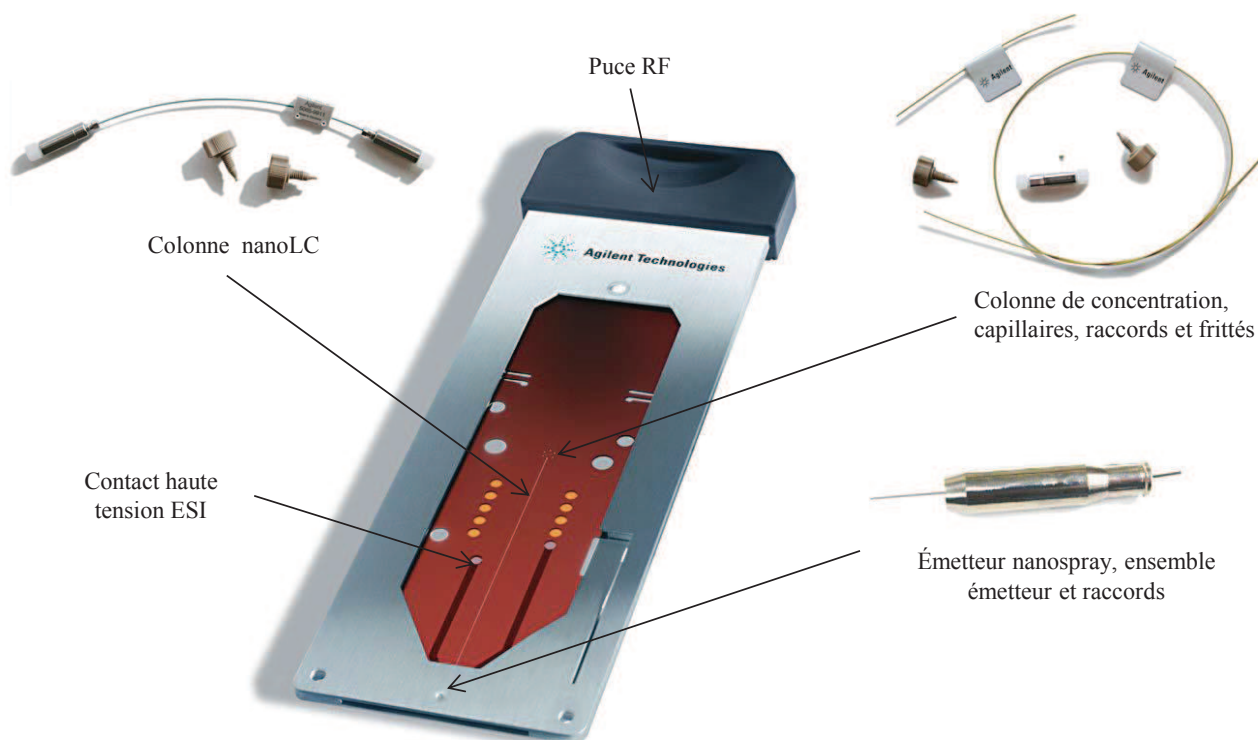


Figure 7. Puce microfluidique HPLC-Chip/MS (Agilent Technologies) intégrant une aiguille nanospray, une nanocolonne et une précolonne de concentration.

2. L'interface

L'interface d'un spectromètre de masse a pour rôle, de transmettre des ions de la source vers l'analyseur sous l'effet d'un champ électrique et d'un gradient de pression. L'interface maximise le rendement de transmission des ions. Les différents instruments utilisés lors de ce travail de thèse possèdent deux géométries d'interface : les guides d'ions multipôles et les guides d'ions de type SRIG "*Stacked ring ion guide*" et IF "*Ion funnels*".

a. Les guides d'ions multipôles

Ce sont des électrodes disposées parallèlement à l'axe de déplacement des ions (Figure 8, (A 1)). En général, les multipôles sont constitués de deux, trois ou quatre paires d'électrodes parallèles (**quadrupôles**, **hexapôle** et **octopôles**) qui sont connectées entre elle de manière à pouvoir appliquer un potentiel de radiofréquence en opposition de phase entre 2 électrodes adjacentes ($U + V\cos(\omega t)$ et $-(U + V\cos(\omega t))$) (Figure 8, (A 2)). La variation du potentiel $V\cos(\omega t)$ lors que $U=0$ "mode RF only" va entraîner une variation du champ électrique et conférer ainsi aux ions, un mouvement oscillatoire vers l'avant²⁰. Le multipôle se comporte alors

exclusivement en guide d'ion et fait déplacer les ions sur l'axe z. Lors que $U \neq 0$ et le quadrupôle se comporte dans cas comme un analyseur de masse.

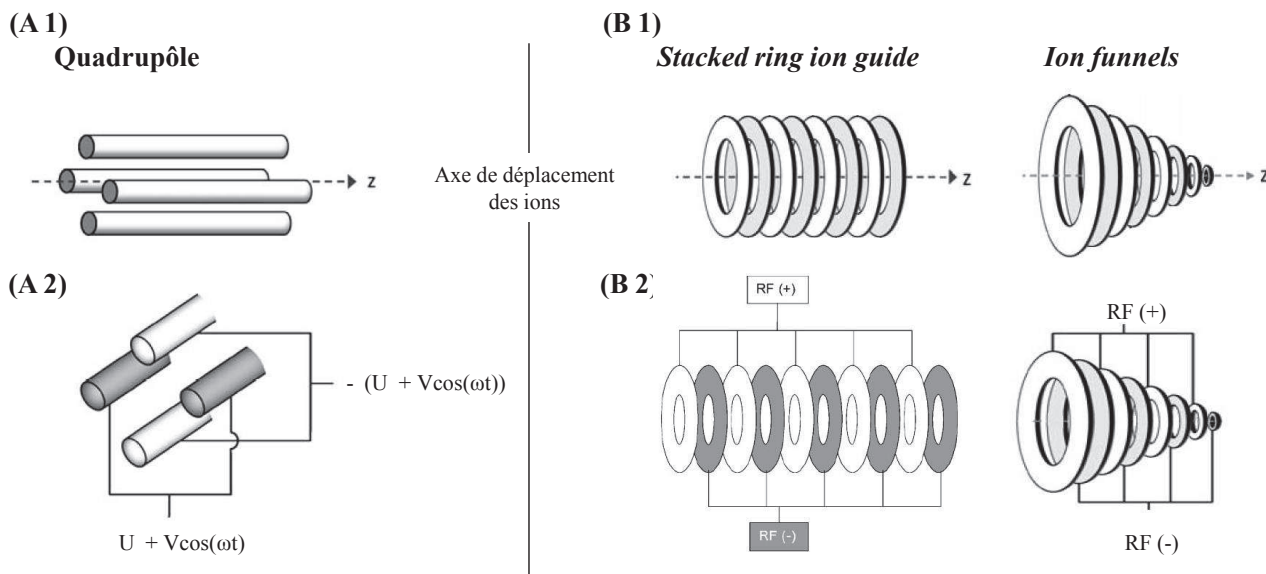


Figure 8. Différents types de guide d'ions servant d'interface de focalisation des ions issus de la source d'ion MS. (A 1, A2) guide d'ion en électrodes parallèles: quadrupôle, (B 1, B2) guide d'ions à lentilles concentriques de type SRIG "Stacked ring ion guide" ²¹ et IF "Ion funnels" ²⁴.

b. Les guides d'ions de type SRIG "Stacked ring ion guide" et IF "Ion funnels"

Ce sont des guides d'ions constitués d'une juxtaposition de lentilles annulaires de même diamètre interne ²¹ formant un tunnel coaxial à la trajectoire des ions (Figure 8, (B 1), SRIG *Stacked ring ion guide*) ou dont le diamètre interne des lentilles diminue progressivement dans le sens de progression des ions (Figure 8, (B 1), *Ion funnels*) ²²⁻²⁴.

Le principe de fonctionnement de SIRG et IF est basé sur l'application d'un potentiel de radiofréquence en opposition de phase sur les différentes lentilles afin obtenir une alternance de lentilles chargées positivement et négativement (Figure 8, (B 2)). La radiofréquence appliquée induit des vagues de champs autour des ions qui restent confinés au centre des lentilles qui les focalisent au fur et à mesure de leur avancement dans le spectromètre de masse. L'IF possède l'avantage de disposer d'une large ouverture permettant de collecter un maximum d'ions d'issus de la source et qui sont progressivement focalisés en un mince faisceau ²⁵.

3. Les analyseurs : trappe ionique, quadrupôle, TOF "Time of Flight"

L'analyseur permet la séparation, le déplacement, la sélection ou l'isolation des ions en phase gazeuse et cela grâce à des champs électriques ou magnétiques ou la combinaison des deux champs. Le rapport de la masse de l'ion sur sa charge (m/z) est le facteur dont va dépendre, la trajectoire, le mouvement et la sélection de l'ion dans un champ électrique ou magnétique. De plus l'analyseur doit être sous un vide poussé pour que le libre parcours moyen des ions leur permette d'atteindre plus facilement le détecteur avec un rendement suffisant.

Cinq types d'analyseurs sont actuellement utilisées pour l'analyse protéomique par spectrométrie de masse: analyseurs quadrupolaires (Q), trappe ionique (IT) ²⁶⁻²⁸, le temps de vol (TOF) ²⁹, la résonance cyclotronique ionique à transformé de fourrier (FT-ICR) ³⁰, et l'Orbitrap (OT) ³¹.

Pour l'analyse protéomique, les caractéristiques fondamentales d'un analyseur sont sa sensibilité, sa résolution, sa précision de mesure de masse, sa gamme de masse et sa capacité à produire des spectres de fragmentation assez informatifs ³². Le coût d'un spectromètre de masse est en partie lié aux analyseurs utilisés Certains spectromètres vont nécessiter la combinaison de plusieurs analyseurs pour tirer partie des avantages de chacun donnant ainsi des spectromètres de masse hybrides avec deux dimensions MS/MS. Cependant d'autres analyseurs de part leur géométrie permettent de réaliser successivement les expériences MS et MS/MS, c'est cas pour : IT, FT-ICR.

Trois types d'analyseurs ont été utilisés au cours de ce travail de thèse : quadripôle (Q), le temps de vol (TOF) et la trappe ionique (IT). Nous détaillons dans les paragraphes suivants ces trois types d'analyseur et leur combinaison pour la réalisation des expériences de MS à deux dimensions.

a. Analyseurs à une dimension MS : quadripôle et le temps de vol (TOF)

Le quadripôle (Q) et le temps de vol (TOF) sont deux analyseurs capables de réaliser une seule dimension MS. C'est-à-dire, ce sont des analyseurs qui vont servir à déterminer le m/z des ions qui y passent, à filtrer des ions sans les retenir dans leur progression vers le détecteur. Puisque c'est la sélection et le maintien de l'ion dans un analyseur qui va permettre par exemple une deuxième dimension MS (MS2 ou MS/MS) au cours de laquelle l'espèce retenue va être fragmentée, les Q et les TOF ne peuvent pas faire de MS/MS sans être associés à une cellule de collision à part entière. Cette raison fait que les Q et les TOF sont combinés pour donner une deuxième dimension de mesure aux spectromètres de masse pour faire la MS/MS : QqQ (triple quadripôle), Q-TOF (quadripôle – temps de vol), TOF-TOF (2 temps de vol)³³⁻³⁵.

i. Quadripôle (Q)

Le quadripôle a été développé dans les années 50 par Paul²⁷ et est aujourd'hui l'analyseur par excellence pour l'analyse du courant d'ion continu provenant d'une source ESI. Le principe de fonctionnement du quadripôle a été largement décrit par Campana^{36,37}. En dépit de son rôle d'interface de guide et focalisation des ions en mode "RF-Only", le quadripôle peut servir d'analyseur par balayage très rapide (quelques millisecondes) des potentiels appliqués entre ces électrodes (Figure 9, (A), les électrodes sur l'axe x et y) qui induisent un champ électrique variable entre celles-ci. Le balayage des potentiels va faire varier la stabilité de la trajectoire des ions entre les électrodes (Figure 9, (A)), pour les séparer en fonction de leur m/z . Les ions de différents m/z vont donc parvenir progressivement au détecteur selon que leur trajectoire soit stable ou instable suivant l'axe z (Figure 9, (A), l'ion \oplus) à un ratio U/V dans le quadripôle au fur et à mesure qu'évolue le balayage du potentiel (Figure 9, (B)). Les équations qui régissent la trajectoire des ions dans le quadripôle ont été établies par le physicien Mathieu (1966). La résolution de ces équations permet de déterminer la zone de stabilité des ions en fonction de U et V (Figure 9, (B), Diagramme de stabilité des ions). Les autres ions qui ont une trajectoire instable (Figure 9, (A), l'ion \oplus) vont être déviés, et ne parviennent pas au détecteur.

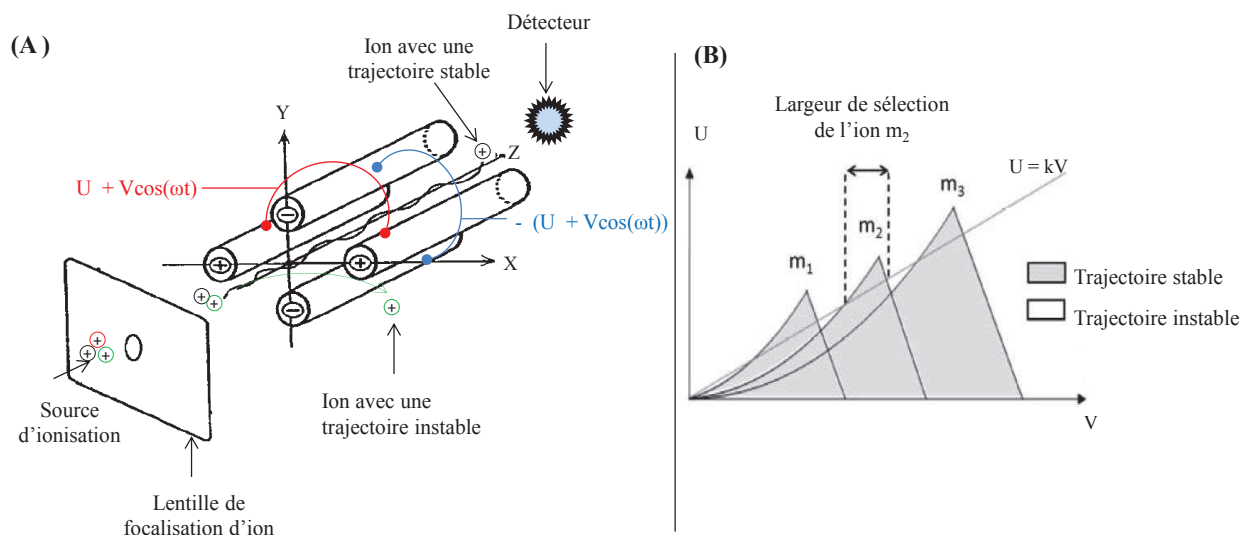


Figure 9. Représentation schématique du fonctionnement d'un analyseur quadripolaire (A) avec la représentation de la trajectoire d'un ion stable \oplus allant jusqu'au détecteur sur l'axe z et un ion avec une trajectoire instable \oplus sortant de l'axe z. (B) Diagramme de stabilité des ions de masse $m_3 > m_2 > m_1$ en fonction des tensions U et V . (Figures adaptées de Campana et al.³⁶).

ii. Le temps de vol (TOF)

Le principe de séparation des ions selon leur rapport m/z en fonction de leur temps de vol date des années 1950²⁹, mais ce n'est qu'avec l'arrivée de la source MALDI en 1987 et les avancées technologiques réalisés sur ce type d'analyseur que celui-ci va être équipé plus de spectromètres de masse.

L'analyseur à temps de vol, TOF "Time of Flight" permet après une accélération des ions par un champ électrique (tension d'accélération U [V]), de les séparer dans un tube de vol libre de champ sous un vide poussé et de mesurer leur rapport m/z en fonction du temps (t [s]) qu'ils mettent pour atteindre le détecteur²⁹.

L'énergie acquise au départ par l'ion ($E_1 = zeU$; U tension électrique [V], ze charge de l'ion [C]), se convertit totalement en énergie cinétique ($E_2 = 1/2mv^2$, m masse de l'ion et v la vitesse de l'ion) qui va contrôler l'avancement de l'ion dans le tube de vol. La trajectoire ou le temps mis par chaque ion (m, z) pour traverser le tube de vol d'un point à un autre va dépendre de sa vitesse (v [$m \cdot s^{-1}$]) obtenue à partir de la conservation de l'énergie: $v = (2zeU/m)^{1/2}$. Le temps de vol t [s] de cet ion sur une distance L correspondant au tube de vol est donné par la relation : $t = L/(m/2zeU)^{1/2}$. Ainsi pour des ions de même charge, ce sont les plus lourds qui mettent le plus de temps pour parvenir au détecteur. Les ions sont donc séparés en fonction de leur m/z . La détermination du rapport m/z de chaque ion, découle du temps de vol mesuré avec précision et de la connaissance des paramètres expérimentaux (position du détecteur L , la tension d'accélération U , e la charge élémentaire de l'électron).

Les premiers analyseurs TOF ont été pendant longtemps handicapés par leur faible résolution, imputable à la distribution temporelle (temps de formation des ions), spatiale (lieu de la formation) et cinétiques (énergie de formation) des ions. Deux améliorations majeures ont permis de résoudre ces problèmes.

- La dispersion temporelle et spatiale ont pu être éliminées grâce à l'utilisation du principe d'extraction retardée (Delayed Extraction)³⁸ à l'aide de lentilles, supplémentaires insérées entre la source d'ion et le TOF. L'application d'un potentiel nettement plus élevé à cette lentille permet de retarder la course des ions en provenance de la source vers le TOF. Ainsi les ions sont non seulement focalisés mais ont un départ synchronisé avant leur pénétration dans le tube de vol.
- La dispersion cinétique a pu être corrigée grâce à la mise au point du "réflectron". Le "réflectron" est un dispositif composé d'une série d'anneaux ou de grilles sur lesquels sont appliqués des potentiels croissants et qui vont faire office de miroirs électrostatiques. Ce dispositif renvoie les ions vers le détecteur en fonction de l'énergie cinétique qu'ils possèdent à leur arrivée dans le dispositif. Les ions de même m/z se retrouvent donc à nouveau focalisés et atteignent le détecteur au même moment (Figure 10).

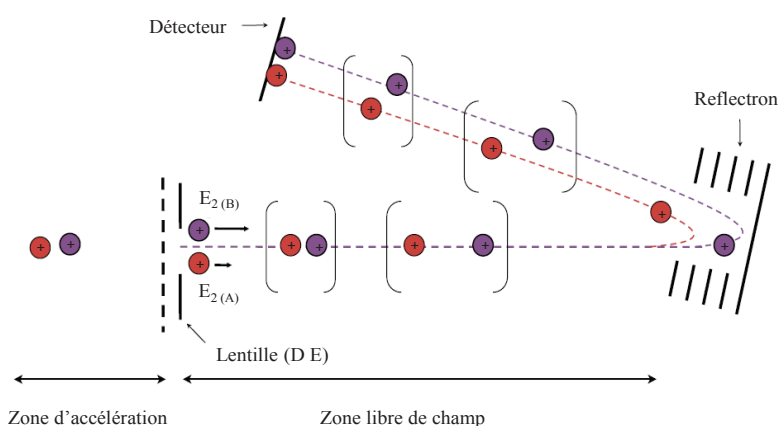


Figure 10. Représentation schématique du fonctionnement d'un analyseur à temps de vol (TOF) possédant la lentille (D E) le dispositif d'extraction retardée de ion et un "réflectron" le dispositif de compensation de la dispersion cinétique. Les ions (●●) sont de même m/z .

On note aujourd'hui plusieurs types de TOF suivant la dimension du tube de vol (par exemple $L = 250$ cm pour le maXis (Bruker)) et suivant la technique de focalisation des ions dans le tube de vol (par exemple 2 "réflectrons" pour la géométrie W du Synapt HDMS Waters).

Les spectromètres récents de type Q-TOF possèdent un quadrupôle en amont qui sert de filtre d'ions. La capacité du quadrupôle à permettre une bonne sélectivité de l'ion est donc sollicitée pour compléter la bonne résolution que possède désormais un TOF. Avec une cellule de collision intercalée entre les deux analyseurs, le quadrupôle permet la sélection d'ions pour une expérience MS de deuxième dimension (MS/MS), et le TOF d'analyser en suite les ions fragments issus de la cellule de collision. Les Q-TOF possèdent une excellente résolution (en moyenne comprise entre 8000 et 40000) en fonction de l'instrument et ont une bonne précision de masse surtout lorsqu'ils sont équipés de systèmes de calibration interne. Ce type d'instrument est un excellent compromis entre précision de masse, résolution, gamme de masse, vitesse de scan, sensibilité et est donc parfaitement adapté pour l'analyse protéomique.

b. Analyseur à plusieurs dimensions MS : la trappe ionique

La trappe ionique 3D découle des travaux de Paul ²⁷ sur le quadropôle qui ont conduit par la suite au développement de la trappe ionique constituée à la différence du quadropôle de 3 électrodes ³⁹. L'aboutissement de la première trappe ionique commercialisée en 1983 a été possible grâce aux travaux de Stafford et Finnigan ⁴⁰. Depuis, de nombreux groupes ont apporté des améliorations à cet analyseur en matière de sensibilité, résolution et de gamme de masse ⁴¹.

La trappe ionique est composée de 3 électrodes, une électrode annulaire en forme de diabolo et deux électrodes chapeaux de géométrie quasi hyperbolique (Figure 11, (A)) qui délimitent une cavité d'environ 1 cm². Les ions entrent et sortent de la trappe par des orifices situés dans les électrodes chapeaux. A la différence des analyseurs précédemment décrits, les trois électrodes vont permettre de confiner les ions à l'intérieur de la trappe au sein de laquelle règne une pression résiduelle d'hélium ($5 \cdot 10^{-5}$ mbar) contribuant à l'efficacité de piégeage des ions (Figure 11, (B)). Une fois dans la trappe, les ions vont subir un champ électrique quadropolaire grâce à la tension oscillante appliquée à l'électrode annulaire ($V \cos(\omega t)$), et leur trajectoire se stabilise dans un puits de potentiel dont la profondeur est directement liée à l'amplitude du courant appliquée (Figure 11, (B)). Les ions vont, alors décrire une trajectoire sous forme de 8 dans l'espace (courbe de Lissajou).

L'éjection des ions de la trappe se fait de deux façons : (1) par diminution de la profondeur du puits de potentiel donc la réduction de l'amplitude de la tension appliquée sur l'électrode annulaire (éjection simple) ou (2) par l'application d'un champ électrique sur les électrodes chapeaux, qui va déstabiliser la trajectoire des ions confinés vers la sortie (éjection raisonnante) (Figure 12). Lors d'une expérience MS de première dimension (MS), l'éjection d'ions se fait par variation de l'amplitude de la tension appliquée sur l'électrode annulaire (Figure 12 (A)) (éjection simple).

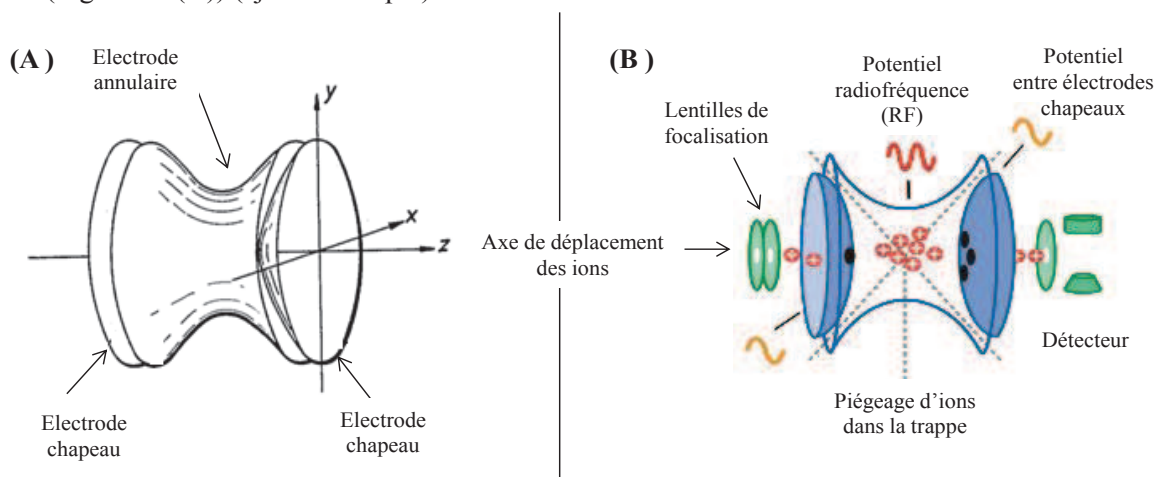


Figure 11. Représentation schématique de la trappe ionique (3D). (A) schéma des composantes d'une trappe ionique ³⁹, (B) l'application du potentiel radiofréquence sur l'électrode annulaire permet de piéger les ions sur des trajectoires stables dans la trappe. L'analyse est réalisée par déstabilisation des trajectoires des ions en appliquant un potentiel sur les électrodes chapeaux ou en réduisant l'amplitude du potentiel de l'électrode annulaire.

Cet analyseur va permettre de faire parvenir les ions en provenance de la source d'ions après un cours instant d'accumulation dans la trappe, mais aussi de sélectionner un ion (ion parent) lors d'une expérience MS de première dimension, en vu d'une expérience MS/MS. L'ion de m/z sélectionné, toujours en mouvement sur sa trajectoire, va subir une excitation avec l'application d'un potentiel contrôlé sur les électrodes chapeaux qui va entraîner son entrée en collision avec les atomes d'hélium et provoquer ainsi sa fragmentation (génération des ions fils). Les ions fils issus de la fragmentation vont pouvoir à leur tour être accumulés et éjectés de la trappe vers le détecteur (Figure 12, (B)).

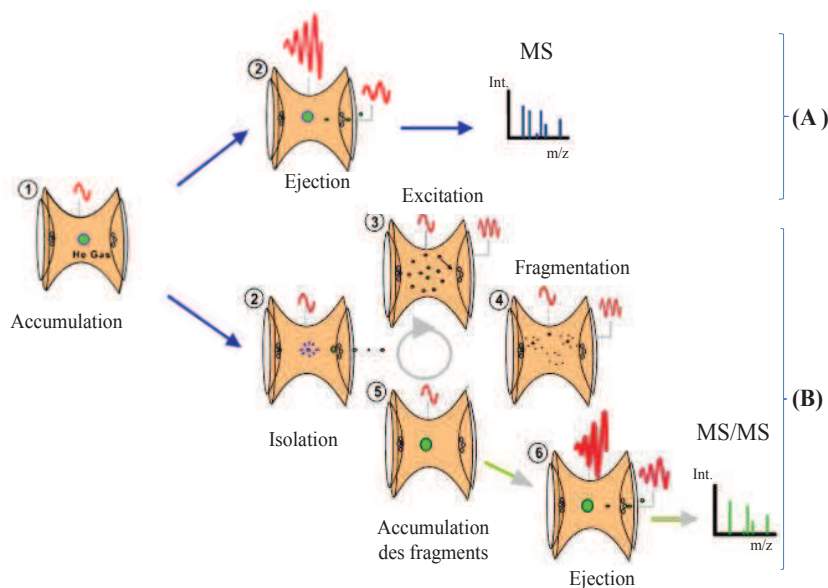


Figure 12. Principe d'isolation, fragmentation et éjection d'ions d'une trappe ionique 3D lors des expériences MS à une dimension (A) et multidimensionnelles MSⁿ (B).

Les expériences MS de troisième dimension peuvent ainsi être réalisées sur des ions fils et ainsi de suite jusqu'à n dimensions permettant ainsi d'avoir des informations de plus en plus détaillées sur l'ion parents d'origine.

Cet analyseur à la fois robuste, sensible et de coût modéré est un instrument de choix pour l'analyse protéomique. Cependant, la précision de mesure de masse est limitante pour certaines applications en protéomique comparé au Q-TOF. En effet, la capacité de stockage des ions dans la trappe est limitée à cause des répulsions coulombiennes trop fortes qui induisent un effet d'espace-charge perturbant ainsi les champs électriques appliqués et donc la bonne estimation des valeurs m/z . Plusieurs améliorations ont été apportées à la première génération de trappe ionique notamment dans le dimensionnement des nouvelles trappes et la proposition de nouvelle géométrie de trappe à l'instar de la trappe linéaire (2D) avec une meilleure capacité de stockage d'ions^{41,42}.

En analyse protéomique, il est important de choisir la bonne combinaison entre source et analyseur pour avoir la meilleure configuration possible du spectromètre de masse pour l'analyse. Il est donc primordial de garder à l'esprit que le choix d'un spectromètre de masse dépendra fortement de la stratégie choisie pour l'identification des protéines en fonction de la question biologique posée.

4. Détecteurs

Comme les analyseurs et les sources d'ionisation, il existe plusieurs types de détecteurs avec des principes de fonctionnement physiques différents. Quelque soit son mode de fonctionnement, le détecteur va compter les ions par conversion de l'impact d'un ion sur une surface, ou son arrivée dans une enceinte en un signal électrique, qui se traduit par un pic (m/z) au quel est associé une intensité. L'enregistrement de plusieurs événements en un laps de temps connu constitue le spectre de masse. Le détecteur est généralement placé sous un vide poussé ($10^{-5} - 10^{-7}$ Torr). Le choix du détecteur est surtout lié à sa capacité à détecter des événements (des ions) suffisamment rapprochés dans le temps et à sa durée de vie.

Le **cylindre de Faraday** est le détecteur le plus simple utilisé en spectrométrie de masse. Dans ce détecteur, le transfert de charge de l'ion est détecté sur une surface conductrice, puis le signal est amplifié. Ce détecteur est précis mais est peu sensible, lent, et possède un bruit de fond assez important, ce qui fait il est rarement utilisé sur les spectromètres de masse dédié à l'analyse protéomique.

Les **multiplicateurs d'électrons (CEMs, Channel Electron Multipliers (CEMs))** sont des détecteurs les plus courants. Dans ce type de détecteur, le signal est amplifié par la formation d'électrons secondaires à l'aide de tube de verre dopé avec un métal ou un semi-conducteur comme le plomb (**dynode**). Le tube de verre peut être linéaire (Linear channel electron multiplier) ou recourbé (curved channel electron multiplier). Ce type de détecteur possède une bonne sensibilité avec une amplification forte du signal cependant avec une durée de vie limitée. On trouve ce type de détecteur sur des spectromètres de type trappe ionique en général (ex : amaZon, Bruker).

La **galette de microcanaux (MCP, *Microchannel plate*)** est un autre détecteur qui peut être considéré comme un assemblage de multiplicateurs d'électrons. La collision d'ions avec une anode chargée, entraîne une cascade d'électrons dans les microcanaux. Et pour augmenter le rendement d'amplification de signal, ces galettes de microcanaux peuvent être superposées. Cette cascade d'électrons est alors détectée par un circuit électrique (Figure 13)⁴³. Cette impulsion de charge électrique peut être enregistrée au moyen d'un convertisseur de type TDC (*Time to digital converter*) par exemple. Le convertisseur TDC enregistre l'arrivée précise d'un ion à un temps précis sous forme de signal. Les instruments de type Q-TOF ont généralement ce type de détecteur (ex : le Synapt HDMS, Waters).

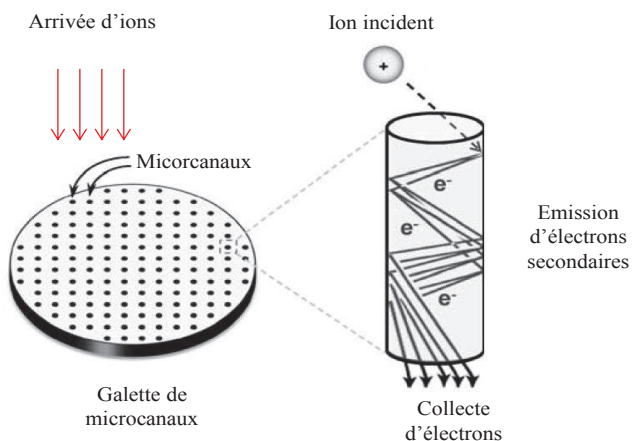


Figure 13. Principe de fonctionnement d'un détecteur MCP.

5. Description des spectromètres utilisés

a. La trappe ionique : amaZon-ETD (Bruker, Daltonics)

Ce spectromètre de masse est une nouvelle génération de trappe ionique avec une grande capacité d'accumulation d'ions, équipé d'une source ESI, d'une interface avec un capillaire de transmission et deux séries de lentilles de focalisation d'ions (*ion funnels*), des octopôles servant de guides d'ions jusqu'à la trappe ionique (3D) et un détecteur de type CEM (Figure 14). A la différence avec l'ancienne génération (HCT Ultra) du même constructeur, les *skimmer* (écrémeurs) ont été remplacés par des lentilles tunnels (*ions funnels*), ce qui a considérablement amélioré la transmission mais également permis de focaliser plus finement les ions arrivant jusqu'à la trappe ionique.

La vitesse de balayage a été également améliorée. Le mode "UltraScan" qui est utilisé pour la génération des spectres MS/MS a vu sa vitesse de balayage passer de 26000 (m/z)/s à 32500 (m/z)/s sur l'amaZon. Ce qui permet donc d'aller plus vite dans l'acquisition des spectres MS et MS/MS. Ainsi nous avons pu noter au laboratoire une amélioration de la sensibilité d'un facteur 10 par rapport à l'ancienne génération HCT Ultra. La gamme de masse de l'amaZon est de 50-3000 m/z avec des vitesses d'acquisition pouvant aller jusqu'à 20 spectres/s pour une précision de mesure de masse de l'ordre de 0,2 Da et une résolution pouvant atteindre 5000 (FWHM). Ce spectromètre de masse possède des performances analytiques en termes de précision, sensibilité et rapidité d'acquisition permettant des expériences MS et MS/MS sur des mélanges complexes.

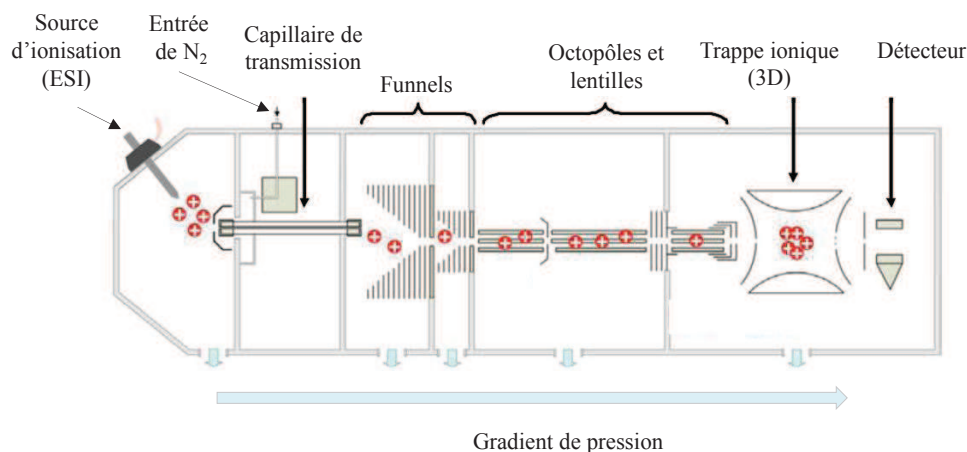


Figure 14. Schéma de la trappe ionique (3D) amaZon ETD (Bruker Daltonics).

b. Les instruments de type Q-TOF: maXis et microTOF-Q (Bruker, Daltonics) et Q-TOF Synapt HDMS (Waters, Corporation)

i. Le maXis et microTOF-Q (Bruker, Daltonics)

Le maXis et le microTOF-Q sont des instruments de type Q-TOF équipés d'une source électrospray (ESI), avec une interface composée de capillaire de transmission, des lentilles (*ion funnels*) et des hexapôles servent de focalisation et de guide d'ions (Figure 15). Un quadrupôle directement placé en amont de la cellule de collision sert de filtre d'ions. Entre la cellule de collision et le TOF, est intercalé sur le maXis un "*Ion cooler*" dont le principal rôle est la re-focalisation des ions vers le TOF donc plus de sensibilité de détection d'ions.

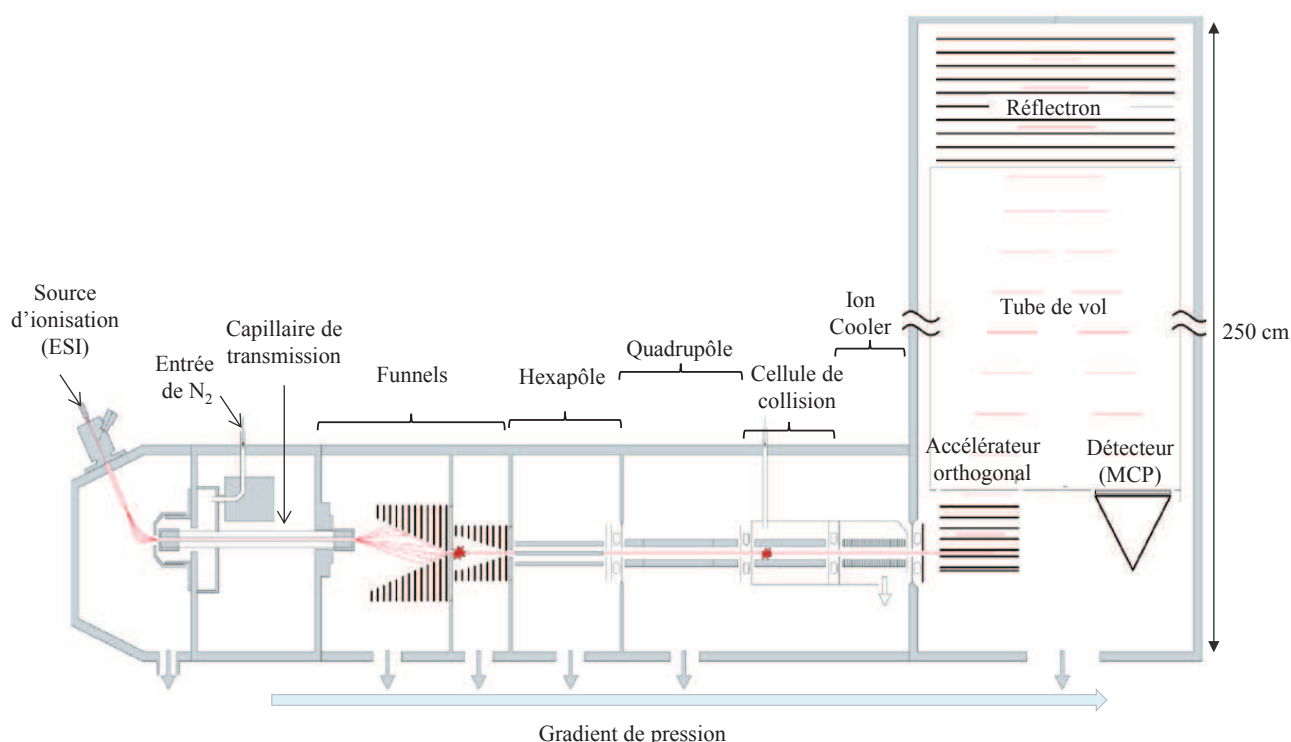


Figure 15. Schéma du maXis (Bruker Daltonics).

Cet *ion cooler* n'existe pas sur le microO-TOF qui a une taille beaucoup plus réduite notamment dans la dimension de son tube de vol comparé à celui du maXis. Le maXis possède ainsi une résolution de masse comprise entre 40000 - 60000 (FWHM), une précision de l'ordre de la ppm en MS et MS/MS et une vitesse de 20 spectres/s (20 GHz) contre une résolution de 15000 (FWHM) et une précision comprise entre 3 et 5 ppm en MS et MS/MS pour le microTOF-Q.

ii. Le Synapt HDMS (Waters, Corporation)

Le synapt HDMS (*High Definition Mass Spectrometry*) est l'instrument de type Q-TOF qui offre la possibilité de faire des expériences de MS à deux dimensions et la mobilité ionique. Il est équipé d'une source d'ionisation ESI en forme de Z (Z-Spray) augmentant l'efficacité de désolvatation du spray, de lentilles de focalisation situées en amont d'un quadropôle qui sert de filtre d'ions et de cellules qui peuvent servir de cellules de collision (Trap, Transfert) et d'une cellule de mobilité ionique (Figure 16). Au cours des expériences MS/MS, la cellule de collision immédiatement située après le Q est utilisée pour fragmenter les ions tandis que la cellule de mobilité et la cellule de transfert servent de guides et focalisation d'ions avant leur entrée dans le tube de vol. Le tube de vol du Synapt HDMS à deux modes de fonctionnement : le mode V (trajectoire en traits plein en rouge) et le mode W (trajectoire en trait discontinu) où la trajectoire des ions est deux fois plus longue ce qui permet à l'instrument de passer d'une résolution de 10000 à 17500 (FWHM).

Le synapt HDMS possède une gamme de masse allant jusqu'à 32 kDa et une vitesse d'acquisition de 10 spectres/s avec une précision de l'ordre de 2 ppm adaptée pour l'analyse protéomique.

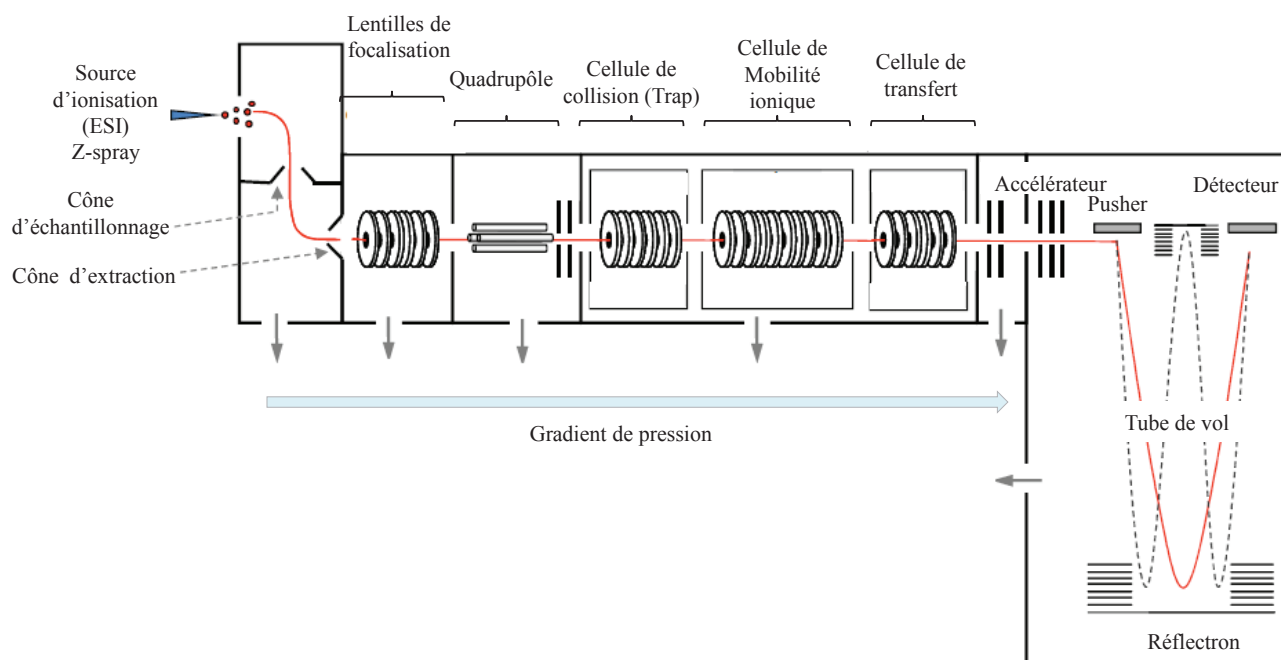


Figure 16. Schéma du Synapt HDMS (Waters Corporation).

6. De l'ion à la séquence protéique : la fragmentation peptidique

Un m/z mesuré par spectrométrie de masse en analyse protéomique peut correspondre à une séquence d'acides aminés (un peptide) d'une protéine ou une autre molécule contenue dans l'échantillon analysé. Or la simple mesure de masse de m/z dans un échantillon complexe peut conduire à plusieurs structures différentes. Par exemple, les peptides, SSLVULK, SLLVSUK, LLVSSUK, au tant de combinaisons possibles avec les mêmes acides aminés, vont conduire aux mêmes rapports m/z ⁴⁴. La fragmentation MS/MS permet de casser les liaisons entre acides aminés en fragments identifiables pour déterminer la succession des acides aminés d'un peptide : le séquençage des protéines ou peptides. Bien sûr cette fragmentation peut se faire sur des protéines ou des peptides. Mais, la fragmentation de protéines (stratégie appelée Top-down) est plus complexe à interpréter et à cela s'ajoute la limite de résolution d'un instrument et donc la mesure de masses imprécises. L'alternative est l'analyse des "morceaux ou bouts" de ces protéines (peptides) après une digestion enzymatique (stratégie appelée Bottom-up), ce qui va donc donner pour chaque peptide des fragments facilement interprétables. Pour ces raisons, la deuxième approche (Bottom-up) est la plus utilisée en analyse protéomique par spectrométrie de masse.

La fragmentation des peptides et des protéines, peut se faire en mode CID (dissociation induite par collision, le mode de fragmentation le plus utilisé), en mode ETD (dissociation par transfert d'électron) ou en ECD (dissociation par capture d'électron). En fonction du mode de fragmentation choisi, la nature du peptide et la géométrie du spectromètre de masse, on va obtenir préférentiellement des fragments spécifiques différents. Le mode de fragmentation utilisé lors de ces travaux de thèse est le CID.

a. La fragmentation peptide : dissociation induite par collision (CID)

La fragmentation des ions est réalisée par collision avec des molécules neutres (Hélium, Argon, azote). On distingue deux types de fragmentation en mode CID⁴⁵ : la fragmentation à basse énergie (quelques eV) sur les instruments de type Q-TOF et trappe ioniques, et la fragmentation à hautes énergies sur les instruments de type TOF-TOF.

Nomenclature de fragmentation des peptides

La nomenclature de la fragmentation de peptide est issue des travaux de Biemann dans les années 90⁴⁶. Deux types d'ions sont observés lors de la fragmentation de peptides : (1) ceux pour lesquels la charge positive est portée par la partie N-terminale (les séries a, b, c) et (2) ceux pour lesquels la charge positive est portée par la partie C-terminale (les séries x, y, z) (Figure 17).

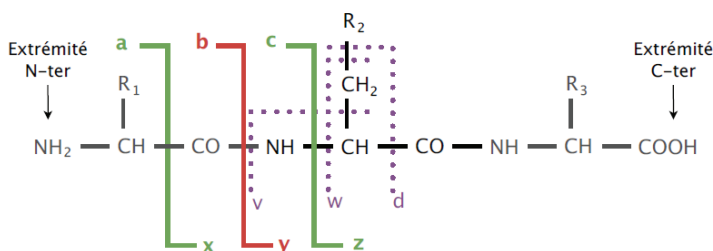


Figure 17. Nomenclature de fragmentation des peptides selon Biemann. R₁, R₂, R₃ sont des chaînes latérales des acides aminés. Les ions des séries (a, b, c et x, y, z) sont obtenus en basse énergie alors que les séries (v, w, d) supplémentaires s'obtiennent à haute énergie.

Les séries b et y émanant de la fragmentation des liaisons peptidiques entre acides aminés beaucoup plus labiles que les autres liaisons, sont généralement majoritaires. Les différences de masse entre des ions consécutifs issus d'une même série permettent de déterminer l'identité des acides aminés successifs et donc de reconstruire la séquence du peptide fragmenté.

b. Modèle de fragmentation : proton mobile

Le transfert d'énergie par activation collisionnelle, implique une délocalisation du proton à différents endroits du squelette peptidique. La présence du proton initie la fragmentation au niveau des liaisons peptidiques (liaisons amides CO-NH) pour former des ions de séries b et y. Cette fragmentation à basse énergie, est dite dirigée par la charge (*charge-directed fragmentation*) et selon le modèle de proton mobile⁴⁷.

Pour un peptide monochargé, la présence d'acides aminés basiques (Arginine, Lysine) dans sa séquence va augmenter considérablement l'énergie nécessaire à sa fragmentation puisque, le proton est séquestré sur la chaîne latérale de l'acide aminé basique.

Pour des peptides tryptiques doublement chargés, les deux charges sont généralement portées par l'extrémité N-terminale et par la chaîne latérale de la Lysine ou Arginine C-terminale qui séquestre un proton. La barrière énergétique nécessaire 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. Ceci va favoriser la génération des ions b et y. Les ions y sont souvent plus intenses puisque le proton séquestré par le résidu C-terminal augmente leur stabilité par rapport aux ions b.

Dans les approches protéomiques, mettant en œuvre des digestions aspécifiques comme la pepsine, les peptides – non tryptiques – subissent des fragmentations moins régulières du fait de la non-présence systématique d'acides aminés N- et C-terminaux basiques comme dans le cas de la trypsine. Les spectres de fragmentation sont dans ce cas difficiles à interpréter ou peu informatif. C'est pourquoi la trypsine est l'enzyme de digestion la plus utilisée en analyse protéomique par spectrométrie de masse.

II. Techniques de séparation et de purification protéiques

L'analyse protéomique se heurte à des difficultés qui sont principalement liées à la complexité des échantillons, aussi que la gamme dynamique assez étendue au niveau des teneurs en protéines dans un échantillon. En effet au niveau d'une seule famille de protéine, on peut noter plusieurs isoformes qui elles-mêmes peuvent être affectées par plusieurs modifications post-traductionnelles, ou encore ces protéines avoir des concentrations différentes. La digestion enzymatique des protéines en peptides va rendre encore plus complexe les échantillons, puisque ces protéines vont générer des centaines voire des milliers de peptides à analyser. Par exemple la banque protéique humaine, compte à elle seule plus 52816 protéines dans la banque IPI (version 2.23). Ces protéines vont générer plus de 89.10^4 peptides tryptiques théoriques d'après la digestion *in silico*. Le spectromètre de masse ne peut pas permettre la caractérisation des séquences de chaque peptide contenu dans ces types de mélanges complexes en même temps. Le fractionnement des échantillons avant leur analyse par spectrométrie de masse est donc nécessaire pour que la complexité soit réduite. On utilise alors différentes techniques de séparation des protéines ou peptides ou encore des techniques de purification ciblant spécifiquement une catégorie de protéines ou peptides (Figure 18).

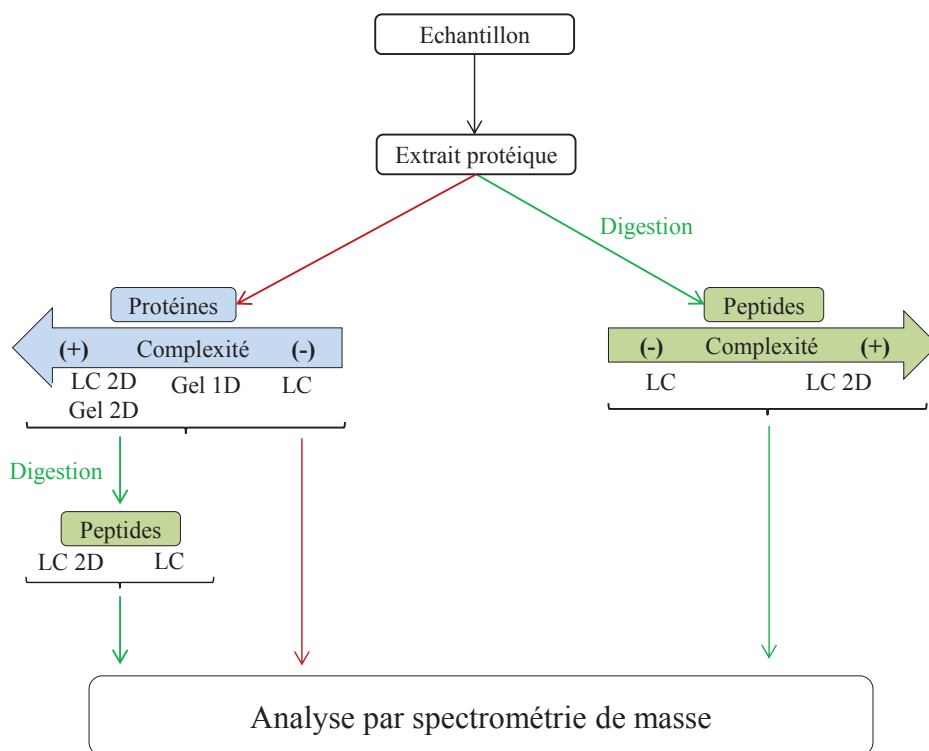


Figure 18. Principales techniques de séparation et purification des protéines et peptides utilisées en analyse protéomique suivant la complexité de l'échantillon. La séparation de protéines se fait par LC, Gel 1D, LC 2D ou Gel 2D et la séparation des peptides par LC ou LC 2D en fonction de la complexité des échantillons.

Les principales techniques de séparation et purification des protéines et/ou peptides utilisés en analyse protéomique sont basées sur l'électrophorèse sur gel (Gel 1D et 2D) et la chromatographie liquide (LC) (de phase inverse et colonne d'affinité). Suivant la complexité de l'échantillon ces techniques peuvent être combinées en plusieurs étapes successives. Il existe plusieurs variantes de ces techniques, mais nous nous limiterons dans les prochains paragraphes à celles utilisées au cours de cette thèse.

1. Séparation des protéines par électrophorèse sur gel : gel 1D et gel 2D

L'électrophorèse est l'une des principales techniques utilisées en biochimie pour la séparation et la caractérisation des biomolécules (protéines). Cette technique est fondée sur le déplacement d'ions (protéines, peptides chargés) sous l'effet d'un champ électrique. En fonction de leur caractéristique physico-chimique propre et les conditions d'électrophorèse, ces molécules chargées acquièrent des vitesses de migration différentes et finissent donc par être séparées les unes des autres. Deux variantes d'électrophorèse sont couramment utilisés en protéomique : l'électrophorèse capillaire et sur gel. Malgré l'efficacité de l'électrophorèse capillaire en termes de séparation de protéines, celle-ci est restée très peu utilisée par rapport à

l'électrophorèse sur gel à cause de sa faible capacité de chargement et l'adsorption des protéines sur les parois du capillaire (phénomène hystérèse). La séparation de protéines par électrophorèse est donc généralement réalisée sur gel, le plus souvent de polyacrylamide. Ces gels offrent l'avantage majeur de permettre une séparation préalable des protéines et sont compatibles avec les traitements échantillons en amont de l'analyse par spectrométrie de masse qui peuvent être nécessaires (réduction, alkylation, digestion). De plus la révélation des protéines à l'aide des colorants (Bleu de Coomassie, Argent) permet une estimation de l'abondance relative des protéines.

a. Electrophorèse sur gel monodimensionnelle (gel 1D)

Cette technique découle des travaux de Laemmli dans les années 70^{48, 49}. Le gel SDS-PAGE (Sodium DodecylSulfate)-(PolyAcrylamide Gel Electrophoresis) ou encore gel 1D, est une électrophorèse sur gel en condition dénaturante avec le Sodium Dodecyl Sulfate (SDS). Le SDS se fixe sur les protéines, les dénature et leur confère des charges négatives qui vont masquer complètement les charges natives des protéines. La densité de charge étant sensiblement constante pour toutes les protéines dans le mélange échantillon - SDS, la séparation dans le champ électrique est donc essentiellement faite en fonction de la masse moléculaire des protéines. En outre, la charge négative de protéines due au SDS favorise la solubilisation des protéines^{50, 51}. La taille des protéines étant très variée dans un échantillon complexe, il est souvent intéressant d'optimiser la résolution en fonction de l'échantillon en : (1) modifiant la taille des pores du gel (liée à la concentration de l'acrylamide), (2) modifiant les tampons d'électrophorèse (pH, par exemple). Cependant il faut préciser que cette technique, peut présenter des limites en termes de résolution pour des échantillons complexes (Figure 19 (A)) par rapport au gel 2D.

Le gel 1D est un excellent compromis entre solubilité des protéines et résolution de séparation. Cette méthode de séparation est très appréciée pour l'analyse de protéines membranaires qui sont généralement difficiles à solubiliser⁵⁰. Elle permet par exemple la séparation des protéines après une purification ciblant spécifiquement des protéines d'intérêts (par exemple la chromatographie d'affinité) pour l'éliminer non seulement les solvants incompatibles avec la MS, mais pour servir d'étape de préfractionnement pour des mélanges relativement peu complexes avant l'analyse par spectrométrie de masse (cf. Résultat, Chapitre 1).

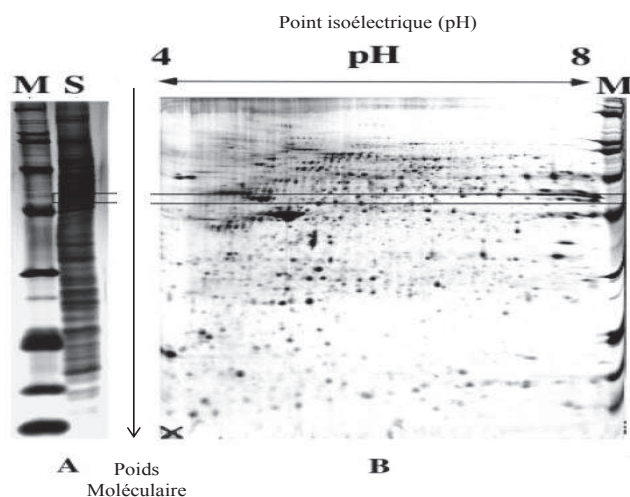


Figure 19. Comparaison de l'électrophorèse sur gel 1D et 2D d'extrait protéiques de mitochondries humaines. (A) Gel 1D : la piste M porte les marqueurs de poids moléculaires, la piste S porte l'échantillon, (B) gel 2D⁵². Le gel 2D présente une meilleure résolution de séparation des protéines par rapport au gel 1D par exemple entre les deux lignes parallèles sur les gels.

b. Electrophorèse sur gel bidimensionnelle (gel 2D)

Le gel 2D est un couplage de l'isoélectrofocalisation (IEF, Isoelectric Focusing) dénaturante avec l'électrophorèse monodimensionnelle gel 1D qui offre deux dimensions de séparation pour les protéines.

La première dimension est une migration horizontale des protéines en fonction de leur point isoélectrique (pI, Figure 19, (B)). Cette première dimension de séparation est mise en œuvre sur une bandelette de gel présentant un gradient de pH et sur laquelle vont migrées les protéines sous l'effet d'un champ électrique^{53, 54}. Ainsi les protéines sont séparées sur le gel en fonction de leur point isoélectrique et s'immobilisent aux endroits où le pI de chacune d'elles a par définition sa charge globale nulle. Les protéines groupées par pI vont subir une deuxième dimension de séparation orthogonale à la première qui est une séparation verticale selon le poids moléculaire des protéines (SDS-PAGE).

Outre sa meilleure résolution par rapport au gel 1D et l'indication sur la quantité relative de chaque protéine, le gel 2D permet de séparer des isoformes et des protéines affectées par des PTMs (phosphorylation, glycosylation...) ⁵².

L'optimisation des deux dimensions de séparation des protéines permet selon l'échantillon d'améliorer la résolution de séparation. On peut étendre la gamme de pH pour avoir une meilleure résolution horizontale (selon l'axe des pI Figure 19 (B)), ou alors modifier la réticulation du gel, la longueur du gel ou le temps de migration pour avoir une meilleure séparation possible.

Cette technique est préférable au gel 1D pour l'identification des protéines sur- ou sous-exprimées (cf., Résultats, Chapitre 2) dans un protéome car elle offre une meilleure résolution et détection grâce à l'intensité des spots analysables par des logiciels d'analyse d'images de gels à l'instar de PD Quest (Bio-Rad Laboratories Inc.), ProgenesisTM (Nonlinear Dynamics Ltd) ⁵⁵.

2. Séparation des protéines et peptides par chromatographie liquide

La chromatographie liquide est une technique de séparation très utilisée en protéomique, car elle permet une séparation rapide de macromolécules (protéines et peptides) et possède une haute résolution et sensibilité appréciable pour la détection des protéines en faible quantité ⁵⁶.

De par son principe, la chromatographie liquide est une séparation des analytes qui s'opère entre deux phases : une phase mobile (mélange de solvants) et une phase stationnaire (par exemple un solide divisé, un gel, une résine). En analyse protéomique, les techniques de séparation chromatographie couramment utilisées sont, la chromatographie d'exclusion stérique (SEC, Size Exclusion Chromatography), la chromatographie d'échange d'ions (IEC, Ion Exchange Chromatography), la chromatographie de phase inverse (RPLC, Reverse Phase Liquid Chromatography), la chromatographie d'interaction hydrophile (HILIC, Hydrophilic Interaction Chromatography) et la chromatographie d'affinité (Affinity Chromatography) ⁵⁶.

a. Séparation et purification des protéines par chromatographie liquide : LC

Pour faciliter l'analyse des protéines d'un mélange complexe par spectrométrie de masse, il est parfois nécessaire de les séparer dans une première étape en plusieurs fractions ou de les isoler et purifier sous forme de complexes protéiques d'intérêt. Plusieurs types de chromatographie liquide peuvent être utilisés dans ce contexte comme par exemple la SEC, IEC, la RPLC ou la chromatographie d'affinité. Ces techniques peuvent être combinées pour avoir plusieurs dimensions de séparation par exemple : IEC+RPLC ou SEC+RPLC ⁵⁶.

Nous nous limiterons dans les paragraphes suivants à la chromatographie de phase inverse (RPLC) et quelques variantes de la chromatographie d'affinité.

i. Séparation des protéines par chromatographie de phase inverse (RPLC)

La séparation des protéines d'un échantillon par RPLC peut être intéressante pour séparer et collecter une protéine afin de mieux la caractériser. Dans ce cas la séparation des protéines se fait généralement sur des colonnes analytiques (colonne C4, C8 ou C18, de diamètre interne 2,1 mm ou 4,6 mm, particules de 5 µm de diamètre). Ces colonnes possèdent de bonne capacité de chargement, efficacité de séparation et une bonne reproductibilité qui vont permettre la séparation des protéines et leurs collectes. De plus les solvants utilisés pour la séparation des protéines en RPLC (l'eau, acétonitrile, méthanol etc.) sont compatibles avec la spectrométrie de masse, ce qui permet donc le couplage entre les deux techniques: LC-MS (cf., Résultats, Chapitre 3).

ii. Séparation et purification des protéines par chromatographie d'affinité

La chromatographie d'affinité est particulièrement utilisée pour isoler ou enrichir des protéines souvent mineures comme par exemple les glycoprotéines, les protéines phosphorylées ou encore des complexes protéiques par immuno-affinité (à partir d'une protéine cible). En effet les protéines portant des PTMs, ou les complexes protéiques d'intérêt peuvent être mineures dans un milieu complexe, et leurs signaux en spectrométrie de masse être entièrement masqués par ceux des protéines majoritaires. La purification par immuno-affinité va dans ce cas, permettre de purifier et concentrer une protéine et ses partenaires d'interaction sous forme de complexes protéiques). Pour une chromatographie affinité classique, c'est une famille de protéines (glycoprotéines, protéines phosphorylées) qui va être enrichie. Ainsi l'échantillon est dé-complexifié par rétention de protéines d'intérêt sur la colonne d'affinité.

Il existe plusieurs catégories de chromatographie d'affinité. Selon les contextes, on distingue par exemple la chromatographie d'affinité pour l'enrichissement de glycoprotéines (lectine, concanavaleine A etc.), pour l'enrichissement des phosphoprotéines (IMAC (Immobilized Metal Ion Affinity Chromatography), TiO₂) et l'immuno-affinité pour l'enrichissement de complexes protéiques (immunoaffinity purification, epitope tagging, GST pulldown, TAP-tag (Tandem Affinity Purification tag))^{57,58}.

Aujourd'hui, on utilise de plus en plus l'immuno-affinité en protéomique pour l'étude des partenaires d'interaction, un enrichissement des complexes protéiques partenaires d'une protéine ou d'une molécule servant de cible. Ces techniques se servent de l'affinité spécifique entre une protéine, un peptide et la colonne d'affinité ou une enzyme et son substrat greffé sur la phase de la colonne ou des billes pour purifier les complexes protéiques dont fait parti la protéine cible.

Il existe deux grandes stratégies de purification de complexes protéiques par immuno-affinité : les méthodes de co-immunopurification et les méthodes de co-précipitation par affinité⁵⁹. Pour purifier alors spécifiquement une protéine et ses partenaires d'interaction, cette protéine va servir d'appât pour ces partenaires d'interaction.

Les méthodes de co-immunopurification

Les complexes protéiques sont purifiés directement du lysat cellulaire en utilisant un anticorps dirigé contre une protéine du complexe (Figure 20, (A)). Cette technique est directe, mais demande un anticorps spécifique à une protéine pour l'expérience, or ce n'est pas toujours le cas lorsqu'on explore pour la première fois un mécanisme biologique par exemple.

Les méthodes de co-précipitation par affinité

Cette méthode est une alternative à la première. A défaut d'un anticorps spécifique, on utilise les propriétés biochimiques d'une étiquette (peptide, protéine etc.) que l'on exprime sur la protéine cible constituant ainsi un appât pour les autres protéines du complexe (Figure 18, (B)). En effet la protéine cible est fusionnée (du côté N-terminale ou C-terminal de la protéine cible) avec l'étiquette. De façon générale, l'étiquette (tag) est choisie en fonction de la protéine cible (par exemple sa taille par rapport à la protéine cible) pour que la fusion ait une interférence minimale sur l'interaction protéine-étiquette (protéine-tag) et les partenaires à co-purifier. Le construit (le gène) est alors introduit *in vivo* dans la cellule cible ou la protéine-tag est mise en incubation *in vitro* avec le lysat cellulaire. La protéine-tag, est alors en contact avec d'autres protéines de l'organisme pour que les complexes multi-protéiques se forment. Pour purifier les complexes protéiques d'intérêts, on utilise la chromatographie d'affinité avec une colonne anti-tag, ou spécifique à l'étiquette. Lors d'une première chromatographie, la colonne va retenir donc spécifiquement les complexes protéiques dont fait partie la protéine-tag. Une deuxième chromatographie permettra d'éluer ("décrocher") les complexes d'intérêt. Il existe plusieurs sortes d'étiquette de taille et nature différentes, et chacune pouvant influencer plus ou moins la conformation de la protéine cible, ou entraîner la non viabilité de l'organisme modifié^{58,60}.

Deux étiquettes peuvent ensuite être utilisées pour bénéficier des avantages de chacune (Figure 20, (C)): TAP-tag (Tandem Affinity Purification tag), une double co-purification par affinité (cf., Résultats, Chapitre 1).

Les complexes protéiques élués peuvent ensuite être séparés sur gel 1D ou 2D avant la digestion pour l'analyse protéomique par spectrométrie de masse.

Les limites des techniques de co-purification par affinité

L'approche d'enrichissement des complexes protéiques par co-purification par affinité permet la purification de nombreux complexes certes, mais souvent avec une perte d'interactions transitoires les plus faibles due aux étapes de lavage⁶¹, des problèmes de contamination dus à la multiplication des sources lors des étapes de production de matériel sont souvent remarqués. Beaucoup d'interactions protéines-protéines, bien qu'elles soient importantes, présentent une faible affinité et sont transitoires ou encore dépendent fortement de l'environnement cellulaire dans lequel elles sont formées. Il est donc indispensable d'avoir des contrôles pour vérifier la spécificité des interactions des différentes protéines du complexe. Les expériences sont ainsi effectuées en plusieurs répliques en parallèle avec des expériences de contrôle pour éliminer les interférences liées au tag (interactions aspécifiques), puisque les tags ou les phases stationnaires de par leur nature peuvent eux-mêmes recruter d'autres protéines qui ne participent pas aux différents complexes.

L'utilisation d'autres méthodes de purification des protéines par pontage chimique (technique appelée, Cross-linking) à l'aide d'agents pontants comme le DSP (Dithiobis(succinimidyl protonate)) peuvent s'avérer intéressants pour lier les partenaires d'interaction entre eux *in vivo* avant la lyse cellulaire et l'ultime étape de

chromatographie d'affinité^{62, 63}. Cette méthode permet de maintenir l'édifice du complexe protéique intact (cf., Résultats, Chapitre 1).

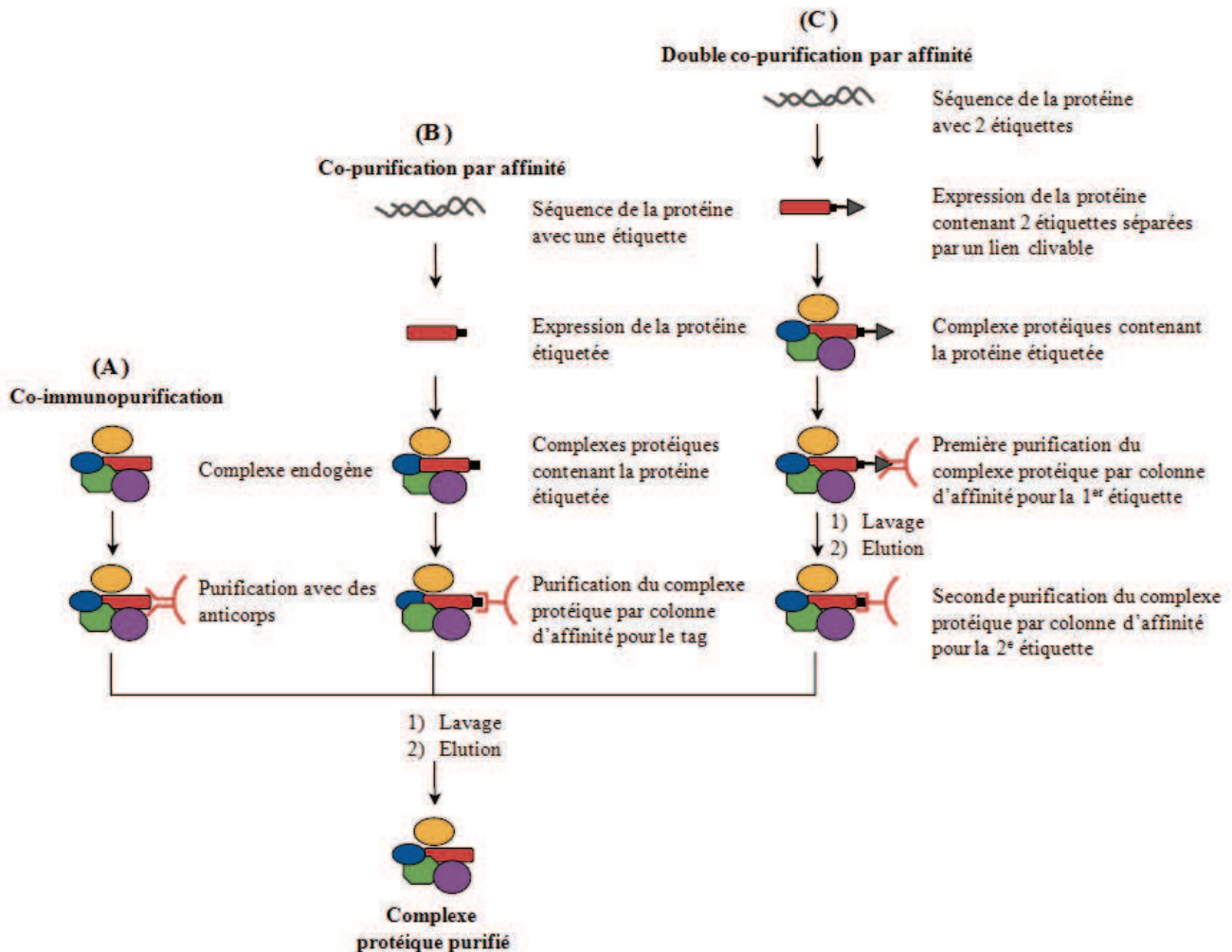


Figure 20. Stratégies de séparation et purification des complexes protéiques par chromatographie d'affinité (la figure a été réadaptée de Kocher and Superti-Furga⁶⁴).

b. Séparation des peptides par nano-chromatographie liquide : nanoLC

L'analyse des protéines entières et leur identification par spectrométrie de masse étant particulièrement délicate et difficile, la majeure partie des analyses de protéines par spectrométrie de masse passe par la digestion enzymatique de ces protéines en peptides (stratégie appelée, bottom up). La nano-chromatographie liquide (nanoLC) est alors utilisée pour séparer les milliers de peptides contenus dans le digestat. La nanoLC est couramment la plus utilisée à cause de sa sensibilité (utilisable sur de faibles quantité de matériels) mais aussi à cause de sa compatibilité avec la spectrométrie de masse (faible débit de solvant). La nano-chromatographie de phase inverse (nanoRPLC) est la plus utilisée à cause de sa reproductibilité (Figure 21, (A)), sa sensibilité, de son efficacité de séparation (Figure 21 (B)). Mais d'autres formes de chromatographie liquide peuvent être couplées avec la chromatographie en phase inversion pour avoir une deuxième dimension de séparation.

Le choix de la phase utilisée en nanoLC (taille de particules, taille des pores, type de phase stationnaire) a plus un impact sur la sélectivité de séparation des peptides. Les colonnes utilisées en nanoLC sont souvent remplies de billes sphériques greffées avec des composés différents selon le type de nanoLC. La taille des particules varie généralement entre 1,7 et 5 μm avec une porosité compris entre 130 - 300 \AA . Pour la nanoRPLC, les billes sphériques sont greffées à leur surface par des chaînes alkyl comme les phase C18 avec 18 carbones⁶⁵. Le faible débit de solvant en nanoLC (100 – 600 nL/min) permet d'avoir un bon compromis entre la séparation chromatographique et une bonne efficacité d'ionisation au niveau de la sources électrospray (nanoESI) du spectromètre de masse et donc un meilleur signal pour les expériences MS ou MS/MS.

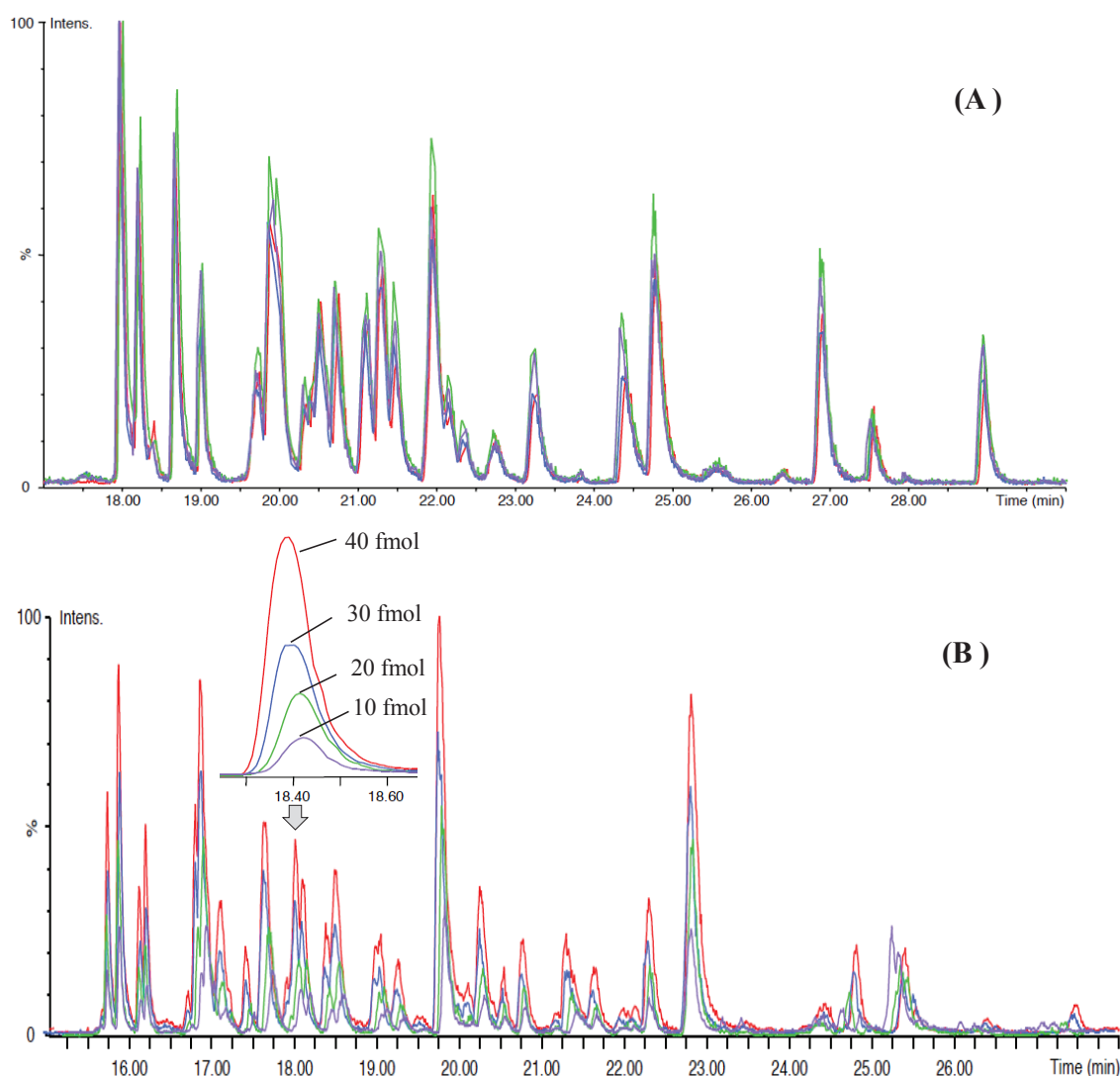


Figure 21. Chromatogramme (BPC, Base peak Chromatogramme) du digeste trypsique de la BSA (Bovin Serum Albumine) sur une colonne nano-RPLC (colonne ACQUITY UPLC BEH30 C18 de 200 mm, 75 μm de diamètre interne, particules de 1,7 μm) en couplage avec le Synapt HDMS (UPLC-Synapt H). (A) Evaluation de la répétabilité avec 4 injections de 50 fmol du digeste trypsique de BSA : pas de décalage entre les pics. (B) Evaluation de la sensibilité avec injection de quantité croissante de digeste trypsique de BSA (10-20-30-40 fmol) : l'intensité des pics croît avec la quantité d'échantillon injectée ⁶⁶.

Les faibles débits utilisés en nanoLC et la faible teneur en analytes demandent une concentration des échantillons avant leur injection sur la colonne pour éviter l'élargissement des pics chromatographiques. C'est pourquoi ce type de système est généralement équipé de colonne de chargement où s'effectuent la concentration et le dessalage des échantillons, et d'une colonne analytique où s'effectue la séparation chromatographique de peptides sous l'effet d'un gradient d'éluion (solvant A (majoritairement de eau), solvant B (majoritairement de l'acétonitrile ou du méthanol).

Les innovations technologiques des dernières années ont permis l'amélioration des capacités de séparation des composés en nanoLC et d'avoir parfois un gain en temps d'analyse. En effet, ces innovations technologiques ont largement contribué à :

- la réduction de la taille des particules (taille de particules $< 2 \mu\text{m}$) pour gagner encore plus en termes de résolution et sensibilité chromatographique (meilleure séparation avec des pics chromatographiques de plus en plus fins FWHM $< 5 \text{ s}$), c'est l'arrivée de l'UHPLC ⁶⁷. La réduction de la taille des particules a permis de travailler à des débits plus élevés ($>300 \text{ nL/min}$) que ceux utilisés classiquement en nanoLC (taille de particules compris entre 5-3 μm) donc des temps d'analyse plus courts, ce qui a imposé une augmentation de pression sur les systèmes et l'apparition de nouveaux équipements pour pallier à ce problème (nanoAcquity UPLC Ultra Performance Liquid Chromatography de Waters Corporation, UHPLC⁺ Ultra High Pressure Liquid Chromatography de Thermo Scientific, UHPLC Agilent Technologies...);

- la miniaturisation des colonnes nanoLC et l'intégration de spray ESI à ces colonnes pour réduire le volume mort et gagner en résolution chromatographique (ex. les puces microfluidiques HPLC-Chip/MS (Agilent Technologies) et TRIZAIC nanoTile (Waters Corporation)) mais aussi le temps d'analyse ⁶⁸;
- le développement de colonnes monolithiques ^{69, 70} permettant de travailler à des débits encore plus élevés à cause de la faible perte de charge de ces colonnes et de gagner ainsi en résolution et en temps d'analyse.

Chromatographie liquide multidimensionnelle

Dans l'approche shotgun où les protéines sont directement digérées sans préfractionnement ni séparation sur gel, la chromatographie multidimensionnelle est souvent utilisée, pour mieux séparer les peptides issus de ces digestes qui sont extrêmement complexes. Cette approche combine plusieurs types de chromatographies dans le but de tirer avantage de chacune des dimensions chromatographiques. Elle permet ainsi de mieux détecter les peptides peu abondants dans l'échantillon. C'est ainsi qu'une approche par chromatographie liquide multidimensionnelle pour l'identification des protéines encore appelée MudPIT (Multi-Dimensional Protein Identification Technology) a été introduite en 1999 ⁷¹. L'approche la plus utilisée consiste à coupler une première étape de séparation des peptides par une chromatographie d'échange de cation (SCX, Strong cation Exchange) directement suivie par une deuxième étape de séparation de peptides sur une colonne nanoRPLC ^{72, 73}. Ces colonnes chromatographiques peuvent être directement emboîtées (MudPIT) ou encore connectées successivement entre elles par des valves (2D on-line). De manière générale, la dernière étape de séparation (la dernière dimension de séparation) reste la RPLC et la première d'autres types de chromatographie liquide comme la IEF ⁷⁴, RPLC à pH basique ⁷⁵.

Le nombre de protéines identifiées par ces approches (shotgun) est parfois supérieur à l'analyse des protéines après séparation sur gel 2D. Cependant, le nombre de peptides ayant servi à l'identification des ces protéines est souvent très faibles et donc ces résultats peuvent comporter de fausses identifications (faux positifs). A cela s'ajoute la difficulté d'interprétation des données d'identification des protéines, puisque, les informations relatives à la protéine (poids moléculaire, pI) sont perdues dès le départ avec la digestion des protéines sans les avoir séparées au préalable ⁷⁶. Il est plus difficile par cette approche de distinguer entre les isoformes ou l'hétérogénéité des modifications post-traductionnelles.

Pour ces raisons que nous venons de citer, toutes les stratégies mise en place pour l'analyse protéomique par spectrométrie de masse utilisées dans ce travail de thèse passent par une première séparation de protéines par l'électrophorèse sur gel 1D ou 2D (cf., Résultats, Chapitre 1 et Chapitre 2), ou encore la séparation par une technique de chromatographie liquide RPLC ou SCX (cf., Chapitre 3 et Chapitre 4).

III. Stratégie d'identification des protéines par spectrométrie de masse : nanoLC-MS/MS

L'analyse protéomique par spectrométrie de masse couplée aux méthodes de séparation des protéines et des peptides, permet d'identifier les protéines contenues dans un échantillon grâce à l'ensemble des spectres MS et MS/MS. En fonction du nombre de protéines contenues dans l'échantillon, des centaines voir des milliers de spectres MS et MS/MS sont alors enregistrés. L'interprétation de ces spectres MS et MS/MS est indispensable pour identifier les protéines correspondantes. Cette interprétation va consister en la mise en relation des spectres de masse avec les séquences peptides et/ou protéiques correspondantes contenues dans la banque protéique. Face à cette quantité de spectres, des outils bioinformatiques ont été développés pour faciliter l'interprétation des données. Selon le type de spectres à interpréter (MS ou MS/MS) on distingue deux approches d'interprétation des données d'analyse protéomique par spectrométrie de masse.

La première approche⁷⁷ consiste à interpréter les données MS (expérience MS de première dimension), c'est-à-dire l'empreinte massique uniquement de la MS. Par exemple dans d'une stratégie "bottom-up", seule la masse des peptides obtenus par analyse en MS, c'est-à-dire l'empreinte peptidique massique (**PMF**, "**Peptide Mass Fingerprinting**") analysés sont prises en compte. Cette approche d'interprétation de données présente beaucoup de limites comme : l'identification de faux positifs, les difficultés d'interprétation de données des échantillons complexes où les empreintes massiques des peptides sont complètement superposées (confusion des peptides de mêmes masses et peptides avec des PTMs par exemple) et la nécessité d'utiliser des spectromètres de masse de haute résolution et précision de masse comme MALDI FT-ICR bien évidemment plus onéreux et surtout difficiles à utiliser en routine^{78, 79}.

La deuxième approche^{80, 81} consiste à interpréter les données MS/MS, c'est-à-dire à la fois l'information sur l'empreinte massique MS et les fragments obtenus dans la deuxième dimension MS (MS/MS). Cette approche va permettre d'identifier non seulement les peptides (ions parents) mais l'enchaînement en acide aminés (ions fragments) de ces peptides permettant ainsi de distinguer entre différents peptides de mêmes masses par leur empreinte de fragmentation peptidique (**PF**, "**Peptide Fragmentation Fingerprinting**").

Cette dernière approche d'identification des protéines basée sur l'interprétation des spectres MS/MS est la plus utilisée car elle offre l'avantage de donner une meilleure fiabilité sur la séquence des peptides, plus d'informations sur la séquence des peptides identifiés et donc des résultats d'identification de protéines beaucoup plus fiables.

1. Acquisition de spectres de fragmentation peptidique (LC-MS/MS)

Il existe deux principaux modes d'acquisition automatique de données d'expériences MS à deux dimensions : le premier mode d'acquisition de données plus ancien est basé sur la sélection individuelle de précurseurs en MS et leur fragmentation dans une cellule collision (Mode DDA, Data-Dependant-Acquisition) et le deuxième mode d'acquisition de données plus récent, évite la sélection des précurseurs et fragmente complètement tous les ions qui arrivent dans un spectromètre de masse en même temps (Mode DIA, Data-Independent-Acquisition). Chacun de ces modes présente des avantages et des inconvénients.

a. Mode d'acquisition données-dépendant : DDA (Data Dependant or Directed Acquisition)

Lors d'une analyse nanoLC-MS/MS, le spectromètre de masse en mode DDA est programmé automatiquement pour réaliser des mesures MS/MS. Cela va donc permettre à l'instrument en temps réel de réaliser lors du premier scan MS (première dimension) une sélection d'un certain nombre d'ions pour une fragmentation en MS/MS. La sélection est dynamique et basée sur des paramètres de sélection physiques (nombre, intensité, charge des ions parents en scan MS) et temporel (temps d'acquisition MS et MS/MS) que l'utilisateur aura défini. Bien évidemment, tous les ions ne sont pas sélectionnés lors d'un scan MS, mais l'optimisation de ces paramètres d'acquisition des spectres MS/MS et les conditions nanoLC permet d'avoir un maximum d'informations sur les peptides de l'échantillon. Ce mode d'acquisition des données LC-MS/MS est la plus utilisé actuellement.

b. Mode d'acquisition données-indépendant: DIA (Data-Independent-Acquisition)

La fragmentation peut se faire sans sélection préalable de précurseur : "shotgun-fragmentation". Il existe plusieurs variantes de "shotgun-fragmentation". On distingue la fragmentation en source (nozzle-Kimmer fragmentation)⁸², la fragmentation dans le cellule de collision CID (Shotgun CID)⁸³ et la MS^E de Waters Cor.⁸⁴. En Shotgun CID par exemple, on effectue une alternance d'acquisition de MS et MS/MS sans isolation de précurseurs. Quand, à la MS^E de Waters Corp, elle permet après un scan MS (à basse énergie) de fragmenter tous les ions en même temps lors de l'étape de fragmentation MS^E (à haute énergie). Il n'y a en théorie pas de perte d'informations sur les peptides liés à la sélection de précurseurs, puisque tous les ions observés sont fragmentés. Les données MS sont alors utilisées pour déterminer la masse exacte et l'intensité du signal des ions parents (peptides). Ces approches de "shotgun-fragmentation" sont intéressantes parce qu'elles permettent d'éviter la sélection aléatoire des précurseurs au cours de l'acquisition et de ce fait améliorent la reproductibilité des peptides et protéines identifiés entre répliquats techniques d'analyse nanoLC-MS/MS. Cependant, les spectres acquis sont très complexes puisque tous les ions observés dans un cycle MS ont leur fils dans un seul spectre "pseudo MS/MS". Néanmoins, comme chaque ions fils suit le profil de son ion parent, chaque ions du spectre acquis peut être associé à son parent par alignement de temps d'élution exact pour reconstituer le profile de fragmentation de celui-ci à l'aide d'un algorithme de calcul⁸⁵.

Une alternative intéressante à la fois pour le mode DDA et le "shotgun-fragmentation" a été reporté par Venable et al⁸⁶. Plutôt que d'effectuer un scan MS complet des précurseurs pour pouvoir déclencher la fragmentation MS/MS comme dans le cas de DDA, Venable et al on rapporté une approche qui procède par une fragmentation des précurseurs dans une fenêtre m/z relativement large (par exemple, des fenêtres de 10 m/z). Tout au long du scan sur la gamme de mass, on effectue des fragmentations à chaque 10 m/z, et ainsi chaque spectre MS/MS est acquis pour chaque m/z à intervalle de temps réguliers. Les spectres MS/MS sont acquis indépendamment des données de scan MS antérieures d'où le nom de mode d'acquisition données-indépendant DIA (Data-Independent-Acquisition)⁸⁶. Plus récemment Gillet et al. ont présenté une stratégie d'acquisition de spectre de fragmentation MS/MS données – indépendantes nommée SWATH⁸⁷. Dans leur étude, les données MS/MS ont été acquises sur QqTOF de haute résolution et vitesse de balayage avec une fenêtre d'isolation de précurseurs de 25 Da, soit 32 MS/MS consécutifs sur une gamme de masse 400 - 1200 m/z. Le mode d'acquisition permet de gagner en sensibilité et reproductibilité par rapport au mode DDA. Le mode DIA permet aussi d'avoir une meilleure quantification relative basée sur les courants d'ions DIA plus précise⁸⁸ que le comptage de spectres (spectral counting) communément effectuée en mode DDA⁸⁹.

Bien évidemment, ces modes d'acquisitions vont générer un grand volume de données très complexes difficilement interprétable par l'utilisateur qui se fie entièrement à l'algorithme de traitement de données. Aucune vérification ou validation ultérieure ne peut donc avoir lieu manuellement comme en mode DDA. De plus, ces modes de fragmentation ne sont pas pris en charge par la plupart moteurs recherche pour l'interprétation des spectres de fragmentation. Ces raisons font que ces modes d'acquisition de données MS/MS sont moins utilisées par rapport au DDA actuellement. Le développement des logiciels de traitement de données et de d'interprétation universel de ce type de données MS/MS devrait permettre dans le futur une utilisation plus répandue de ces méthodes d'acquisition.

2. Interprétation de données MS/MS

a. Approche par empreinte de fragmentation peptidique : Peptide Fragment Fingerprinting (PFF)

Cette approche est basée sur la comparaison des listes de masses des ions parents (mesurés lors de la première dimension MS (les spectres MS) et de leurs fragments associés (mesurés lors de la deuxième dimension MS (les spectres MS/MS)) avec des listes de masses issues de digestion théorique des protéines *in silico* présentes dans des banques de séquences protéiques et de la fragmentation théorique des peptides associés (Figure 22). Cette comparaison est réalisée grâce à un moteur de recherche (Sequest⁸⁰, Mascot⁹⁰, OMSSA⁹¹, X!tandem⁹²).

Il existe aujourd'hui de nombreux moteurs de recherche de données d'analyse protéomique qui fonctionnent sur l'approche par PFF et chacun avec ses spécificités (par exemple les critères de score de qualité de similitude en données expérimentales et théoriques issus de la banque, etc.)⁹³.

Quelque soit le moteur de recherche celui-ci fonctionne à partir des paramètres utilisateur qui sont : la erreur de masse tolérée sur l'ion parent (précurseur) et ions fils (fragments), l'enzyme de digestion utilisée (coupures spécifiques ou aspécifiques), les modifications des acides aminés, le type de série de fragmentation recherché (a, x, b, y, c, z etc.) et les formats de la liste de masses (mgf, pkl, mzxml etc.).

Pour avoir plus de fiabilité dans les résultats d'identification, des recherches peuvent être faites en combinant plusieurs moteurs de recherche⁹⁴. En effet des tests de comparaison effectués entre les moteurs de recherche (Mascot, Sequest, X!tandem, Phenyx) et (Mascot, Sequest, X!tandem, OMSSA) ont montré des variations en termes de résultats d'identification sur des mêmes données LC-MS/MS. La combinaison de moteur de recherche est un bon moyen de confirmation des résultats et d'identification des peptides candidats supplémentaires⁹⁵⁻⁹⁷.

Les moteurs de recherche sont en perpétuel amélioration pour augmenter leur vitesse, capacité et fiabilité à traiter de grands volumes de données LC-MS/MS⁹⁸.

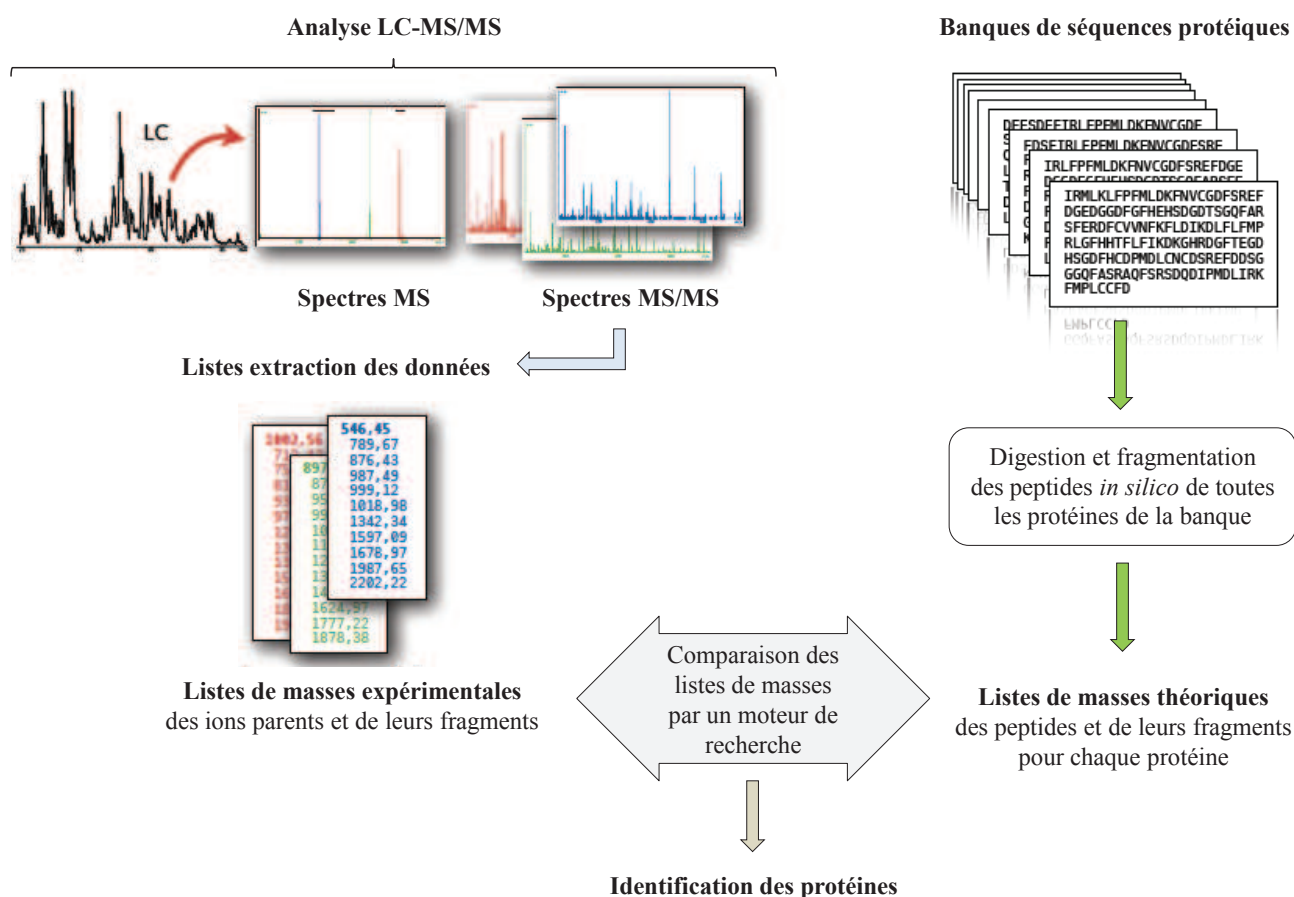


Figure 22. Principe d'identification des protéines par nanoLC-MS/MS

Limites de la recherche conventionnelle par les moteurs de recherche

Cette approche d'interprétation de données LC-MS/MS dite conventionnelle est "non tolérante aux erreurs", c'est-à-dire les algorithmes se basent sur des identités strictes entre données expérimentales et données théoriques et donc cette approche n'est adaptée que pour les protéines présentes dans les banques de séquences protéiques (Figure 23 (A)). Si nous considérons deux peptides quasi identiques comme TRYPSINE et TRYQSINE dont uniquement TRYPQSINE se trouve dans la banque, le peptide TRYPSINE ne sera pas identifié même si certains ions fragments sont communs entre les deux peptides. Les banques protéiques doivent donc être bien fournies (séquences correctes et complètes de protéines des génomes étudiés). Or dans certains cas les organismes étudiés n'ont pas leurs génomes ou alors le séquençage du génome est inachevé. Cette approche banque protéique-dépendante va entraîner une interprétation incomplète de données (Figure 23, (B)), puisque seules les protéines provenant de génomes séquencés vont pouvoir être identifiées dans une étude mettant en jeu des organismes séquencés et non séquencés (cf., Résultats, Chapitre 2).

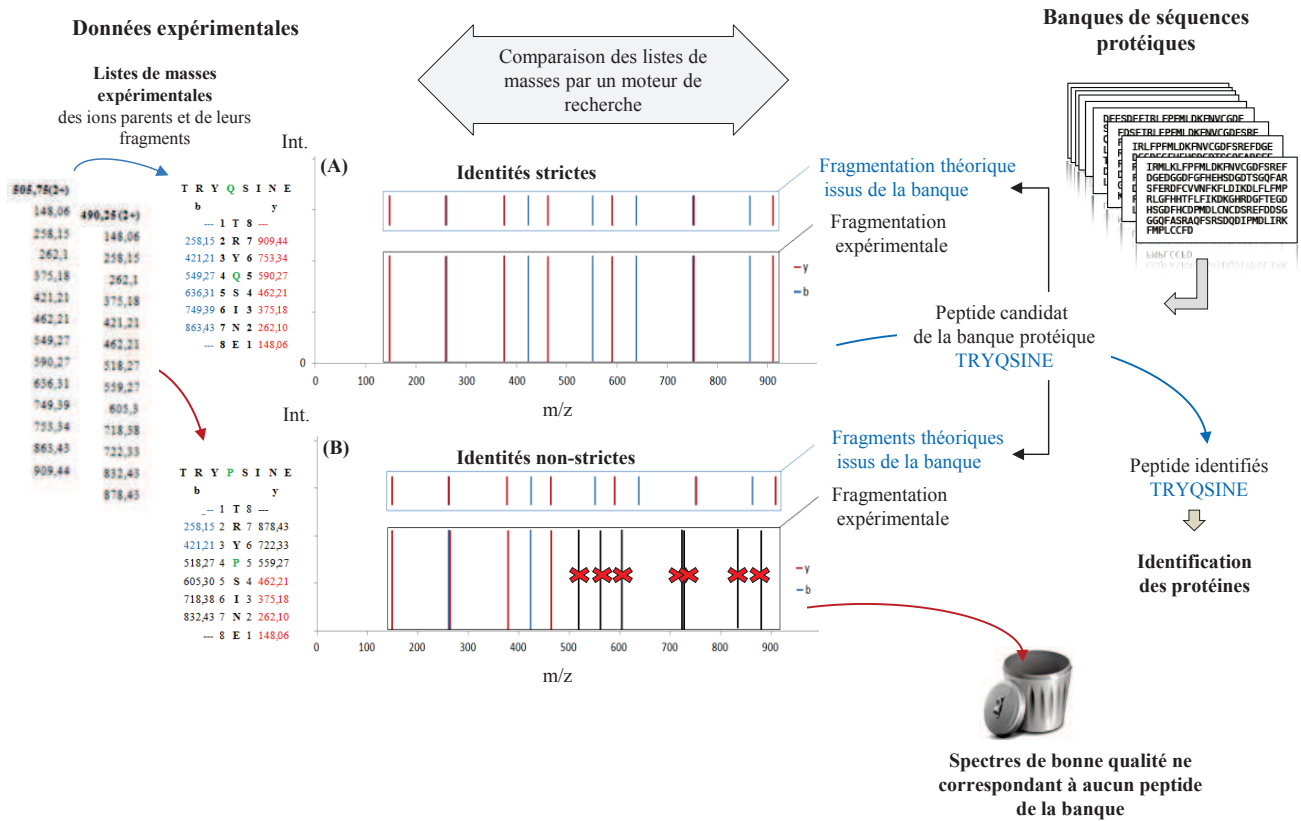


Figure 23. Principe de fonctionnement des moteurs de recherche : fonctionnement par stricte identité. (A) Les protéines identifiées sont celles ayant des peptides avec la même empreinte de fragment théorique que les empreintes de fragments dans les données LC-MS/MS. (B) Les autres spectres ne correspondant pas strictement à des données dans la banque de séquences protéiques ne sont pas interprétés et ne sont pas considérés, même s'ils sont de bonne qualité. Ces peptides peuvent correspondre à des protéines qui ne sont pas dans la banque de données.

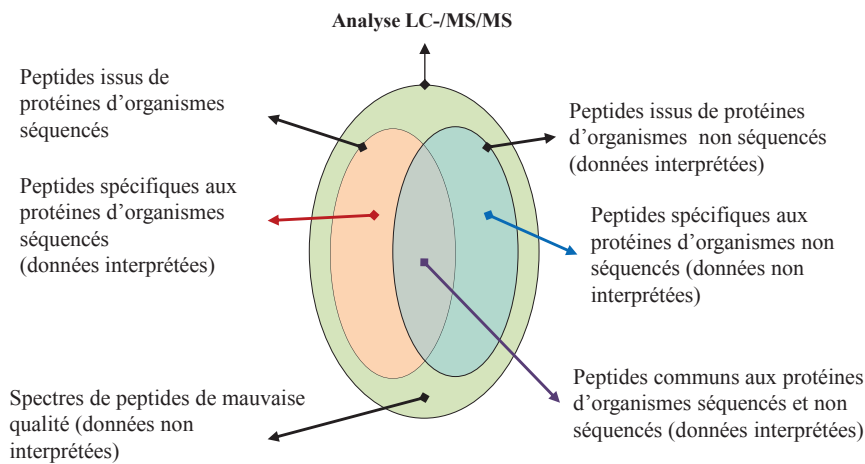


Figure 24. Schématisation de la composition des données MS/MS : une partie des données n'est pas été interprétées lors de la recherche par moteur de recherche (Mascot, OMSSA etc.) pour de principales raisons que : les spectres sont de mauvaise qualité ou ne sont pas des spectres de peptides (autres molécules) ; les spectres correspondent à des fragmentations de peptides dont les protéines ne sont pas dans les banques de séquences protéiques utilisées. Les données interprétées sont celles exclusivement contenues dans la banque de séquences protéiques.

Seul donc une partie des spectres va servir à l'identification. Les autres des spectres restant, même s'ils sont de bonne qualité et assez informatifs sur la séquence d'un peptide issue d'une protéine absente de la banque protéique, ne sont pas utilisés (Figure 24). Une autre approche d'interprétation de données LC-MS/MS permet de pallier à ce problème : le séquençage *de novo* et l'identification inter-espèces par recherche de similarité (homologie de séquence).

b. Approche par séquençage *de novo* et identification inter-espèces par recherche de similarités de séquences (homologie de séquence)

Le séquençage *de novo* des données MS/MS est une alternative pour l'identification des protéines. Dans cette approche réalisée manuellement au début du séquençage *de novo*, les séquences en acides aminés sont directement déduites des spectres MS/MS par identification des différences de masse entre pics successifs pouvant correspondre à la masse des acides aminés. La différence de masse entre deux fragments consécutifs de la même série de fragmentation d'un peptide correspond à l'acide aminé compris entre les deux fragments. Il est donc possible à partir des différences de masse entre fragments successifs de la même série de reconstruire la totalité ou en partie (tags de peptide) la séquence du peptide fragmenté (Figure 25). Le résultat du séquençage *de novo* dépend de la qualité du spectre (nombre de fragments d'une série (a, b, c, x, y) et de la précision de mesure de masse.

Face au volume des données d'analyse LC-MS/MS, des outils de séquençage *de novo* automatisés ont été développés.

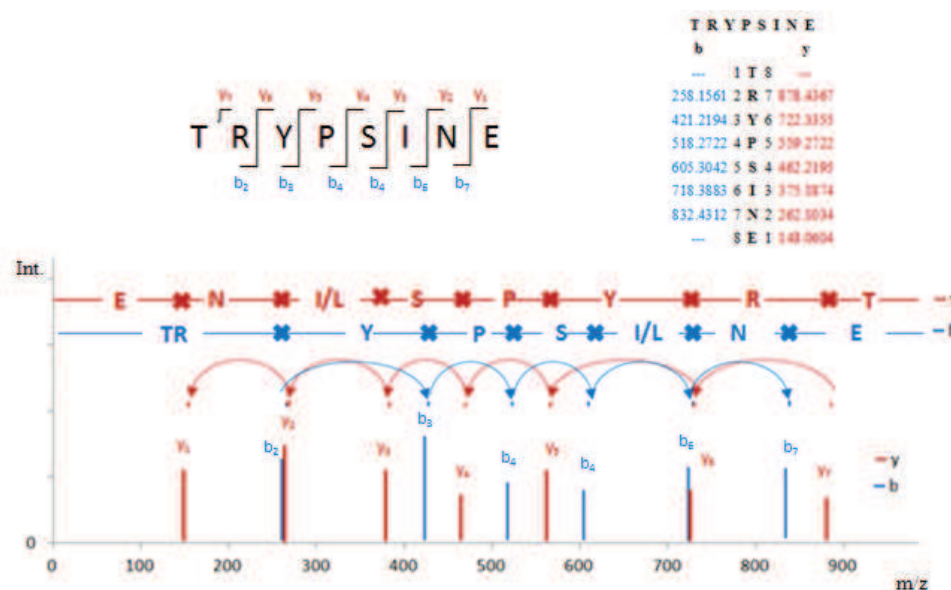


Figure 25. Principe de séquençage *de novo* (interprétation manuelle) : détermination des écarts de masse successifs entre des pics de fragmentation de la même série pour reconstituer la séquence en acides aminés du peptide fragmenté (exemple du peptide TRYPSINE). Le spectre est simplifié et présente uniquement les ions de la série b et y.

i. Séquençage *de novo* automatisé des spectres MS/MS

A l'instar des moteurs de recherche dans les banques protéiques, il existe aussi plusieurs algorithmes de séquençage *de novo* des données expérimentales de MS/MS. Les principaux algorithmes de séquençage *de novo* sont : RapidDenovo (Biotoools, Bruker), PepSeq (MassLynx, Waters), Sherenga⁹⁹, AUDENS¹⁰⁰, ADEPTS¹⁰¹, Lutefisk, PEAKS¹⁰², PepNovo¹⁰³. Différentes approches ont été développées pour l'automatisation de l'interprétation des spectres par séquençage *de novo*. Deux grandes familles d'approches sont les plus utilisées : l'approche par pseudo-PFF et la reconstruction itérative qui repose sur la notion de graphe spectral¹⁰⁴.

Approche par pseudo-PFF (Exhaustive search approche)

A partir de la connaissance de la masse du peptide analysé (ion parent), ce type d'approche va consister à générer toutes les combinaisons possibles d'acides aminés permettant de retrouver cette masse, et donc à créer une banque de tous les peptides possibles (Figure 26, (A), Pseudo banque). Les spectres théoriques de ces peptides sont ensuite comparés aux spectres expérimentaux de la même manière que dans une approche PFF (Figure 26, (B) identité stricte) (PASS3,¹⁰⁵).

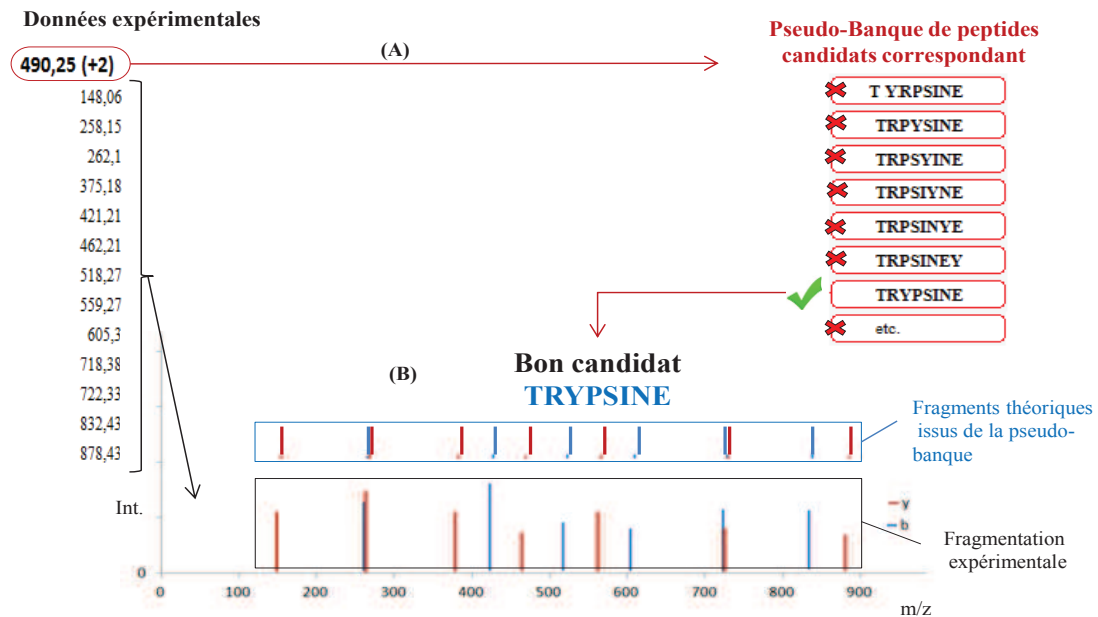


Figure 26. Principe de fonctionnement du séquençage *de novo* automatisé en mode pseudo-PFF.

Le principal défaut de ces méthodes est lié à l'explosion combinatoire du nombre de peptides artificiels générés. Différentes méthodes de séquençage *de novo* basées sur ce type d'approche existent et se distinguent par leur manière de limiter l'impact de l'explosion combinatoire.

- L'usage par exemple des ions immonium permet de contraindre la composition des peptides. L'immonium est un fragment présent occasionnellement dans le spectre, et qui signale la présence d'un acide aminé précis sans pour autant donner d'information sur sa position au sein du peptide. Il est donc possible de ne générer que des séquences d'acides aminés contenant au moins les acides aminés désignés par la présence d'ions immonium, ce qui réduit ainsi fortement le nombre de séquences candidates à générer¹⁰⁶.
- Il est aussi possible de ne pas générer tous les peptides candidats. Dans ce cas un certain nombre de peptides sont générés aléatoirement à l'aide de l'algorithme et évalués à l'aide des critères de score pour retenir uniquement les bons peptides candidats avec de meilleurs scores¹⁰⁷, ce qui va réduire l'espace de recherche de solution.

Exemple de PEAKS (Bioinformatics Solutions Inc.)

PEAKS¹⁰², est un logiciel pour le séquençage *de novo* automatisé, qui fonctionne en deux étapes :

Dans un premier temps, un algorithme dynamique est utilisé pour générer les 10000 meilleurs peptides candidats à partir de la masse de l'ion parent. Durant cette génération de candidats, le score utilisé ne tient compte que d'assez peu d'information (abondance des ions a, b, c, x, y ainsi que des variantes b/y présentant une perte de H₂O et NH₃).

Dans un second temps, les 10000 candidats générés sont réévalués en utilisant une méthode de score plus performante, qui va cette fois utiliser une tolérance de masse plus stricte et considérer la présence d'autres types d'ions. Cet outil a pour particularité de donner un score de confiance pour chaque acide aminé en plus de celui donné à la séquence complète.

Cette méthode de séquençage *de novo* bien que très performante pour des peptides non modifiés (peptides sans PTMs) conduit souvent à des résultats complètement erronés lorsqu'il s'agit de peptides modifiés. En effet, cette approche conduit à la proposition des séquences peptidiques même si certaines séquences (tags du peptide) proposés sont peu probables (faible degré de confiance). Il existe une autre approche de séquençage *de novo* beaucoup plus fiable.

Reconstruction itérative d'une séquence de peptide (Spectrum graph approach)

A lieu de tester un grand nombre de peptides candidats sur chaque spectre expérimental, certaines méthodes procèdent par reconstruction du peptide de manière itérative, tout comme dans la méthode manuelle. Les fragments contenus dans les spectres sont directement utilisés avant de générer des peptides candidats. Ces

méthodes s'appuient sur une représentation du spectre sous la forme d'un graphe spectral (Sherenga⁹⁹, Lutefisk¹⁰⁸, PEAKS¹⁰², PepNovo¹⁰³).

Dans un graphe spectral, les ions issus d'un même fragment de peptide sont regroupés en un nœud. Un score est alors attribué à chaque nœud en fonction de la probabilité que la fragmentation ait lieu à l'emplacement de ce nœud. Deux nœuds sont reliés par un arc si leur différence de masse est égale à la masse d'un ou plusieurs acides aminés. Chaque arc est étiqueté avec le nom du ou des acides aminés qu'il représente. Il est par la suite plus facile de rechercher le chemin de score maximal dans ce graphe. Le score d'un chemin correspond à la somme des scores de tous les nœuds qu'il emprunte. La lecture des étiquettes le long des arcs empruntés par ce chemin va donc donner la séquence qui correspond et interprète le mieux le spectre. La Figure 27 représente un exemple de graphe spectral. Dans ce graphe, le chemin de couleur rouge est le chemin de score maximal. Par exemple, **Sherenga** est une méthode évaluant chaque nœud grâce à un ratio de vraisemblance entre deux hypothèses: (i) les pics sont issus d'une fragmentation, (ii) les pics sont du bruit.

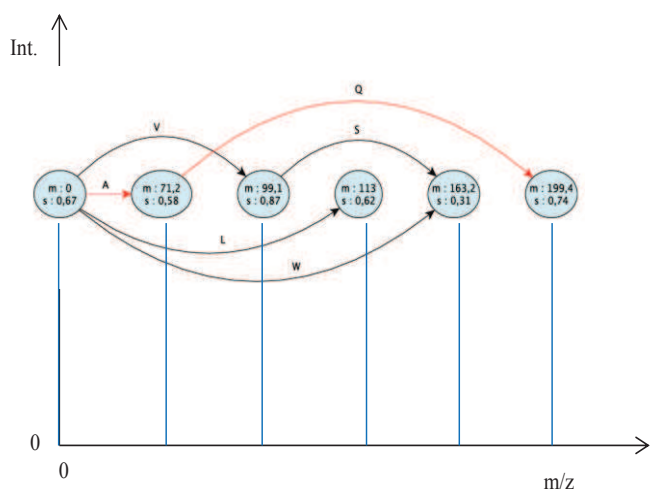


Figure 27. Principe de fonctionnement du séquençage *de novo* itérative d'une séquence de peptide (*Spectrum graph approach*). Exemple de graphe spectral : Chaque nœud contient comme information la position (masse *m*) qu'il représente ainsi qu'un score (*s*). Plusieurs arcs sont représentés et portent tous l'étiquette de l'acide aminé qu'ils représentent. Chaque chemin de ce graphe représente une séquence d'acides aminés possible. La somme des scores de chaque chemin permet de trouver le meilleur peptide candidat (chemin de score maximal en rouge).

Exemple de PepNovo

En 2005, le concept de Sherenga a été repris dans PepNovo avec un calcul de score totalement redéfini. PepNovo utilise un raisonnement probabiliste pour modéliser les événements de fragmentation de peptides dans un spectromètre de masse¹⁰³. Les deux hypothèses utilisées pour calculer le ratio de vraisemblance y ont été modifiées pour prendre en compte de nombreux aspects qui sont liés à la séquence protéique, aux conditions expérimentales ou encore à l'appareil utilisé (basés sur les observations expérimentales du comportement de peptides en MS/MS). Pour paramétrer chacun des éléments de ce score, PepNovo utilise un apprentissage basé sur la fragmentation des peptides modèles.

L'intérêt des graphes spectraux "Spectrum graph" par rapport aux méthodes de type Pseudo-PFF réside dans la réduction de la taille de l'espace de recherche. Les graphes spectraux ont réduit cet espace de recherche de l'ensemble des peptides possibles à un sous-ensemble correspondant à l'ensemble des chemins de ces graphes. De plus les résultats présentés sont ceux correspondant aux chemins qui expliquent le mieux possible les spectres. Les spectres sont ainsi complètement annotés si tous les pics sont présents dans les spectres, sinon seulement les tags (séquence partielle d'acides aminés consécutifs d'un peptide) justes sont identifiés. Cette approche est bien plus réaliste puisque tous les fragments ne sont toujours pas présents dans les spectres MS/MS (dépendant des conditions expérimentales et de la nature de l'acide aminé). Comme le montrent plusieurs différentes expériences indépendantes comparant PepNovo et d'autres algorithmes de séquençage *de novo*, PepNovo a donné de meilleurs résultats en terme de peptides correctement séquencés par rapport à plusieurs autres algorithmes de séquençage *de novo*^{109, 110}.

Cet algorithme de séquençage automatique des spectres MS/MS basé sur les graphes spectraux "Spectrum graph" a été utilisé dans l'un des projets de ces travaux de thèse (cf. Résultats, Chapitre 2).

ii. Identification inter-espèces par recherche de similitude (homologie de séquences)

Cette approche, contrairement à la recherche conventionnelle via les moteurs de recherche (cf., III.2.a.) est tolérante aux erreurs de séquence : elle permet l'identification des protéines non présentes dans les banques de données par homologie de séquence avec des protéines d'organismes phylogénétiquement proches. Cette

approche offre donc la possibilité de travailler sur des organismes dont les génomes ne sont pas encore séquencés. L'approche se base sur le fait que, les organismes de différentes espèces, expriment des protéines homologues (des protéines dont les séquences d'acides aminés sont quasi identiques, d'un organisme à l'autre) (Figure 28), qui présentent une forte similitude dans leurs séquences d'acides aminés.

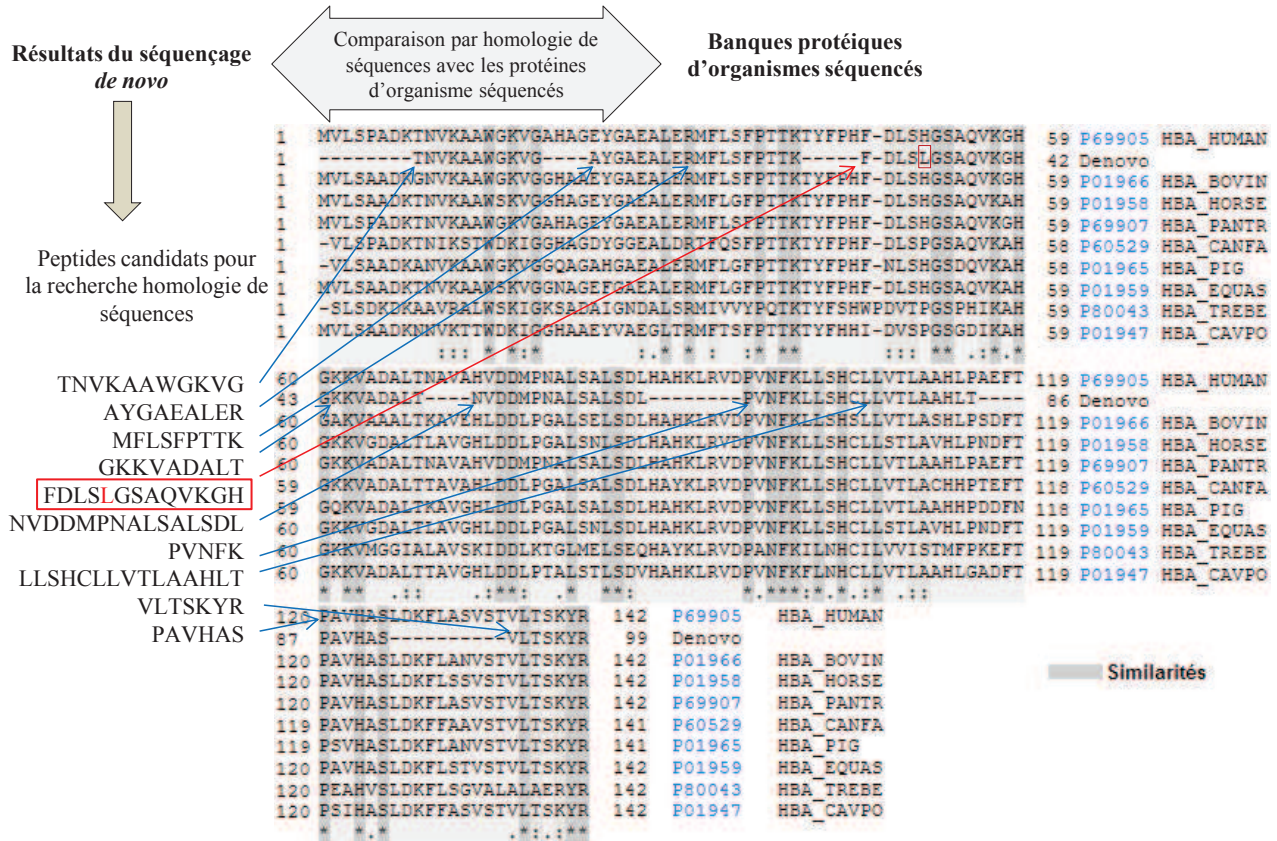


Figure 28. Alignement de séquences protéiques des chaînes de globines entre différentes espèces (*Homo sapiens*, *Bos taurus*, *Equus caballus*, *Pan troglodytes*, *Canis familiaris*, *Sus scrofa*, *Equus asinus*, *Trematomus bernacchii*, *Cavia porcellus*) et les peptides séquencés par *de novo*. L'alignement montre que les peptides candidats sont issus d'une ou des protéines homologues à la globine alpha. Dans ce cas il s'agit d'une variante de globine alpha dont le peptide normal FDSLH*GAQVVKGH porte une mutation au niveau de (H*) qui est remplacé par L/I (FDLSL*GAQVVKGH).

L'alignement des tags ou peptides obtenus par séquençage *de novo* avec d'autres protéines dans les banques protéiques (organismes séquencés) va donc permettre de déterminer par similitude (homologie), la ou les protéines homologues de ces organismes séquencés.

Comme les moteurs de recherche, il existe plusieurs outils de comparaison de séquences protéiques et la plupart de ces outils fonctionnent par alignement de séquence les unes contre les autres. Cet alignement s'effectue de manière à avoir la meilleure similitude possible entre les différentes séquences comparées. Le score de similitude (homologie) entre les séquences est défini alors comme l'optimum de l'alignement de séquences. Les algorithmes d'alignement de séquences les plus utilisés sont : BLAST¹¹¹, FASTA¹¹². Ces outils au départ destinés à l'alignement des séquences protéiques ont été modifiés pour prendre en compte la taille des tags de séquences plus courts (CIdentify¹¹³, MS-BLAST¹¹⁴, FASTS et FASTF¹¹⁵).

L'un des principaux atouts de ces outils est l'automatisation des recherches d'homologie de séquences à partir des tags et la soumission de plusieurs requêtes à la fois, ce qui constitue un gain de temps énorme devant le volume des analyses LC-MS/MS. La combinaison de plusieurs de ces outils avec les algorithmes de séquençage *de novo* ont été cités dans la littérature (PEAKS, PepNovo et MS-BLAST ou FASTS),¹¹⁶⁻¹¹⁸.

c. Les banques de séquences protéiques

L'identification des protéines et l'interprétation biologique qui découle d'une analyse protéomique dépend directement de la disponibilité de la banque de séquences protéiques et de sa qualité d'annotation. Le développement des techniques de séquençage d'ADN et ARNm ces dernières années ont conduit à l'explosion des banques génomiques et protéomiques (Figure 29) qui ont ainsi donné un grand essor à l'analyse protéomique.

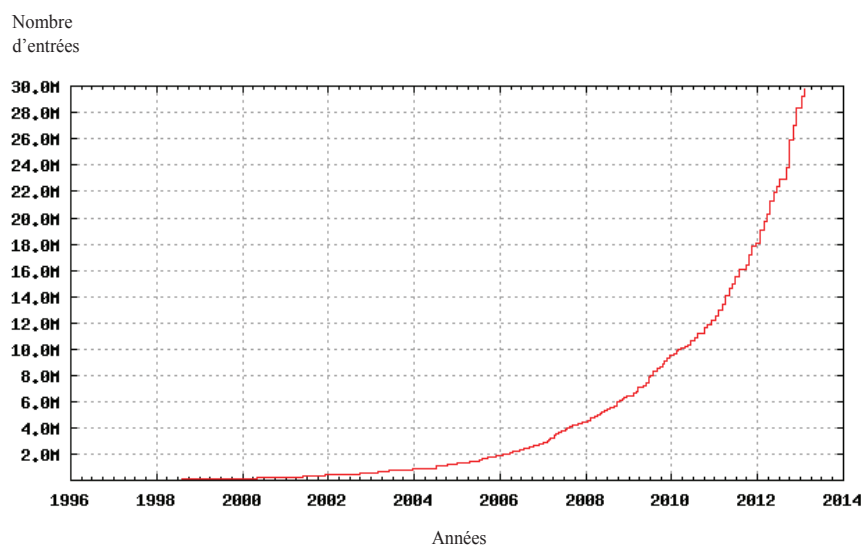


Figure 29. Évolution du nombre d'entrées des séquences protéiques (en millions) dans UniProt/TrEMBL au cours des 15 dernières années. *Source :* European Bioinformatics Institute (EBI) (<http://www.ebi.ac.uk/uniprot/TrEMBLstats/>). (7 février 2013)

La disponibilité croissante des banques de séquences protéiques a permis l'émergence des stratégies d'identification automatique des protéines. De nombreuses banques sont disponibles aujourd'hui¹¹⁹⁻¹²², et chacune d'elle est caractérisée par son exhaustivité, son degré de redondance, la qualité de ses annotations. On distingue deux catégories de banques protéiques :

- les banques généralistes de dépôt contenant essentiellement des séquences protéiques issues de la traduction des gènes prédits *in silico* par des algorithmes de prédiction de protéines sans vérification et avec des annotations fonctionnelles limitées, voire quasi inexistantes. Comme exemple on peut citer la banque GenPept (GenBank, Gene products Data Bank) qui est produite par le NCBI (National Center for Biotechnology Information) et qui regroupe l'ensemble des protéines traduites à partir des banques nucléotidiques et avec une forte redondance des séquences.
- les banques protéiques spécialisées couramment utilisées en protéomiques sont des banques pour lesquelles un travail de nettoyage, de tri, de documentation et d'annotation a été réalisé par des experts¹²³. La liste de ces banques est longue mais néanmoins les plus utilisées sont Swiss-Prot (UniprotKB/SwissProt), TrEMBL (UniprotKB/TrEMBL).

Les protéines sont classées par taxonomie (organismes), en fonction de l'organisme étudié, ces banques peuvent être restreintes à la taxonomie ou des protéines de l'organisme étudié être extraites de ces banques pour constituer une banque moins volumineuse. La prise en compte soigneuse de la taille de la banque (restriction de la banque en fonction de l'étude) va permettre non seulement de réduire l'espace de recherche et donc gagner en temps de recherche mais aussi, permettre une meilleure interprétation des résultats d'identification et limiter le taux de faux positifs.

3. Evaluation des résultats

Face à la diversité des moteurs de recherche, les scores calculés par ces moteurs de recherche ne constituent pas des indicateurs suffisamment fiables pour considérer la pertinence des identifications. Une étape supplémentaire de vérification est particulièrement nécessaire lorsque les études nécessitent l'interprétation d'un grand volume de données de LC-MS/MS. Il est en effet indispensable dans ce cas d'associer une évaluation de la qualité et de la confiance accordée aux résultats d'identification afin de ne pas biaiser les conclusions de l'étude. C'est pourquoi les journaux spécialisés dans le domaine de l'analyse protéomique par spectrométrie de masse comme MCP (Molecular and Cellular Proteomics) et JP (Journal of Proteomics) ont établi des directives sur les informations et la documentation à associer aux résultats d'analyse protéomique afin de faciliter le suivi et la reproductibilité de l'interprétation des résultats publiés.

Cette vérification des résultats de recherche peut se faire manuellement, mais devant le volume de données LC-MS/MS, il est devenu indispensable de disposer d'outils d'estimation empirique de la confiance à accorder aux identifications. Des stratégies d'évaluation de taux de faux positifs (mauvaises identifications) sont alors apparues.

Une première approche d'estimation de taux de faux positifs est l'approche "Target-Decoy" ¹²⁴. Dans cette stratégie, la recherche d'identification est faite dans une banque protéique classique "Target / cible" combinée à 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 "Target" et "Decoy" suivent la même distribution. Les résultats d'identification réalisés dans la banque Target-Decoy contiennent alors des vrais et faux positifs mélangés à des identifications de la banque leurre (Decoy). Les identifications des 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 30). Cette stratégie permet ainsi d'évaluer le taux de faux positifs (FDR), et la probabilité qu'une identification soit correcte dans un ensemble d'identification (peptides ou protéines). Comme le montre la Figure 30, le compromis qui doit être fait entre le nombre d'identifications et le taux de faux positifs (Figure 30, (A)). Plus le seuil de score appliqué est bas, plus le nombre d'identification mais également le nombre de faux positifs est grand (FDR élevé) (Figure 30, (B)). Par contre, avec un seuil trop élevé, les identifications obtenues sont faibles mais sont fiables (FDR, bas) et le taux de faux négatifs est important et l'on perd des identifications (Figure 30, (C)).

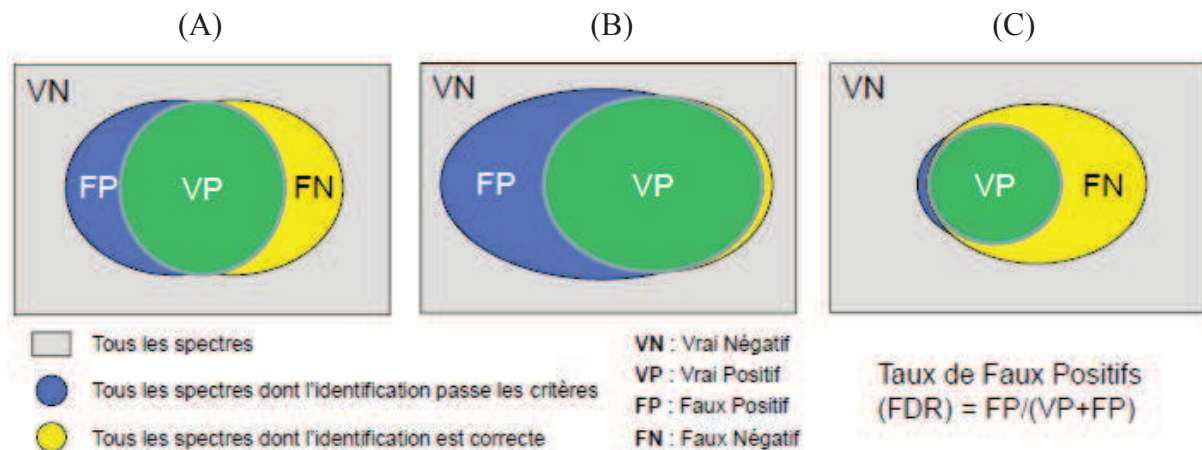


Figure 30. Illustration de la validation des identifications en fonction du seuil de score.

La seconde approche, utilise les approches empirique de Bayes ^{125, 126} et consiste à utiliser des scores probabilistes obtenus pour l'ensemble des listes de masses candidates à une identification. Ces scores sont calculés par des algorithmes lors de la comparaison des listes de masses théoriques aux listes de masses expérimentales. Deux populations de scores sont alors considérées : les scores des listes de masses théoriques ayant conduit à une identification de rang 1 (identifications correctes) et les scores de listes de masses ayant conduit à une identification de rang 2 ou supérieur (mauvaises identifications). Les distributions des scores des deux populations sont modélisées par deux lois binomiales. Par comparaison des deux lois binomiales, il est alors possible d'attribuer à chaque score probabiliste une probabilité d'appartenir à l'une ou l'autre des populations (Figure 31).

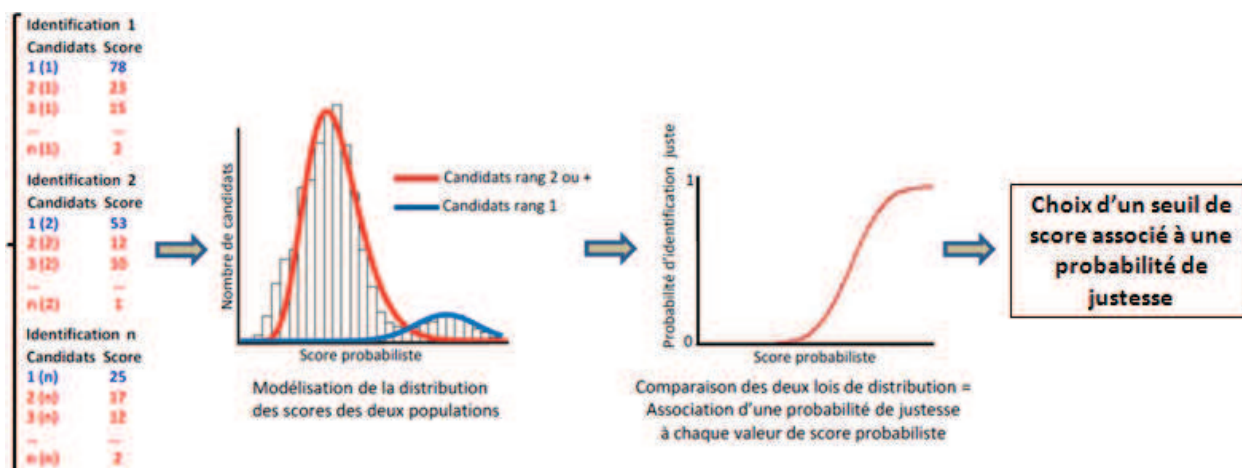


Figure 31. Illustration de la distribution des vrais positifs (candidats de rang 1) et faux positifs (candidats de rang 2) et principe de fixation du seuil pour l'utilisation des approches empiriques de Bayes^{125, 126}.

Bien que ces outils permettent d'évaluer les identifications et assurent une certaine confiance dans la qualité des résultats d'analyse protéomique par spectrométrie de masse, il est important de compléter l'étude avec des tests biologiques sur des protéines candidates identifiées par analyse protéomique. L'utilisation de méthodes d'analyse différentes permet d'établir un ensemble de preuves sur la validation définitive de l'identification d'une protéine (cf., Résultats, Chapitre 1 et 2).

Conclusion

Comme nous avons pu le voir tout au long de ce chapitre, l'analyse protéomique comprend plusieurs étapes importantes qui nécessitent l'utilisation de plusieurs techniques et outils de protéomique : techniques de préparation des échantillons, de purification et de séparation des protéines, la spectrométrie de masse et l'interprétation des données à l'aide d'outils bioinformatiques. Ce dernier point particulièrement nous a permis de mettre en lumière certaines limites de l'analyse protéomique pour l'étude des organismes non ou mal séquencés, ce qui a mis en évidence la nécessité de développer les approches par séquençage *de novo*. Chacune de ces étapes de l'analyse protéomique doit être soigneusement choisie et réalisée de manière optimale pour ne pas biaiser les étapes suivantes.

En fonction de la question biologique posée et de la nature de l'échantillon, le "protéomiste" évalue le parc instrumental et les outils bioinformatiques pour choisir les mieux adaptés au contexte d'étude. Il développe alors des méthodologies analytiques, et met au point des stratégies d'interprétation de résultats pour avoir les informations de qualité afin d'apporter des éléments de réponse moléculaire à la question posée. Puisque, chaque technique utilisée est susceptible d'induire des biais, il est donc indispensable de maîtriser parfaitement les techniques utilisées. Et lorsque les outils existant ne permettent pas d'apporter de réponses à la question posée, le "protéomiste" les adapte, ou développe avec l'aide des bioinformaticiens, de nouveaux outils bioinformatiques pour pallier aux limites.

En dépit des étapes importantes de l'analyse protéomique comme la préparation des échantillons, la qualité et la quantité d'information d'analyse protéomique par spectrométrie de masse (MS et MS/MS) sont étroitement liées au couplage LC-MS. Il faut donc s'assurer des bonnes performances du système chromatographique et du spectromètre de masse qui sont l'une des clés du succès de l'analyse protéomique et plus particulièrement lorsque les échantillons sont disponibles en faible quantité.

Références bibliographiques du chapitre 1

1. Reynolds, T. For proteomics research, a new race has begun. *Journal of the National Cancer Institute* **2002**, 94, (8), 552-4.
2. Chinery, M. Insectes de France et d'Europe occidentale. *Flammarion*: 2005; p 320.
3. Bensimon, A.; Heck, A. J.; Aebersold, R. Mass spectrometry-based proteomics and network biology. *Annual review of biochemistry* **2012**, 81, 379-405.
4. Sabido, E.; Selevsek, N.; Aebersold, R. Mass spectrometry-based proteomics for systems biology. *Current opinion in biotechnology* **2012**, 23, (4), 591-7.
5. Griffiths, J. A brief history of mass spectrometry. *Analytical chemistry* **2008**, 80, (15), 5678-83.
6. Alafuzoff, I.; Arzberger, T.; Al-Sarraj, S.; Bodi, I.; Bogdanovic, N.; Braak, H.; Bugiani, O.; Del-Tredici, K.; Ferrer, I.; Gelpi, E.; Giaccone, G.; Graeber, M. B.; Ince, P.; Kamphorst, W.; King, A.; Korkolopoulou, P.; Kovacs, G. G.; Larionov, S.; Meyronet, D.; Monoranu, C.; Parchi, P.; Patsouris, E.; Roggendorf, W.; Seilhean, D.; Tagliavini, F.; Stadelmann, C.; Streichenberger, N.; Thal, D. R.; Wharton, S. B.; Kretschmar, H. Staging of neurofibrillary pathology in Alzheimer's disease: a study of the BrainNet Europe Consortium. *Brain pathology* **2008**, 18, (4), 484-96.
7. Gelpi, E. From large analogical instruments to small digital black boxes: 40 years of progress in mass spectrometry and its role in proteomics. Part II 1985-2000. *Journal of mass spectrometry* **2009**, 44, (8), 1137-61.
8. Gelpi, E. From large analogical instruments to small digital black boxes: 40 years of progress in mass spectrometry and its role in proteomics. Part I 1965-1984. *Journal of mass spectrometry* **2008**, 43, (4), 419-35.
9. Schaeffer-Reiss, C. A brief summary of the different types of mass spectrometers used in proteomics. *Methods in molecular biology* **2008**, 484, 3-16.
10. Mann, M.; Kelleher, N. L. Precision proteomics: the case for high resolution and high mass accuracy. *Proceedings of the National Academy of Sciences of the United States of America* **2008**, 105, (47), 18132-8.
11. Karas, M.; Hillenkamp, F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Analytical chemistry* **1988**, 60, (20), 2299-301.
12. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. Electrospray ionization for mass spectrometry of large biomolecules. *Science* **1989**, 246, (4926), 64-71.
13. Dole, M.; Mack, L. L.; Hines, R. L.; Mobley, R. C.; Ferguson, L. D.; Alice, M. B. Molecular Beams of Macroions. *The Journal of Chemical Physics* **1968**, 49, (5), 2240-2249.
14. Iribarne, J. V.; Thomson, B. A. On the evaporation of small ions from charged droplets. *The Journal of Chemical Physics* **1976**, 64, (6), 2287-2294.
15. Wilm, M.; Mann, M., Analytical properties of the nanoelectrospray ion source. *Analytical chemistry* **1996**, 68, (1), 1-8.
16. Li, J.; Tremblay, T. L.; Harrison, J.; Thibault, P. Integrated system for rapid proteomics analyses using microfluidic devices coupled to nanoelectrospray mass spectrometry. *Methods in molecular biology* **2004**, 276, 305-24.
17. Shen, Y.; Zhao, R.; Berger, S. J.; Anderson, G. A.; Rodriguez, N.; Smith, R. D. High-efficiency nanoscale liquid chromatography coupled on-line with mass spectrometry using nanoelectrospray ionization for proteomics. *Analytical chemistry* **2002**, 74, (16), 4235-49.
18. Koster, S.; Verpoorte, E. A decade of microfluidic analysis coupled with electrospray mass spectrometry: an overview. *Lab on a chip* **2007**, 7, (11), 1394-412.
19. Sikanen, T.; Franssila, S.; Kauppila, T. J.; Kostianen, R.; Kotiaho, T.; Ketola, R. A. Microchip technology in mass spectrometry. *Mass spectrometry reviews* **2010**, 29, (3), 351-91.

20. He, J.; Yu, Q.; Li, L.; Hang, W.; Huang, B. Characteristics and comparison of different radiofrequency-only multipole cooling cells. *Rapid communications in mass spectrometry* **2008**, *22*, (20), 3327-33.
21. Giles, K.; Pringle, S. D.; Worthington, K. R.; Little, D.; Wildgoose, J. L.; Bateman, R. H. Applications of a travelling wave-based radio-frequency-only stacked ring ion guide. *Rapid communications in mass spectrometry* **2004**, *18*, (20), 2401-14.
22. Shaffer, S. A.; Prior, D. C.; Anderson, G. A.; Udseth, H. R.; Smith, R. D. An ion funnel interface for improved ion focusing and sensitivity using electrospray ionization mass spectrometry. *Analytical chemistry* **1998**, *70*, (19), 4111-9.
23. Shaffer, S. A.; Tolmachev, A.; Prior, D. C.; Anderson, G. A.; Udseth, H. R.; Smith, R. D. Characterization of an improved electrodynamic ion funnel interface for electrospray ionization mass spectrometry. *Analytical chemistry* **1999**, *71*, (15), 2957-64.
24. Kelly, R. T.; Tolmachev, A. V.; Page, J. S.; Tang, K.; Smith, R. D. The ion funnel: theory, implementations, and applications. *Mass spectrometry reviews* **2010**, *29*, (2), 294-312.
25. Shaffer, S. A.; Tang, K.; Anderson, G. A.; Prior, D. C.; Udseth, H. R.; Smith, R. D. A novel ion funnel for focusing ions at elevated pressure using electrospray ionization mass spectrometry. *Rapid Communications in Mass Spectrometry* **1997**, *11*, (16), 1813-1817.
26. Schwartz, J. C.; Senko, M. W.; Syka, J. E. A two-dimensional quadrupole ion trap mass spectrometer. *Journal of the American Society for Mass Spectrometry* **2002**, *13*, (6), 659-69.
27. Paul, W.; Steinwedel, H. Ein neues Massenspektrometer ohne Magnetfeld. *Zeitschrift für Naturforschung A* **1953**, *8*, (7), 448-450.
28. Riter, L. S.; Gooding, K. M.; Hodge, B. D.; Julian, R. K., Jr. Comparison of the Paul ion trap to the linear ion trap for use in global proteomics. *Proteomics* **2006**, *6*, (6), 1735-40.
29. Wolff, M. M.; Stephens, W. E. A Pulsed Mass Spectrometer with Time Dispersion. *Review of Scientific Instruments* **1953**, *24*, (8), 616-617.
30. Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S. Fourier transform ion cyclotron resonance mass spectrometry: a primer. *Mass spectrometry reviews* **1998**, *17*, (1), 1-35.
31. Perry, R. H.; Cooks, R. G.; Noll, R. J. Orbitrap mass spectrometry: instrumentation, ion motion and applications. *Mass spectrometry reviews* **2008**, *27*, (6), 661-99.
32. Aebersold, R.; Mann, M. Mass spectrometry-based proteomics. *Nature* **2003**, *422*, (6928), 198-207.
33. Pozo, O. J.; Van Eenoo, P.; Deventer, K.; Elbardissy, H.; Grimalt, S.; Sancho, J. V.; Hernandez, F.; Ventura, R.; Delbeke, F. T. Comparison between triple quadrupole, time of flight and hybrid quadrupole time of flight analysers coupled to liquid chromatography for the detection of anabolic steroids in doping control analysis. *Analytica chimica acta* **2011**, *684*, (1-2), 98-111.
34. Shevchenko, A.; Loboda, A.; Shevchenko, A.; Ens, W.; Standing, K. G. MALDI quadrupole time-of-flight mass spectrometry: A powerful tool for proteomic research. *Analytical Chemistry* **2000**, *72*, (9), 2132-2141.
35. Suckau, D.; Resemann, A.; Schuerenberg, M.; Hufnagel, P.; Franzen, J.; Holle, A. A novel MALDI LIFT-TOF/TOF mass spectrometer for proteomics. *Analytical and bioanalytical chemistry* **2003**, *376*, (7), 952-65.
36. Campana, J. E. Elementary theory of the quadrupole mass filter. *International Journal of Mass Spectrometry and Ion Physics* **1980**, *33*, (2), 101-117.
37. Campana, J. E.; Jurs, P. C. Computer-Simulation of the Quadrupole Mass Filter. *International Journal of Mass Spectrometry and Ion Processes* **1980**, *33*, (2), 119-137.
38. Vestal, M. L.; Juhasz, P.; Martin, S. A. Delayed Extraction Matrix-Assisted Laser-Desorption Time-of-Flight Mass-Spectrometry. *Rapid Communications in Mass Spectrometry* **1995**, *9*, (11), 1044-1050.
39. Paul, W.; Steinwedel, H. Apparatus For Separating Charged Particles Of Different Specific Charges. US Patent 2,939,952, June 7, 1960 1960.

40. Stafford Jr, G. C.; Kelley, P. E.; Syka, J. E. P.; Reynolds, W. E.; Todd, J. F. J. Recent improvements in and analytical applications of advanced ion trap technology. *International Journal of Mass Spectrometry and Ion Processes* **1984**, 60, (1), 85-98.
41. March, R. E. Quadrupole ion trap. *Mass Spec Rev* **2009**, 28, 961–989.
42. March, R. E. Quadrupole ion trap mass spectrometry: a view at the turn of the century. *Int J Mass Spectrom* **2000**, 200, 285–312.
43. Wiza, J. Microchannel plate detectors. *Nuclear Instruments and Methods* **1979**, 162, 587-601.
44. Kassel, D. B.; Biemann, K. Differentiation of hydroxyproline isomers and isobars in peptides by tandem mass spectrometry. *Analytical chemistry* **1990**, 62, (15), 1691-5.
45. Biemann, K. Sequencing of peptides by tandem mass spectrometry and high-energy collision-induced dissociation. *Methods in enzymology* **1990**, 193, 455-79.
46. Biemann, K. Appendix 5. Nomenclature for peptide fragment ions (positive ions). *Methods in enzymology* **1990**, 193, 886-7.
47. Dongre, A. R.; Somogyi, A.; Wysocki, V. H. Surface-induced dissociation: an effective tool to probe structure, energetics and fragmentation mechanisms of protonated peptides. *Journal of mass spectrometry* **1996**, 31, (4), 339-50.
48. Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **1970**, 227, (5259), 680-5.
49. Klose, J. From 2-D electrophoresis to proteomics. *Electrophoresis* **2009**, 30 Suppl 1, S142-9.
50. Di Girolamo, F.; Ponzi, M.; Crescenzi, M.; Alessandrini, J.; Guadagni, F. A simple and effective method to analyze membrane proteins by SDS-PAGE and MALDI mass spectrometry. *Anticancer research* **2010**, 30, (4), 1121-9.
51. Zhang, L.; Xie, J.; Wang, X.; Liu, X.; Tang, X.; Cao, R.; Hu, W.; Nie, S.; Fan, C.; Liang, S. Proteomic analysis of mouse liver plasma membrane: use of differential extraction to enrich hydrophobic membrane proteins. *Proteomics* **2005**, 5, (17), 4510-24.
52. Rabilloud, T.; Vaezzadeh, A. R.; Potier, N.; Lelong, C.; Leize-Wagner, E.; Chevallet, M. Power and limitations of electrophoretic separations in proteomics strategies. *Mass spectrometry reviews* **2009**, 28, (5), 816-43.
53. Gorg, A.; Weiss, W.; Dunn, M. J. Current two-dimensional electrophoresis technology for proteomics. *Proteomics* **2004**, 4, (12), 3665-85.
54. Rabilloud, T.; Lelong, C. Two-dimensional gel electrophoresis in proteomics: a tutorial. *Journal of proteomics* **2011**, 74, (10), 1829-41.
55. Raman, B.; Cheung, A.; Marten, M. R. Quantitative comparison and evaluation of two commercially available, two-dimensional electrophoresis image analysis software packages, Z3 and Melanie. *Electrophoresis* **2002**, 23, (14), 2194-2202.
56. Shi, Y.; Xiang, R.; Horvath, C.; Wilkins, J. A. The role of liquid chromatography in proteomics. *Journal of chromatograph. A* **2004**, 1053, (1-2), 27-36.
57. Azarkan, M.; Huet, J.; Baeyens-Volant, D.; Looze, Y.; Vandebussche, G. Affinity chromatography: a useful tool in proteomics studies. *Journal of chromatography B* **2007**, 849, (1-2), 81-90.
58. Banks, C. A.; Kong, S. E.; Washburn, M. P. Affinity purification of protein complexes for analysis by multidimensional protein identification technology. *Protein expression and purification* **2012**, 86, (2), 105-119.
59. Isono, E.; Schwechheimer, C. Co-immunoprecipitation and protein blots. *Methods in molecular biology* **2010**, 655, 377-87.
60. Waugh, D. S. Making the most of affinity tags. *Trends in biotechnology* **2005**, 23, (6), 316-20.

61. Puig, O.; Caspary, F.; Rigaut, G.; Rutz, B.; Bouveret, E.; Bragado-Nilsson, E.; Wilm, M.; Seraphin, B. The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods* **2001**, *24*, (3), 218-29.
62. Petrotchenko, E. V.; Borchers, C. H. Crosslinking combined with mass spectrometry for structural proteomics. *Mass spectrometry reviews* **2010**, *29*, (6), 862-76.
63. Stengel, F.; Aebersold, R.; Robinson, C. V. Joining forces: integrating proteomics and cross-linking with the mass spectrometry of intact complexes. *Molecular & cellular proteomics* **2012**, *11*, (3), R111 014027.
64. Kocher, T.; Superti-Furga, G. Mass spectrometry-based functional proteomics: from molecular machines to protein networks. *Nature methods* **2007**, *4*, (10), 807-15.
65. Neverova, I.; Van Eyk, J. E. Role of chromatographic techniques in proteomic analysis. *Journal of chromatograph. B* **2005**, *815*, (1-2), 51-63.
66. Bechade, G. Avancées en analyse protéomique par spectrométrie de masse: recherche de biomarqueurs et caractérisation de complexes covalents. *Thèse de l'Université de Strasbourg*, Strasbourg, 2009.
67. Shen, Y.; Smith, R. D.; Unger, K. K.; Kumar, D.; Lubda, D. Ultrahigh-throughput proteomics using fast RPLC separations with ESI-MS/MS. *Analytical chemistry* **2005**, *77*, (20), 6692-701.
68. Yin, H.; Killeen, K.; Brennen, R.; Sobek, D.; Werlich, M.; van de Goor, T. Microfluidic chip for peptide analysis with an integrated HPLC column, sample enrichment column, and nanoelectrospray tip. *Analytical chemistry* **2005**, *77*, (2), 527-33.
69. Premstaller, A.; Oberacher, H.; Walcher, W.; Timperio, A. M.; Zolla, L.; Chervet, J. P.; Cavusoglu, N.; van Dorsselaer, A.; Huber, C. G. High-performance liquid chromatography-electrospray ionization mass spectrometry using monolithic capillary columns for proteomic studies. *Analytical chemistry* **2001**, *73*, (11), 2390-6.
70. Liang, Y.; Zhang, L.; Zhang, Y. [Preparation of monolithic materials and their applications in proteomic analysis]. *Chinese journal of chromatography* **2011**, *29*, (9), 805-15.
71. Link, A. J.; Eng, J.; Schieltz, D. M.; Carmack, E.; Mize, G. J.; Morris, D. R.; Garvik, B. M.; Yates, J. R. Direct analysis of protein complexes using mass spectrometry. *Nature biotechnology* **1999**, *17*, (7), 676-82.
72. Washburn, M. P.; Wolters, D.; Yates, J. R. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature biotechnology* **2001**, *19*, (3), 242-7.
73. Wolters, D. A.; Washburn, M. P.; Yates, J. R. An automated multidimensional protein identification technology for shotgun proteomics. *Analytical chemistry* **2001**, *73*, (23), 5683-90.
74. Chen, J.; Lee, C. S.; Shen, Y.; Smith, R. D.; Baehrecke, E. H. Integration of capillary isoelectric focusing with capillary reversed-phase liquid chromatography for two-dimensional proteomics separation. *Electrophoresis* **2002**, *23*, (18), 3143-8.
75. Delmotte, N.; Lasaosa, M.; Tholey, A.; Heinzle, E.; Huber, C. G. Two-dimensional reversed-phase x ion-pair reversed-phase HPLC: an alternative approach to high-resolution peptide separation for shotgun proteome analysis. *Journal of proteome research* **2007**, *6*, (11), 4363-73.
76. Nesvizhskii, A. I.; Aebersold, R. Interpretation of shotgun proteomic data: the protein inference problem. *Molecular & cellular proteomics* **2005**, *4*, (10), 1419-40.
77. Pappin, D. J.; Hojrup, P.; Bleasby, A. J. Rapid identification of proteins by peptide-mass fingerprinting. *Current biology* **1993**, *3*, (6), 327-32.
78. Clauser, K. R.; Baker, P.; Burlingame, A. L. Role of accurate mass measurement (± 10 ppm) in protein identification strategies employing MS or MS/MS and database searching. *Analytical chemistry* **1999**, *71*, (14), 2871-82.
79. Horn, D. M.; Peters, E. C.; Klock, H.; Meyers, A.; Brock, A. Improved protein identification using automated high mass measurement accuracy MALDI FT-ICR MS peptide mass fingerprinting. *International Journal of Mass Spectrometry* **2004**, *238*, (2), 189-196.

80. Eng, J.; McCormack, A.; Yates, J. 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* **1994**, *5*, (11), 976-989.
81. Mann, M.; Wilm, M. Error-tolerant identification of peptides in sequence databases by peptide sequence tags. *Analytical chemistry* **1994**, *66*, (24), 4390-9.
82. Hakansson, K.; Zubarev, R.; Hakansson, P. Combination of nozzle-skimmer fragmentation and partial acid hydrolysis in electrospray ionization time-of-flight mass spectrometry of synthetic peptides. *Rapid communications in mass spectrometry* **1998**, *12*, (11), 705-11.
83. Purvine, S.; Eppel, J. T.; Yi, E. C.; Goodlett, D. R. Shotgun collision-induced dissociation of peptides using a time of flight mass analyzer. *Proteomics* **2003**, *3*, (6), 847-50.
84. Plumb, R. S.; Johnson, K. A.; Rainville, P.; Smith, B. W.; Wilson, I. D.; Castro-Perez, J. M.; Nicholson, J. K. UPLC/MS(E); a new approach for generating molecular fragment information for biomarker structure elucidation. *Rapid communications in mass spectrometry* **2006**, *20*, (13), 1989-94.
85. Chakraborty, A. B.; Berger, S. J.; Gebler, J. C. Use of an integrated MS--multiplexed MS/MS data acquisition strategy for high-coverage peptide mapping studies. *Rapid communications in mass spectrometry* **2007**, *21*, (5), 730-44.
86. Venable, J. D.; Dong, M. Q.; Wohlschlegel, J.; Dillin, A.; Yates, J. R. Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nature methods* **2004**, *1*, (1), 39-45.
87. Gillet, L. C.; Navarro, P.; Tate, S.; Rost, H.; Selevsek, N.; Reiter, L.; Bonner, R.; Aebersold, R. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Molecular & cellular proteomics* **2012**, *11*, (6), O111 016717.
88. Turck, C. W.; Falick, A. M.; Kowalak, J. A.; Lane, W. S.; Lilley, K. S.; Phinney, B. S.; Weintraub, S. T.; Witkowska, H. E.; Yates, N. A. The Association of Biomolecular Resource Facilities Proteomics Research Group 2006 study: relative protein quantitation. *Molecular & cellular proteomics* **2007**, *6*, (8), 1291-8.
89. Liu, H.; Sadygov, R. G.; Yates, J. R. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Analytical chemistry* **2004**, *76*, (14), 4193-201.
90. Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **1999**, *20*, (18), 3551-67.
91. Geer, L. Y.; Markey, S. P.; Kowalak, J. A.; Wagner, L.; Xu, M.; Maynard, D. M.; Yang, X.; Shi, W.; Bryant, S. H. Open mass spectrometry search algorithm. *Journal of proteome research* **2004**, *3*, (5), 958-64.
92. Craig, R.; Beavis, R. C. TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* **2004**, *20*, (9), 1466-7.
93. Brosch, M.; Swamy, S.; Hubbard, T.; Choudhary, J. Comparison of Mascot and X!Tandem performance for low and high accuracy mass spectrometry and the development of an adjusted Mascot threshold. *Molecular & cellular proteomics* **2008**, *7*, (5), 962-70.
94. Vaudel, M.; Barsnes, H.; Berven, F. S.; Sickmann, A.; Martens, L. SearchGUI: An open-source graphical user interface for simultaneous OMSSA and X!Tandem searches. *Proteomics* **2011**, *11*, (5), 996-9.
95. Dagda, R. K.; Sultana, T.; Lyons-Weiler, J. Evaluation of the Consensus of Four Peptide Identification Algorithms for Tandem Mass Spectrometry Based Proteomics. *Journal of proteomics & bioinformatics* **2010**, *3*, 39-47.
96. Balgley, B. M.; Laudeman, T.; Yang, L.; Song, T.; Lee, C. S. Comparative evaluation of tandem MS search algorithms using a target-decoy search strategy. *Molecular & cellular proteomics* **2007**, *6*, (9), 1599-608.
97. Bakalarski, C. E.; Haas, W.; Dephoure, N. E.; Gygi, S. P. The effects of mass accuracy, data acquisition speed, and search algorithm choice on peptide identification rates in phosphoproteomics. *Analytical and bioanalytical chemistry* **2007**, *389*, (5), 1409-19.

98. Bjornson, R. D.; Carriero, N. J.; Colangelo, C.; Shifman, M.; Cheung, K. H.; Miller, P. L.; Williams, K. X!!Tandem, an improved method for running X!tandem in parallel on collections of commodity computers. *Journal of proteome research* **2008**, 7, (1), 293-9.
99. Dancik, V.; Addona, T. A.; Clauser, K. R.; Vath, J. E.; Pevzner, P. A. De novo peptide sequencing via tandem mass spectrometry. *Journal of computational biology* **1999**, 6, (3-4), 327-42.
100. Grossmann, J.; Roos, F. F.; Cieliebak, M.; Liptak, Z.; Mathis, L. K.; Muller, M.; Gruissem, W.; Baginsky, S. AUDENS: a tool for automated peptide de novo sequencing. *Journal of proteome research* **2005**, 4, (5), 1768-74.
101. He, L.; Ma, B. ADEPTS: advanced peptide de novo sequencing with a pair of tandem mass spectra. *Journal of bioinformatics and computational biology* **2010**, 8, (6), 981-94.
102. Ma, B.; Zhang, K.; Hendrie, C.; Liang, C.; Li, M.; Doherty-Kirby, A.; Lajoie, G. PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid communications in mass spectrometry* **2003**, 17, (20), 2337-42.
103. Frank, A.; Pevzner, P. PepNovo: de novo peptide sequencing via probabilistic network modeling. *Anal Chem* **2005**, 77, (4), 964-73.
104. Chong, K. F.; Leong, H. W. Tutorial on De Novo Peptide Sequencing Using Ms/Ms Mass Spectrometry. *Journal of bioinformatics and computational biology* **2012**, 1231002.
105. Sakurai, T.; Matsuo, T.; Matsuda, H.; Katakuse, I. Paas-3 - a Computer-Program to Determine Probable Sequence of Peptides from Mass-Spectrometric Data. *Biomedical mass spectrometry* **1984**, 11, (8), 396-399.
106. Spengler, B. De novo sequencing, peptide composition analysis, and composition-based sequencing: a new strategy employing accurate mass determination by fourier transform ion cyclotron resonance mass spectrometry. *Journal of the American Society for Mass Spectrometry* **2004**, 15, (5), 703-14.
107. Heredia-Langner, A.; Cannon, W. R.; Jarman, K. D.; Jarman, K. H. Sequence optimization as an alternative to de novo analysis of tandem mass spectrometry data. *Bioinformatics* **2004**, 20, (14), 2296-304.
108. Johnson, R. S.; Taylor, J. A. Searching sequence databases via de novo peptide sequencing by tandem mass spectrometry. *Molecular biotechnology* **2002**, 22, (3), 301-15.
109. Pevtsov, S.; Fedulova, I.; Mirzaei, H.; Buck, C.; Zhang, X. Performance evaluation of existing de novo sequencing algorithms. *Journal of proteome research* **2006**, 5, (11), 3018-28.
110. Bringans, S.; Kendrick, T. S.; Lui, J.; Lipscombe, R. A comparative study of the accuracy of several de novo sequencing software packages for datasets derived by matrix-assisted laser desorption/ionisation and electrospray. *Rapid communications in mass spectrometry* **2008**, 22, (21), 3450-4.
111. Altschul, S. F.; Madden, T. L.; Schaffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research* **1997**, 25, (17), 3389-402.
112. Pearson, W. R.; Lipman, D. J. Improved tools for biological sequence comparison. *Proceedings of the National Academy of Sciences of the United States of America* **1988**, 85, (8), 2444-8.
113. Taylor, J. A.; Johnson, R. S. Sequence database searches via de novo peptide sequencing by tandem mass spectrometry. *Rapid communications in mass spectrometry* **1997**, 11, (9), 1067-75.
114. Shevchenko, A.; Sunyaev, S.; Loboda, A.; Bork, P.; Ens, W.; Standing, K. G. Charting the proteomes of organisms with unsequenced genomes by MALDI-quadrupole time-of-flight mass spectrometry and BLAST homology searching. *Analytical chemistry* **2001**, 73, (9), 1917-26.
115. Mackey, A. J.; Haystead, T. A.; Pearson, W. R. Getting more from less: algorithms for rapid protein identification with multiple short peptide sequences. *Molecular & cellular proteomics* **2002**, 1, (2), 139-47.
116. Waridel, P.; Frank, A.; Thomas, H.; Surendranath, V.; Sunyaev, S.; Pevzner, P.; Shevchenko, A. Sequence similarity-driven proteomics in organisms with unknown genomes by LC-MS/MS and automated de novo sequencing. *Proteomics* **2007**, 7, (14), 2318-29.

117. Catusse, J.; Strub, J. M.; Job, C.; Van Dorsselaer, A.; Job, D. Proteome-wide characterization of sugarbeet seed vigor and its tissue specific expression. *Proceedings of the National Academy of Sciences of the United States of America* **2008**, 105, (29), 10262-7.
118. Hatano, N.; Hamada, T. Proteome analysis of pitcher fluid of the carnivorous plant *Nepenthes alata*. *Journal of proteome research* **2008**, 7, (2), 809-16.
119. Apweiler, R.; Bairoch, A.; Wu, C. H. Protein sequence databases. *Current opinion in chemical biology* **2004**, 8, (1), 76-80.
120. Apweiler, R.; Bairoch, A.; Wu, C. H.; Barker, W. C.; Boeckmann, B.; Ferro, S.; Gasteiger, E.; Huang, H.; Lopez, R.; Magrane, M.; Martin, M. J.; Natale, D. A.; O'Donovan, C.; Redaschi, N.; Yeh, L. S. UniProt: the Universal Protein knowledgebase. *Nucleic acids research* **2004**, 32, D115-9.
121. Bairoch, A.; Apweiler, R.; Wu, C. H.; Barker, W. C.; Boeckmann, B.; Ferro, S.; Gasteiger, E.; Huang, H.; Lopez, R.; Magrane, M.; Martin, M. J.; Natale, D. A.; O'Donovan, C.; Redaschi, N.; Yeh, L. S. The Universal Protein Resource (UniProt). *Nucleic acids research* **2005**, 33, D154-9.
122. Wu, C. H.; Apweiler, R.; Bairoch, A.; Natale, D. A.; Barker, W. C.; Boeckmann, B.; Ferro, S.; Gasteiger, E.; Huang, H.; Lopez, R.; Magrane, M.; Martin, M. J.; Mazumder, R.; O'Donovan, C.; Redaschi, N.; Suzek, B. The Universal Protein Resource (UniProt): an expanding universe of protein information. *Nucleic acids research* **2006**, 34, D187-91.
123. UniProt. Ongoing and future developments at the Universal Protein Resource. *Nucleic acids research* **2011**, 39, D214-9.
124. Elias, J. E.; Gygi, S. P. Target-decoy search strategy for mass spectrometry-based proteomics. *Methods in molecular biology* **2010**, 604, 55-71.
125. Storey, J. D.; Tibshirani, R. Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences of the United States of America* **2003**, 100, (16), 9440-5.
126. Keller, A.; Nesvizhskii, A. I.; Kolker, E.; Aebersold, R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Analytical chemistry* **2002**, 74, (20), 5383-92.



Chapitre 2 :

Modifications post-traductionnelles: N-glycosylation

Ce chapitre présente l'état de l'art de l'analyse de glycoprotéines par spectrométrie de masse et les difficultés analytiques rencontrées lors de ces analyses.

Introduction

Les protéines contrôlent divers processus biologiques dans les organismes vivants grâce à leurs propriétés physico-chimiques. En effet, les 20 acides aminés codés génétiquement possèdent des propriétés chimiques suffisantes pour donner aux protéines certaines caractéristiques comme le repliement. Mais, des rôles beaucoup plus particuliers comme ceux liés au contrôle et à la régulation des processus biologiques demande que les protéines soient complétées par des modifications covalentes qui sont généralement introduites après la traduction: les modifications post-traductionnelles (PTMs) ¹.

Les PTMs sont des processus primordiaux de la biosynthèse des protéines au cours desquels, les protéines issues de la traduction des ARN (Acides Ribonucléiques) sont modifiées pour acquérir des fonctionnalités. Les PTMs affectent les protéines de façon réversible ou définitive après la traduction et étendent leurs propriétés et fonctionnalités en les rattachant à d'autres groupements fonctionnels biochimiques (par exemple, les phosphates, divers lipides et glucides), ou en changeant la nature chimique des acides aminés (par exemple, par la déamidation, la citrullination). Les PTMs vont fonctionnaliser les protéines afin qu'elles soient facilement reconnues par d'autres protéines ou pour être dirigées vers des cibles dans les différents compartiments de la cellule ².

Aujourd'hui plus de 400 PTMs ont été rapportées dans la littérature (www.uniprot.org/docs/ptmlist), et à ces modifications se rajoutent des modifications qui peuvent être introduites *in vitro* artificiellement dans les laboratoires (<http://www.unimod.org>). Bien que certaines modifications se produisent dans les cellules *in vivo*, il arrive parfois que quelques unes se produisent artificiellement (artéfacts) *in vitro* au cours de la manipulation des échantillons. Il est dans ces cas difficiles de déterminer parfois si la modification est intervenue dans la cellule (ou tissu) ou après extraction des protéines ³. Les PTMs peuvent consister en :

- (1) une modification de la structure primaire de la protéine par addition ou clivage d'un ou plusieurs acides aminés (par exemple l'élimination de la méthionine N-terminal de certaines protéines) ;
- (2) une modification de l'extrémité N- et C-terminale de la protéine par la fixation d'un groupement (par exemple l'acétylation N-terminale des protéines) ;
- (3) une modification de la chaîne latérale des acides aminés par fixation covalente d'un groupement (par exemple la glycosylation ou la phosphorylation) ;
- (4) un clivage protéolytique.

Bien que l'importance des PTMs ait été reconnue très tôt en biologie, l'ampleur des PTMs sur les processus biologiques a été surtout révélée grâce à l'analyse protéomique et l'apparition des techniques d'enrichissement des protéines. En effet, en utilisant des stratégies d'enrichissement sélectives différentes sur un même échantillon, il n'est pas rare d'identifier plusieurs modifications dans un seul échantillon comme le cas des phosphorylations et des glycosylations ⁴. Le caractère réversible de certaines PTMs et l'occupation non stœchiométrique de nombreux sites comme l'illustre le cas de la glycosylation où plusieurs sites peuvent ou non être occupés sur une même protéine, rend l'analyse des protéines affectées par des PTMs particulièrement difficile à cause de cette hétérogénéité. En plus de l'hétérogénéité des PTMs, il peut toujours exister des problèmes de distinction entre certaines PTMs comme la sulfatation et la phosphorylation qui ont la même masse (PTMs isobariques). Il y a aussi de nombreuses PTMs qui ont été observées souvent dans des expériences de protéomique et qui n'ont pas encore été clairement caractérisées (PTMs inconnues). En revanche, la distinction, l'identification précise et fiable du site occupé par des PTMs reste un réel challenge analytique et bien souvent constitue une source potentielle d'erreurs dans les résultats publiés. Mais en règle générale, des données de spectrométrie de masse de bonne qualité sont nécessaires pour identifier sans ambiguïté les résidus exactement modifiés dans la séquence des protéines modifiées.

La diversité et l'étendue des PTMs conduisent donc à des défis analytiques supplémentaires par rapport à l'analyse d'une protéine non modifiée, qui vont au-delà de la simple identification d'une protéine lors d'analyse protéomique par spectrométrie de masse ⁵. Aujourd'hui la meilleure connaissance des certaines modifications post-traductionnelles comme la N-glycosylation ou la phosphorylation a permis de connaître des motifs d'acides aminés (3-10 acides aminés) sur lesquels certaines d'entre elles peuvent préférentiellement se produire (sites consensus ⁶) et de prédire ainsi les sites potentiels possibles sur des protéines ^{7, 8} à l'aide d'outils bioinformatiques. Mais la prédiction reste toujours théorique, et seule l'identification expérimentale des PTMs sur une protéine permet de confirmer son existence.

Nous nous intéresserons dans les prochains paragraphes à la glycosylation et plus particulièrement la N-glycosylation qui est l'une des plus importantes PTMs qui affectent les protéines des procaryotes et eucaryotes.

I. La glycosylation : une PTM courante

Une glycosylation est le résultat d'une réaction enzymatique conduisant à une liaison covalente de glucides (sucres, glycanes) sur la chaîne latérale d'un acide aminé d'une protéine. Les enzymes qui catalysent la glycosylation de protéines sont en général des glycosyltransférases⁹. Ce processus enzymatique, qui affecte plus souvent les protéines membranaires ainsi que les protéines sécrétées, a lieu dans le RER (Réticulum Endoplasmique Rugueux) et/ou dans l'appareil de Golgi. Les glucides impliqués dans les glycosylations peuvent être des monosaccharides ou une succession de monosaccharides qui forment des structures polysaccharidiques (oligosaccharides) variées. Les monosaccharides les plus connus sont indiqués dans le Tableau 1.

Tableau 1. Récapitulatifs des noms, des abréviations, des symboles et des masses isotopiques des résidus des monosaccharides les plus cités en littérature et observés dans les structures des oligosaccharides. La nomenclature et les symboles ont été établis par le "Consortium for functional Glycomics" (www.functionalglycomics.org)¹⁰.

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

Une étude récente réalisée dans la banque SWISS-PROT a permis de montrer que les glycosylations figurent parmi les trois PTMs les plus répandues affectant le plus de protéines¹¹. On estime à plus de 20% la proportion des protéines affectées par des glycosylations. Cette PTM participe à divers processus biologiques dont la transduction de signal, le repliement des protéines, la stabilité des interactions cellules-cellules et virus-cellules ou encore les réponses immunitaires¹²⁻¹⁸. Les défauts de glycosylation sont responsables de maladies congénitales sévères ou peuvent dans certains cas conduire à des anomalies de développement embryonnaire ou à des cas de pathologies comme le cancer, des maladies inflammatoires ou du foie¹⁹⁻²¹. La caractérisation des glycosylations est donc importante non seulement pour le développement des vaccins et des médicaments mais aussi pour le diagnostic de certaines pathologies^{20,22}.

Il existe plusieurs types de glycosylations, mais les deux types les plus fréquents retrouvés chez les procaryotes et les eucaryotes sont : les O-glycosylations (O-linked glycosylation) et les N-glycosylations (N-linked glycosylation) (Figure 1).

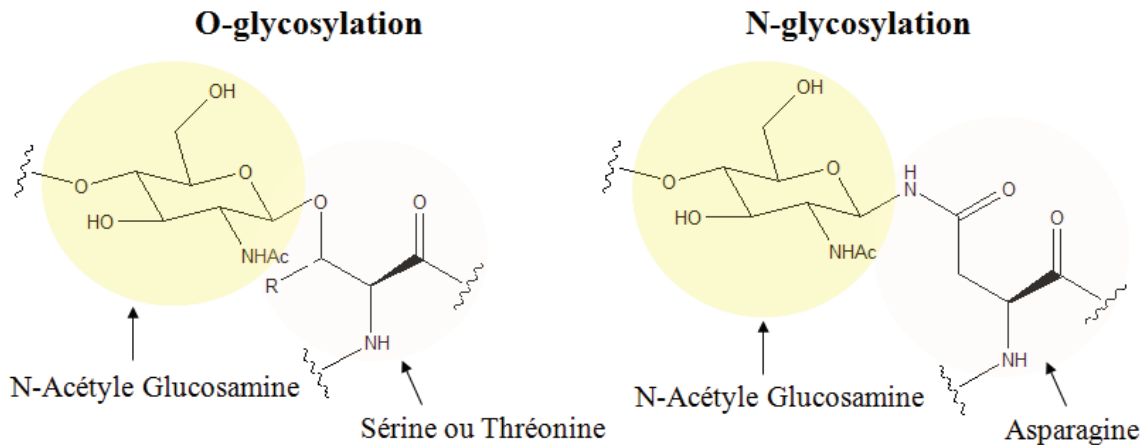


Figure 1. Structures chimiques des liaisons sucre-protéine impliquées dans les O- et N-glycosylations.

Les O-glycosylations : dans une O-glycosylation, les résidus de monosaccharides sont liés aux protéines par le groupement hydroxyle des résidus de sérine (Ser/S), de la thréonine (Thr/T), de l'hydroxyproline (Hyp) ou l'hydroxylysine. La plus répandue des O-glycosylation est de type mucin où le premier sucre lié à la sérine ou à la thréonine est le N-Acétyleglucosamine (GlcNAc). Cette glycosylation est un processus biologique qui a lieu en général dans l'appareil de Golgi et peut conduire à de longues chaînes de sucres souvent linéaires²³. Il n'existe pas de sites consensus pour la O-glycosylation, c'est-à-dire que la O-glycosylation peut se produire sur n'importe quelle sérine ou thréonine d'une protéine²⁴.

Les N-glycosylations : dans une N-glycosylation, les résidus de monosaccharides sont rattachés aux protéines par le groupement amide d'une asparagine (Asn/N) via le premier monosaccharide qui est un N-Acétyleglucosamine (GlcNAc). Contrairement aux O-glycosylations qui n'ont pas de sites consensus, les N-glycosylations s'effectuent sur l'asparagine située dans un motif Asn-X-Ser/Thr où X peut être n'importe quel acide aminé à l'exception de la proline (Pro/P). Cependant dans de rares cas de N-glycosylations la Ser/Thr du motif est remplacée par une cystéine²⁵. Les sucres des N-glycosylations ont une structure beaucoup plus complexe que les sucres des O-glycosylations avec plusieurs ramifications appelées : les antennes. La voie de biosynthèse étant commune pour les N-glycoprotéines, tous les oligosaccharides des N-glycosylations ont un squelette identique (conserved core) (Figure 2, encadré en rouge) qui est constitué de trois mannoses et deux N-acétyleglucosamines (Man₃GlcNAc₂)^{12,26}.

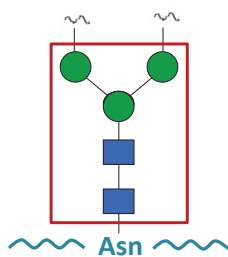


Figure 2. Structure de base (conserved core, squelette commun) des oligosaccharides des N-glycosylations : Man₃GlcNAc₂ (encadré en rouge). ■ N-acétyleglucosamine, ● Mannose.

Dans les prochains paragraphes de ce chapitre nous nous focaliserons exclusivement sur le cas de la N-glycosylation qui a fait l'objet d'une étude dans la partie Résultats de Chapitre 4 de ce travail de thèse.

II. Difficultés d'analyse des glycoprotéines : cas des N-glycosylations

1. Complexité de la N-glycosylation : Micro- et Macro-hétérogénéité.

Les N-glycosylations sont classées en 3 types suivant la composition des antennes ²⁷. On distingue ainsi (Figure 3):

- les sucres de type **high-mannose** où les antennes sont constituées uniquement de mannoses ;
- les sucres de type **complexe** qui correspondent à des antennes composées des N-Acétylglucosamines, des galactoses, des fucoses et/ou des acides sialiques ;
- les oligosaccharides de **type hybride**, qui sont composés des antennes de type highmannose d'un coté et des antennes de type complexe de l'autre.

Selon le nombre d'antennes que porte le squelette identique (conserved core), on parle d'oligosaccharides bi-, tri- ou tétra-antennés respectivement pour deux, trois, et quatre antennes.

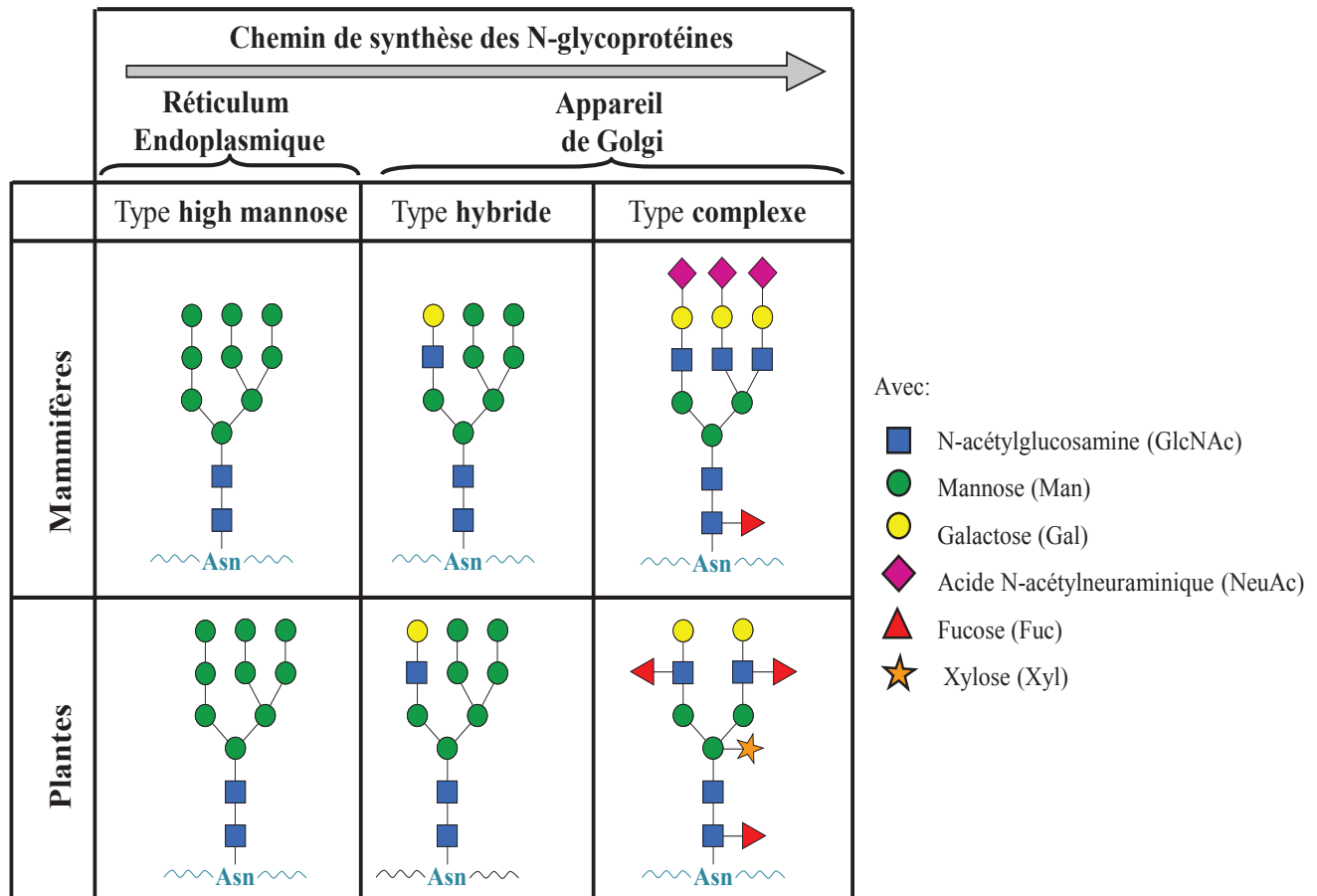


Figure 3. Exemple de structure de N-glycosylation chez les mammifères et les plantes.

En outre, ces structures ne sont pas statiques ni stœchiométriques, puisqu'elles varient de manière qualitative et quantitative en fonction de l'avancement du processus de biosynthèse dans le RER et/ou appareil de Golgi ²⁸. Une glycoprotéine (protéine glycosylée) va donc être un mélange de la protéine qui est dans plusieurs étapes de sa maturation (chaînes glycaniques plus ou moins longues). Les glycoprotéines vont posséder par exemple sur un même site (Asn) plusieurs structures d'oligosaccharides différentes. Les glycoprotéines vont donc être hétérogènes au niveau de ces structures oligosaccharidiques. Les différentes formes de glycosylations peuvent avoir des rôles très différents et entraîner des changements de conformation des protéines affectées qui vont totalement changer leur fonctionnalité biologique.

On distingue deux types d'hétérogénéité : la macro- et la micro-hétérogénéité ²⁹. Dans le cas de la **macro-hétérogénéité**, un ou plusieurs sites de glycosylation d'une glycoprotéine, vont ou non être glycosylés. Pour la **micro-hétérogénéité**, un site glycosylé va porter plusieurs structures d'oligosaccharides, et suivant l'avancement de la biosynthèse, les antennes (ramifications) vont être très variées et être plus ou moins longues (Figure 4).

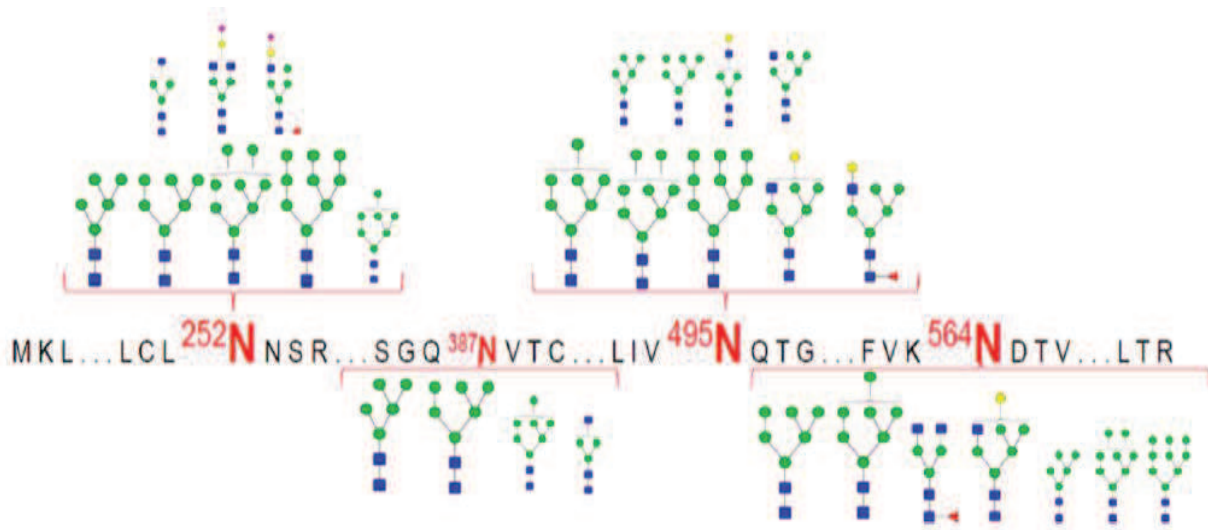


Figure 4. Détail de l'hétérogénéité des N-glycosylation et de l'abondance des glycoformes caractérisées sur la lactoferrine bovine ³⁰.

La Figure 4 indique les différents glycanes qui sont identifiés sur chaque Asn (micro-hétérogénéité), et l'ensemble des sites de glycosylation (macro-hétérogénéité) présents sur la lactoferrine bovine. Le nombre de glycoformes peut fortement varier sur un seul site de glycosylation, par exemple, la glycoprotéine CD59 (MAC-IP) humaine a plus de 100 différentes structures oligosaccharidiques sur un site de glycosylation ^{15, 31}. Les glycoprotéines peuvent devenir ainsi très hétérogènes et complexes. Ces raisons expliquent pourquoi l'analyse de glycoprotéines est souvent très complexe et pourquoi il est parfois difficile de corrélérer les profils d'expression d'une glycoprotéine à des états pathologiques ²⁹.

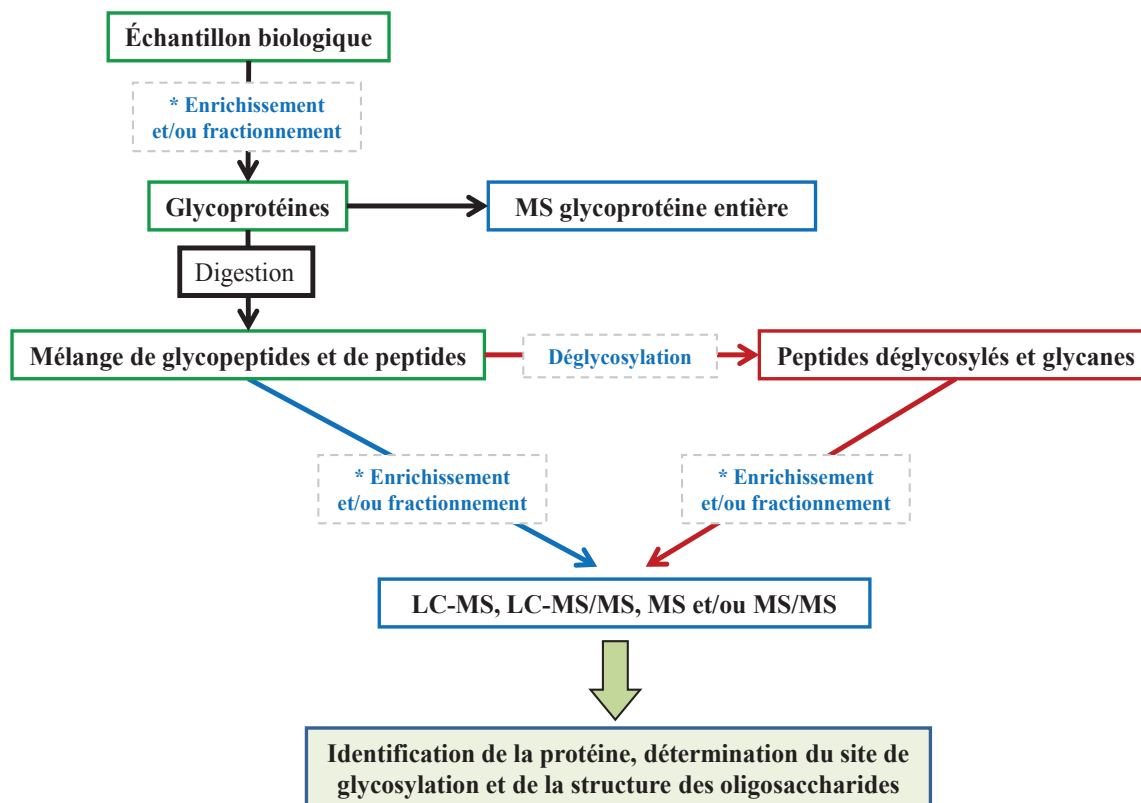
Le rôle des glycosylations dans de nombreux processus biologiques n'est pas encore complètement élucidé à cause de cette extrême complexité de la glycosylation et à cause des difficultés analytiques rencontrées au cours de la caractérisation de cette PTM. Le développement de nouvelles méthodologies d'analyse est donc indispensable pour parvenir à cette caractérisation et à la compréhension du rôle des glycosylations dans les processus biologiques.

2. Stratégie d'analyse de glycoprotéines par spectrométrie de masse

L'analyse d'une glycoprotéine va consister à (1) identifier et de caractériser les séquences d'acides aminés de la structure primaire de la glycoprotéine, mais aussi à (2) caractériser les oligosaccharides et d'identifier les sites Asn qu'occupent les oligosaccharides. L'analyse d'une glycoprotéine va donc être plus difficile que celle d'une protéines non glycosylée puisque plusieurs autres paramètres vont entrer en jeu lors de l'analyse par spectrométrie de masse ³² comme :

- **la complexité de l'échantillon** : les différentes glycoprotéines peuvent être mineures dans un mélange de protéines, ou les différentes glycoformes être à différents ratio de concentration. Dans ce cas l'analyse directe de l'échantillon par spectrométrie de masse sans étapes d'enrichissement ne va pas permettre de détecter toutes les glycoprotéines mineures qui auront leurs signaux complètement masqués par ceux des protéines majoritaires;
- **la complexité de la glycosylation** : plusieurs glycoprotéines peuvent exister dans le même échantillon, et ces glycoprotéines posséder plusieurs sites qui sont glycosylés, et les oligosaccharides au niveau de chaque site être hétérogènes. Cette hétérogénéité va rendre encore plus complexe l'échantillon à analyser, et entraîner une difficulté supplémentaire qui est la répartition du signal de la glycoprotéine entre toutes ces glycoformes.

Pour ces principales raisons, les stratégies couramment utilisées en analyse des glycoprotéines passent par l'enrichissement ou fractionnement des glycoprotéines ou glycopeptides pour réduire la complexité de l'échantillon ou alors par la déglycosylation des protéines pour réduire la complexité liée à la glycosylation³² afin de faciliter la caractérisation de glycoprotéines (Figure 5).



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

Figure 4. Stratégie d'analyse de glycoprotéines par spectrométrie de masse. Les glycoprotéines entières peuvent être analysées par MS généralement si elles sont enrichies et comportent un nombre réduit de glycoformes, elles peuvent être digérées en glycopeptides ou peptides déglycosylés et être analysés pour déterminer leur micro- et macro-hétérogénéité. Selon la complexité des échantillons, ceux-ci peuvent être enrichis/fractionnés en glycoprotéines ou glycopeptides par chromatographie d'affinité avant l'analyse par spectrométrie de masse (MS). Les oligosaccharides libérés par déglycosylation peuvent être analysés par MS ou MS/MS pour déterminer la composition des oligosaccharides.

La Figure 5 présente les techniques et les méthodologies couramment utilisées pour l'analyse de glycoprotéines par spectrométrie de masse. Ces approches comportent plusieurs étapes qui peuvent être résumées de la manière suivante:

- la purification des glycoprotéines à partir de l'échantillon biologique (étape souvent facultative si l'échantillon n'est pas complexe);
- l'analyse des glycoprotéines entières par MS (étape possible lorsque la glycoprotéine est pure et possède peu de glycoformes);
- la digestion des glycoprotéines par des endoprotéases afin d'obtenir un ensemble de peptides et glycopeptides. Cette étape va conduire souvent à l'analyse des glycopeptides entiers après enrichissement ou non et/ou par la suite à la déglycosylation de glycopeptides pour obtenir des peptides "nus" et des oligosaccharides libres qui vont être analysés par MS ou MS/MS;
- l'identification de la protéine, la détermination des sites de glycosylation et la structure des oligosaccharides.

3. Techniques d'enrichissement des glycoprotéines et glycopeptides

L'analyse protéomique par LC-MS et MS/MS est une méthode de choix pour identifier et caractériser des milliers de peptides issus de la digestion des protéines. Cependant, seul un nombre limité de précurseurs (ions parents) peuvent être sélectionnés et fragmentés dans un cycle MS et MS/MS (en mode d'acquisition

DDA le plus utilisé). Les espèces moins abondantes comme par exemple les peptides issus de protéines mineurs, sont défavorisées lors de cette sélection puisque leurs signaux en MS sont faibles par rapport aux espèces abondantes. Il est donc parfois important de procéder à une étape d'enrichissement des glycopeptides et glycoprotéines pour faciliter leur détection au cours de l'analyse par spectrométrie de masse. Cette étape d'enrichissement présente plusieurs avantages pour l'analyse de glycoprotéines. (1) L'élimination de peptides non glycosylés va permettre de décomplexifier et de purifier les glycoprotéines qui généralement moins abondante dans les échantillons ³³. (2) Les digestions à l'aide d'endoprotéases vont être plus efficaces sur les échantillons enrichis puisque les glycoprotéines sont plus abondantes et exposées à ces endoprotéases ³⁴. Les techniques d'enrichissement ou purification des glycoprotéines et glycopeptides couramment utilisées sont basées sur la chromatographie affinité et d'autres types de chromatographie comme l'échangeuse d'ions. Selon le type de chromatographie, il existe une panoplie de phases stationnaires qui peuvent être utilisées et les résultats obtenus peuvent varier en fonction de l'échantillon étudié. Le choix de la phase est avant tout basé sur le type de glycosylation que l'on veut enrichir.

a. La chromatographie d'affinité

Selon le type de phase stationnaire utilisée, les glycoprotéines ou glycopeptides vont se lier spécifiquement avec cette phase de la colonne de manière réversible ou irréversible.

• Les phases de lectines

Les lectines sont des protéines qui se lient spécifiquement et de façon réversible à certains glucides. Les propriétés des lectines sont utilisées pour purifier spécifiquement les glycoprotéines qui portent ces glucides. Les lectines sont alors fixées sur de la sépharose ou de l'agarose dans une colonne de chromatographie pour purifier des glycoprotéines contenus dans les échantillons. Il existe plusieurs lectines avec des spécificités très différentes. Par exemple, la ConA (concanavaline A, une lectine extraite du haricot sabre) est spécifique aux glucides contenant des mannoses ou des glucoses ³⁵, la WGA (Wheat germ agglutinin, une lectine extraite du soja) est spécifique à des sucres contenant les N-Acétyleglucosamines et les acides sialiques ^{36, 37}. Plusieurs lectines peuvent être combinées aussi dans une colonne ou des colonnes séparées pour augmenter non seulement l'efficacité d'enrichissement mais aussi la variété de glycoformes à enrichir lors de l'enrichissement ³⁸.

Malgré l'avantage qu'offre la spécificité des lectines pour les oligosaccharides portés par les glycoprotéines et glycopeptides, leur utilisation s'accompagne de quelques inconvénients majeurs qui peuvent entacher la qualité de données de l'analyse de glycoprotéines. En effet, vu que les lectines ont une affinité limitée à certains oligosaccharides de glycoprotéines ou glycopeptides et que cette affinité est transitoire et dépendante des tampons dans lesquels sont les échantillons, tous les types de glycoformes ne vont pas être enrichis sur les colonnes de lectines (les lectines sont sélectives). Des études récentes ont mis en évidence des problèmes relatifs à la spécificité et l'efficacité des lectines lors de l'enrichissement glycoprotéines provenant des mélanges de protéines extrêmement complexes ³⁹. Ces limitations de l'utilisation des lectines en particulier pour les mesures quantitatives ou qualitative des glycoprotéines ou glycopeptides sont donc à prendre en considération lors des analyses par spectrométrie de masse.

• Les phases hydrazides

Au cours de cette méthode d'enrichissement sur les phases hydrazides, l'oxydation des diols vicinaux cis d'oligosaccharides en aldéhyde est d'abord effectuée avec le periodate de sodium (NaIO₄). Ensuite, les groupements aldéhydes formés sont utilisés pour purifier les glycopeptides et les glycoprotéines par réaction chimique donnant lieu à une liaison covalente avec les phases hydrazides sous des conditions de pH contrôlées ⁴⁰⁻⁴² (Figure 5).

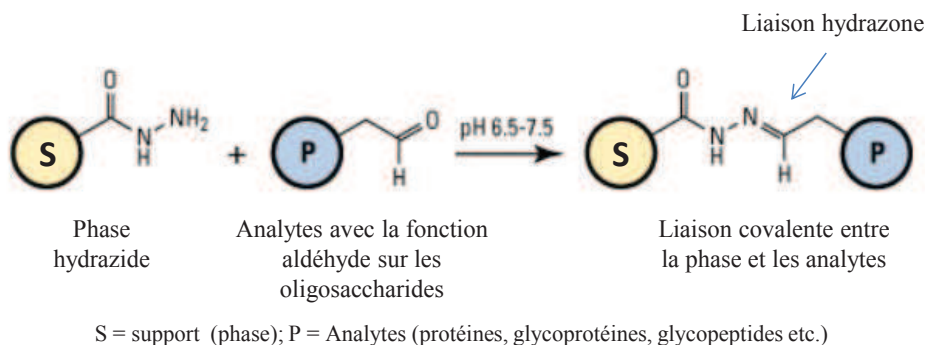


Figure 5. Schéma de la réaction entre les phases hydrazides et les glycopeptides ou glycopeptides. Les fonctions aldéhydes obtenues de l'oxydation des oligosaccharides vont être utilisées pour former des liaisons covalentes de types hydrazone avec les phases hydrazides sous des conditions de pH contrôlés. La formation de cette liaison va permettre de retenir les glycopeptides sur la phase et d'éliminer les protéines non glycosylées.

Un grand nombre de glycoprotéines et glycopeptides vont pouvoir être purifiés par cette méthode, puisque les groupements aldéhydes sont généralement absents dans les peptides et les protéines non glycosylées. Ce qui rend la technique assez spécifique des glycopeptides et glycoprotéines de façon générale. En outre, cette technique permet d'avoir des étapes de lavage plus rigoureuses sans risque de décrocher les analytes d'intérêt de la colonne puisqu'ils sont liés de façon covalente sur la phase. Ces étapes de lavage vont minimiser ainsi la co-purification par exemple des protéines non glycosylées. Les analytes d'intérêt vont ensuite être libérés par réactions enzymatiques (avec la PNGase A ou PNGase F) qui vont couper par exemple dans la structure des oligosaccharides pour libérer le reste des analytes (peptides déglycosylés) de la phase.

Dans la plupart des cas, les informations sur la micro-hétérogénéité sont perdues au cours de cette méthode d'enrichissement. De plus ces méthodes d'enrichissement comprennent plusieurs étapes au cours desquelles des quantités non négligeables de protéines peuvent être perdues. A cela s'ajoutent les modifications que peut induire la réaction d'oxydation sur certains acides aminés N-terminaux comme la sérine ou la théorine qui vont se fragmenter différemment par spectrométrie de masse⁴³ et donc difficilement identifiables.

- **Phase à base d'oxyde de métaux**

Des phases solides à base d'oxydes de titane (TiO₂) ont été utilisées récemment pour l'enrichissement des glycopeptides dont les oligosaccharides portent des acides sialiques⁴⁴. Le mécanisme de fonctionnement de cette chromatographie d'affinité avec les oxydes de métaux reste encore mal connu. Il faut noter que ces méthodes d'enrichissement sont aussi utilisées pour l'enrichissement d'autres PTMs comme le cas de la phosphorylation⁴⁵. L'utilisation de cette méthode peut entraîner des interférences par co-purification des protéines ou peptides phosphorylés⁴⁶. Pour l'instant l'utilisation de cette méthode d'enrichissement des glycoprotéines et glycopeptides n'est pas très répandue.

Comme dans le cas des phases hydrazides, l'information sur la micro-hétérogénéité sont souvent perdues avec l'utilisation de cette méthode d'enrichissement, puisque les peptides déglycosylés par exemple sont directement élués de la phase par l'action des endoprotéases qui clivent les oligosaccharides du leur peptides d'origine⁴⁷.

b. La chromatographie d'interaction hydrophile

Grâce aux groupements -OH portés par les oligosaccharides et les glycopeptides, ceux-ci vont être plus polaires et hydrophiles que la plupart des peptides issus de la digestion enzymatique des échantillons. Ce caractère des glycopeptides va pouvoir être exploité lors de la chromatographie d'interaction hydrophile (HILIC, hydrophilic liquid chromatography)⁴⁸ où la phase stationnaire est revêtue de groupements polaires considérée comme liquide. Ces groupements polaires sur la phase stationnaire vont interagir plus spécifiquement avec les glycopeptides pour les séparer des peptides non glycosylés qui sont en général moins polaires. Il se produit alors un phénomène semblable à l'extraction liquide-liquide où les composés hydrophiles vont passer de la phase mobile généralement concentrée en acétonitrile à la phase stationnaire polaire et hydrophile. Par la suite ces composés vont pouvoir être élués séparément pour l'analyse par spectrométrie de masse. Plusieurs phases peuvent être utilisées dans ce type de chromatographie comme par exemple les phases

ZIC-HILIC (Zwitterionic Hydrophilic Interaction Liquid Chromatography), ou des billes de silice ou sépharose modifiées^{40, 49, 50}.

Cette approche d'enrichissement des glycopeptides permet de conserver la structure intacte des oligosaccharides. Cependant, une co-purification de peptides hydrophiles non glycosylés restent toujours possible avec cette technique d'enrichissement. Mais, l'efficacité de purification peut toutefois être améliorée en modifiant la concentration des ions d'appariement tels que l'acide trifluoroacétique (TFA) ce qui va ainsi réduire la co-purification avec d'autres espèces non glycosylées⁵¹. Il faut noter également que les effets secondaires de la procédure doivent être pris en compte, puisqu'avec des concentrations élevées des solvants organiques, les protéines peuvent précipiter. Il faut donc garder à l'esprit que les protéines ont des degrés de solubilité différents selon la concentration de solvants organiques utilisée⁵².

D'autres types de chromatographies comme par l'exemple la chromatographie d'exclusion stérique⁵³, la chromatographie échangeuse d'ion ou l'utilisation des micro-colonnes ou colonne avec des phases en carbone graphite⁵⁴⁻⁵⁶ ou des phase stationnaires utilisant l'acide phénylborique immobilisé⁵⁷ ont été citées dans la littérature.

L'utilisation d'une phase par rapport à une autre va surtout dépendre du type d'échantillon, de l'étude que l'on effectue et de la question posée, en fonction donc des phases utilisées les résultats peuvent varier. C'est pourquoi dans certains cas plusieurs phases sont combinées pour avoir un maximum d'information pour la caractérisation.

4. Analyse de glycoprotéines entières

La caractérisation des glycoprotéines intactes par spectrométrie de masse pose un certain nombre de difficultés qui sont liées à leur hétérogénéité. Etant donné que la glycoprotéine est un mélange d'espèces moléculaires avec la même séquence en acides aminés mais des chaînes glycaniques différentes, l'analyse d'une glycoprotéine par spectrométrie de masse conduit à un spectre très hétérogène et l'intensité des ions est répartie entre les différentes glycoformes. Ces spectres sont d'autant plus complexes 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 ou possédant peu de glycoformes (cas des anticorps). Pour les glycoprotéines très hétérogènes, l'analyse MALDI-TOF donne un spectre très complexe et difficilement interprétable. Cependant, en optimisant le type de matrice et la préparation du dépôt, il est possible d'obtenir de meilleurs résultats⁵⁸.

Les glycoprotéines intactes peuvent être également analysés par ESI-MS⁵⁹. Cependant, contrairement à l'ionisation MALDI où les espèces observées sont majoritairement monochargées, l'ionisation ESI génère des ions multichargés. Avec ce type de source, il est donc nécessaire d'utiliser un analyseur de résolution suffisante pour séparer des glycoprotéines de masse très proche. L'ESI-MS permet d'évaluer l'hétérogénéité glycanique pour des glycoprotéines possédant un nombre limité de site glycosylés et est plus généralement utilisé pour l'analyse de protéines recombinantes (Figure 6).

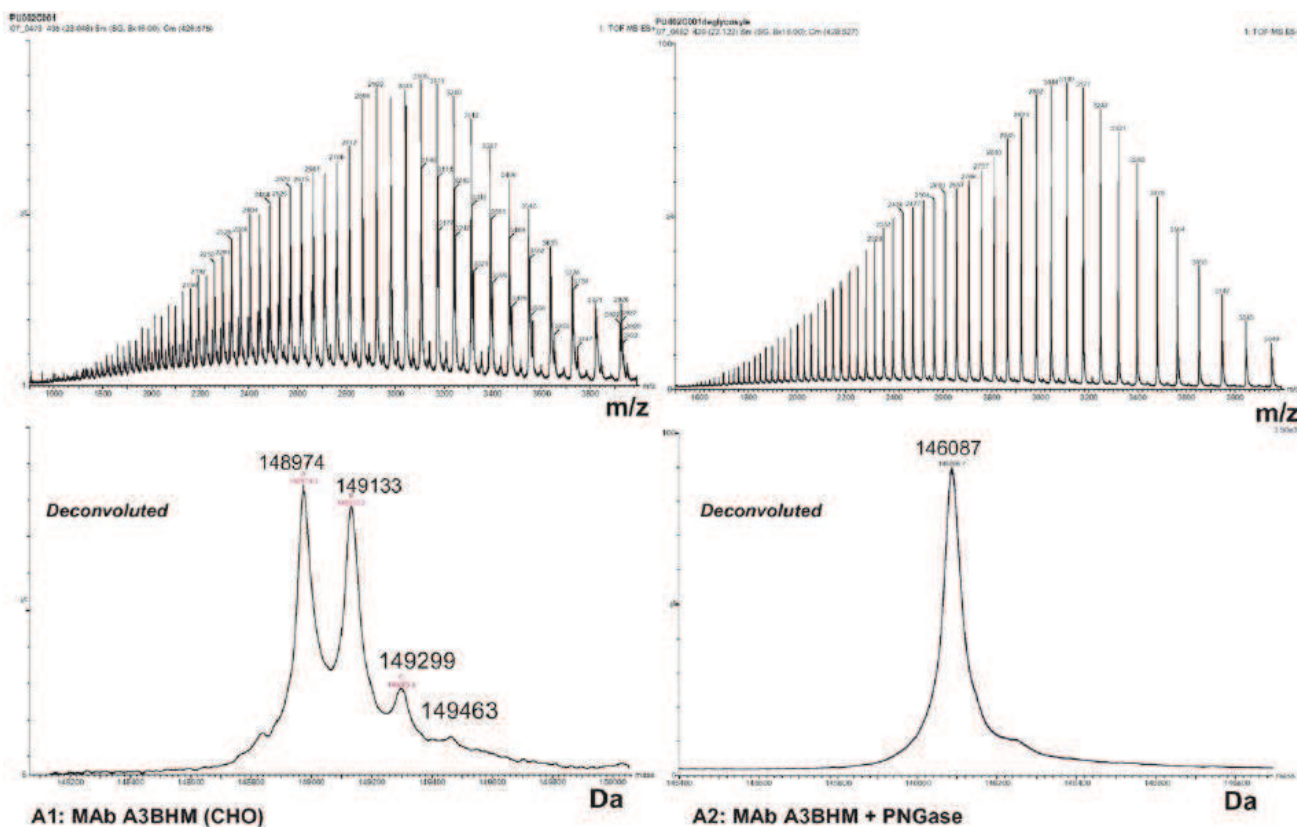


Figure 6. Spectre correspondant à l'analyse par électrospray en mode positif d'un anticorps monoclonal. A1 : protéine intacte; A2 : protéine après déglycosylation à la PNGase F. En haut les ions multichargés et en bas les spectres déconvolués qui représentent la masse moléculaire de l'anticorps et de ses différentes glycoformes ⁵⁹.

De manière générale, l'analyse des glycoprotéines intactes aboutit à très peu d'informations significatives dès lors que la glycoprotéine analysée comporte plus d'un site glycosylé. C'est pourquoi il est nécessaire d'utiliser d'autres approches qui passent par la digestion de la glycoprotéine à l'aide glycosydases ou d'endoprotéases de façon à générer des glycopeptides portant chacun un seul site de glycosylation et aussi caractériser la macro- ou micro-hétérogénéité.

5. Analyse des peptides déglycosylés et glycopeptides

a. Analyse des peptides déglycosylés

L'une des stratégies d'identification des sites de glycosylation d'une glycoprotéine consiste à décrocher partiellement ou totalement les structures oligosaccharidiques de cette glycoprotéine pour obtenir des peptides "nus" et des oligosaccharides libres qui sont plus faciles à analyser séparément en spectrométrie de masse.

La déglycosylation d'un glycopeptide ou d'une glycoprotéine peut se faire chimiquement (β -élimination en milieu alcalin, hydrazinolyse, acide trifluoro-methanesulfonique (TFMS)) ou alors à l'aide d'enzymes de déglycosylation comme la PNGase F (Peptide-N-Glycosydase F ou N-glycosydase F) et dans des cas rares avec la PNGase A (Peptide-N-glycosydase A ou N-glycosydase A) et l'endoH (Endo-Beta-N-acetylglucosaminidase H ou endoglycosydase H) ^{60, 61}.

La déglycosylation enzymatique avec la PNGase F et la PNGase A, lorsqu'elle fonctionne va libérer les oligosaccharides sur les glycopeptides ou sur les glycoprotéines et provoquer le changement de l'asparagine (Asn/N) porteur des glycanes en acide aspartique (Asp/D, $\Delta m = 1$ Da).

L'endoH quant à elle laisse uniquement le GlcNAc sur le peptide porteur de l'oligosaccharide initial (Asp/N + $\Delta m = 203$ Da) (Figure 7).

Cette déglycosylation va simplifier l'analyse par spectrométrie de masse (nanoLC-MS/MS) car elle permet d'éliminer l'hétérogénéité des glycopeptides et glycoprotéines.

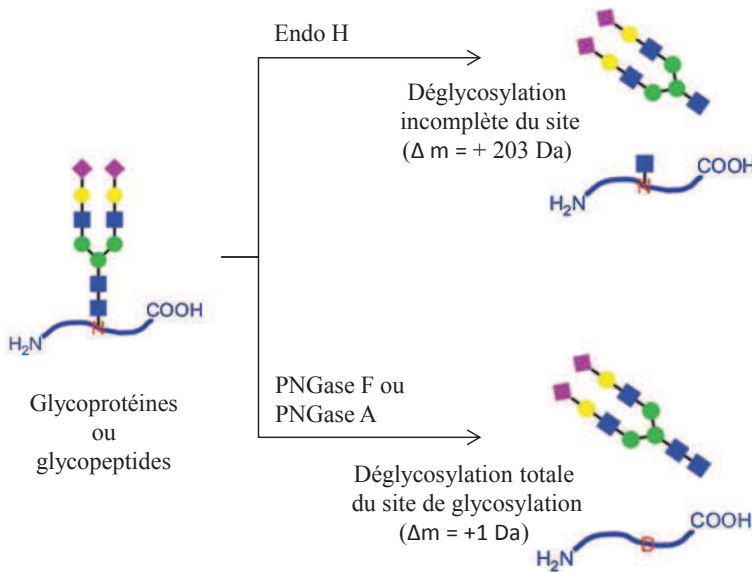


Figure 7. Représentation schématique de la déglycosylation d'un glycopeptide ou d'une glycoprotéine par une enzyme de déglycosylation (exemple de l'endoH, de la PNGase F et la PNGase A).

Les peptides nus ou peptides-GlcNAc obtenus ont des structures quasi similaires que les peptides avec des modifications variables généralement analysés par spectrométrie de masse. La position de l'Asn initialement glycosylée va être retrouvée (Figure 8) lors de l'identification à l'aide de moteurs de recherche par indication de la modification induite par le déglycosylation sur l'asparagine (par exemple la déamidation de l'Asn (Asn + 1 Da). La Figure 8 montre l'interprétation d'un spectre de fragmentation MS/MS en mode CID d'un peptide issu de la digestion trypsique d'une protéine membranaire de l'*E. Coli.* déglycosylée avec la PNGase A ⁶².

Cependant, il faut mentionner que certaines enzymes de déglycosylation sont spécifiques de certains types de polysaccharides, par exemple l'endoH et la PNGase F fonctionnent uniquement sur les polysaccharides de type hybride et high mannose et non sur les polysaccharides de type complexe où le GlcNAc proximal de la chaîne glycanique possède un fucose en position $\alpha(1,3)$ ^{63, 64}.

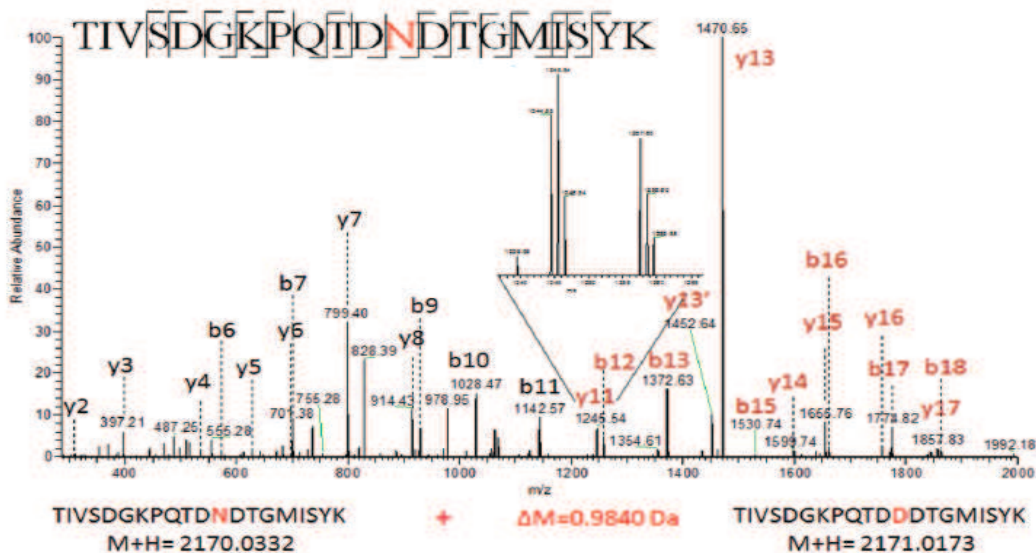


Figure 8. Spectre de fragmentation CID annoté du peptide TIVSDGKPKQTDN*DTGMISYK (M+H= 2171.01) où D* est une Asn/N déamidée du peptide TIVSDGKPKQTDN*DTGMISYK (M+H= 2170,03) soit une différence de $\Delta m \approx 1$ Da entre les deux peptides ⁶².

Le choix de l'enzyme de déglycosylation va donc être déterminant dans cette stratégie d'identification de site de glycosylation, de même que l'instrument MS utilisé puisque celui-ci doit posséder par exemple une précision et une résolution suffisantes pour pouvoir observer une modification de 1 Da sur les peptides déglycosylés. De plus, la déamidation de l'asparagine peut se faire naturellement à pH basique, et donc dans certains cas les résultats d'identification de site de glycosylation peuvent être entachés d'erreurs. C'est pourquoi il est important d'utiliser une technique d'enrichissement assez spécifique aux glycopeptides et glycoprotéines pour éviter la co-purification d'autres espèces non glycosylées. Dans certains cas, pour distinguer la déamidation naturelle et de celle induite par la déglycosylation, la digestion avec le PNGase A ou

PNGase F est faite dans de l'eau lourde ($H_2^{18}O$)⁶². La déamidation naturelle entraîne dans ce cas une différence de masse de 1 Da sur l'Asn alors que celle induite par la digestion enzymatique à l'aide de PNGase A ou PNGase F induit une différence de masse de 3 Da. Il est possible de réduire le risque de déamidation *in vitro* en optimisant les conditions de préparation des échantillons par réduction de pH de digestion trypsique et la déglycosylation⁶⁵.

Cette stratégie d'identification des sites de glycosylation par analyse de peptides déglycosylés a été utilisée dans la partie Résultats Chapitre 4 de ce travail de thèse. Cependant l'analyse de peptides déglycosylés ne donne pas l'information sur la micro-hétérogénéité et la structure des oligosaccharides.

b. Analyse MSⁿ des oligosaccharides

Les oligosaccharides obtenus de la déglycosylation par exemple avec la PNGase A ou F peuvent être analysés, déterminer l'information sur leur composition en monosaccharide. En effet, Domon et Castello ont montré dans leurs travaux⁶⁶ que, tout comme les peptides, la fragmentation des oligosaccharides en spectrométrie de masse suit une règle et des séries d'ions pouvaient être obtenues au cours de cette fragmentation. L'identification d'ions successifs d'une même série permet alors de reconstituer la composition de l'oligosaccharide fragmenté. Ils ont établi ainsi la nomenclature pour la fragmentation des sucres (Figure 9).

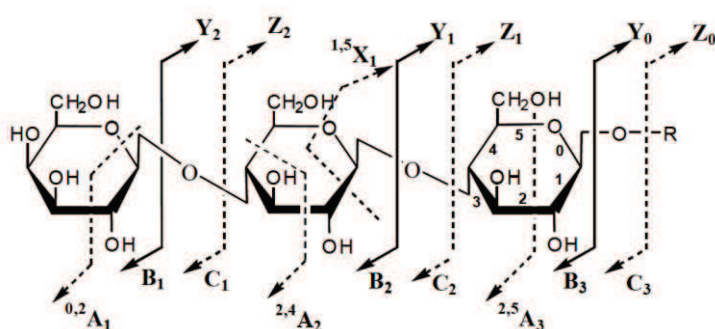


Figure 9. Nomenclature de fragmentation des polysaccharides⁶⁶. Pour les fragments de type Ai et Xi, les numéros en exposants correspondent à des fragmentations intracycliques.

Cependant la caractérisation structurale d'oligosaccharides présente des difficultés analytiques différentes de celles d'un peptide. En effet, les oligosaccharides peuvent être constitués de plusieurs monosaccharides de natures différentes et avoir plusieurs antennes. De plus, les monosaccharides de même masse peuvent correspondre à des structures différentes. La mesure précise de la masse des oligosaccharides suffit quelque fois pour déduire la composition de ceux-ci, lors qu'ils ont des structures assez simples (puisque le nombre de monosaccharides entrant dans leur composition est limité). Cependant, la MS/MS est parfois indispensable pour identifier chaque monosaccharide des oligosaccharides de taille plus volumineuse. Mais, le caractère hydrophile des oligosaccharides est responsable d'une diminution de leur efficacité d'ionisation et donc d'un manque de sensibilité lorsqu'on les analyse par spectrométrie de masse.

Pour contourner ce problème, les oligosaccharides peuvent être dérivatisés par perméthylation⁶⁷ ou analysés sous forme d'adduits de métaux alcalins comme par exemple les adduits de sodium⁶⁸ pour pouvoir être mieux détecté en spectrométrie de masse. Mais il faut préciser que la fragmentation des oligosaccharides non dérivatisés (Figure 10) est moins complexe que celle obtenue par exemple avec des adduits de sodium ou par la perméthylation (Figure 10). En effet, alors que les oligosaccharides non dérivatisés produisent principalement les ions de série Bi et Yi (Figure 10), les oligosaccharides avec des adduits de métaux alcalins vont donner des fragments supplémentaires comme les fragments intercycliques Xi et Ai (Figure 11). Ces fragments sont informatifs lorsqu'on recherche la structure détaillée des oligosaccharides, mais augmentent aussi considérablement la complexité des spectres (Figure 11) et rendent difficile l'interprétation des spectres de fragmentation lors qu'on s'intéresse uniquement à la composition des glycanes.

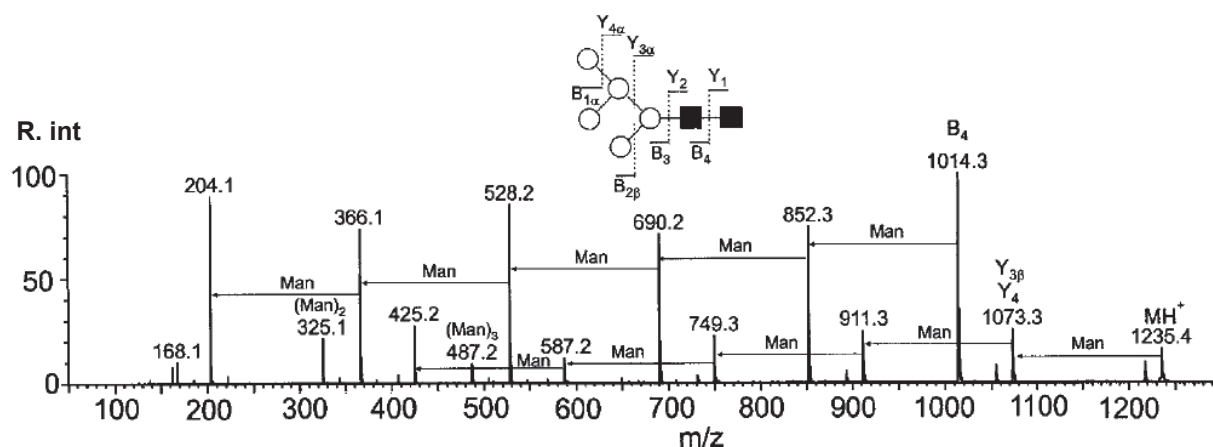


Figure 10. Spectre de fragmentation CID d'un oligosaccharide ($\text{Man}_5\text{GlcNac}_2$) ($\text{MH}^+ = 1235,4$). Les ions observés sont de la série Bi et Y_i ⁶⁸.

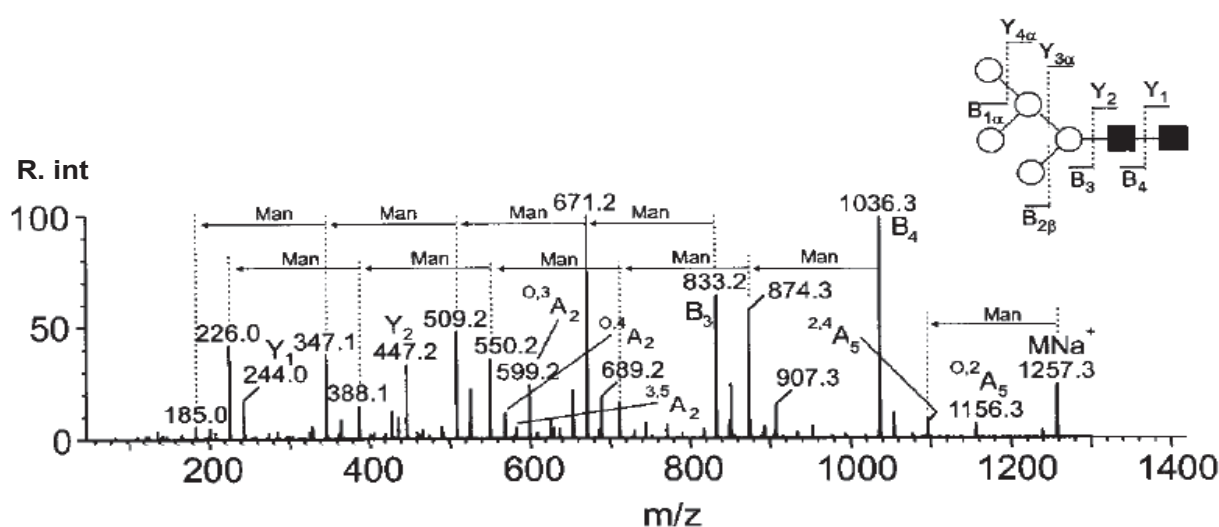


Figure 11. Spectre de fragmentation CID d'un oligosaccharide ($\text{Man}_5\text{GlcNac}_2$) ($\text{MNa}^+ = 1235,4$). Les ions observés sont de la série Bi, Y_i et les ions intracycliques (ions A_i) ⁶⁸.

En conclusion l'approche par déglycosylation permet d'identifier les sites de glycosylation sur la protéine et de déterminer la composition des oligosaccharides, mais l'information sur la micro-hétérogénéité de chaque site est perdue. Pour pallier à cet inconvénient, l'identification et la caractérisation des sites et des oligosaccharides peuvent se faire sans déglycosylation des glycoprotéines aux niveaux des glycopeptides.

c. Analyse de glycopeptides par MS

L'analyse des glycopeptides permet d'obtenir à la fois l'information sur le site de glycosylation et la composition des oligosaccharidiques portés par les glycopeptides. Dans ce cas les glycopeptides sont générés par digestion enzymatique des glycoprotéines. Différentes enzymes par exemple la trypsine, la GluC, l'AspN ou la chymotrypsine sont alors utilisées pour générer des séquences d'acides aminés plus ou moins longues et par conséquent des glycopeptides de tailles différentes pour l'analyse nanoLC-MS/MS ³⁰.

Les glycopeptides sont composés d'acides aminés et d'oligosaccharides et leurs fragmentations en MS/MS sont différentes de celles des peptides analysés classiquement en spectrométrie de masse. Deux types de fragmentation vont pouvoir être observés en théorie sur un spectre de fragmentation des glycopeptides : (1) la fragmentation de la séquence d'acides aminés ⁶⁹ et (2) la fragmentation de la chaîne de glycanes ⁶⁶. Mais en pratique, il est difficile d'obtenir les deux types de fragmentation et selon le mode de fragmentation choisie (CID ou ETD), une voie de fragmentation ou l'autre va être favorisée. De plus, les glycopeptides sont difficiles à ioniser que les peptides par la technique d'électrospray habituellement utilisée à cause de leurs propriétés physico-chimiques. En outre puisqu'un même site peut porter plusieurs glycanes, le signal en spectrométrie de masse est réparti sur de nombreuses espèces moléculaires dont l'intensité est bien plus faible que celle des peptides correspondant non glycosylés.

En réponse à ce problème de détection des glycopeptides, une approche a été développée ⁷⁰. Cette approche permet de détecter les glycopeptides dans les spectres de masse à l'aide d'ions spécifiques de la fragmentation des glycosylations : "les ions diagnostiques" (Figure 12).

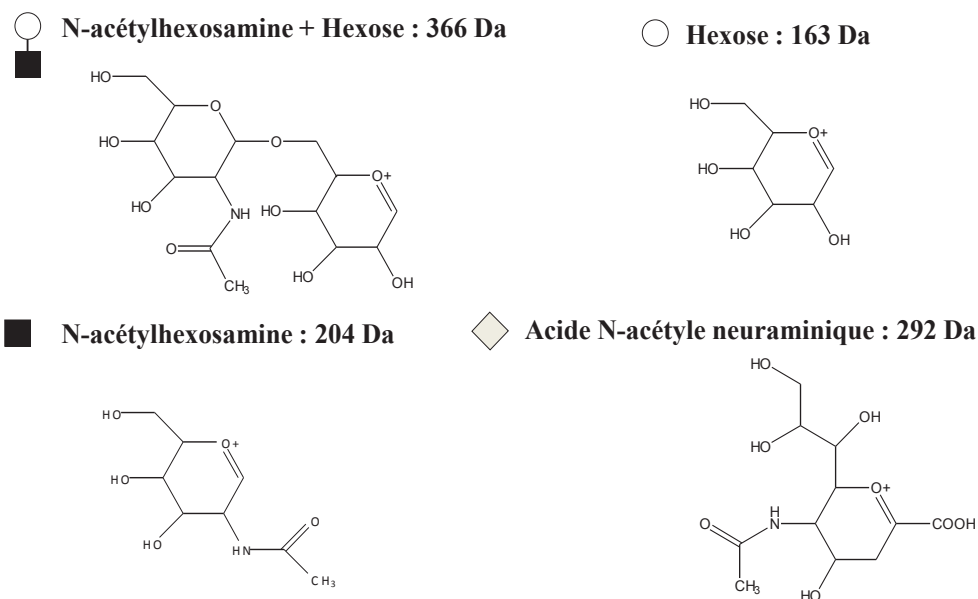


Figure 12. Structures chimiques des ions oxoniums servant d'ions diagnostiques

Les ions diagnostiques peuvent être générés de deux manières: soit par fragmentation dans la cellule de collision ⁷⁰, soit par fragmentation en source ^{70, 71}. Ces ions sont issus de la fragmentation interne des glycanes et ont donc des masses moléculaires différentes en fonction du type de glycosylation. Par exemple, dans le cas des glycanes de type high mannoses, on observe l'ion m/z 366 (ManGlcNAc^+) dans des proportions moins importantes que dans le cas de glycopeptides complexes ⁷⁰. En effet, comme le montre la Figure 13, la génération de cet ion nécessite la coupure de 3 liaisons dans le cas de glycanes de type high mannose et seulement d'une ou deux liaisons dans le cas des glycanes complexes.

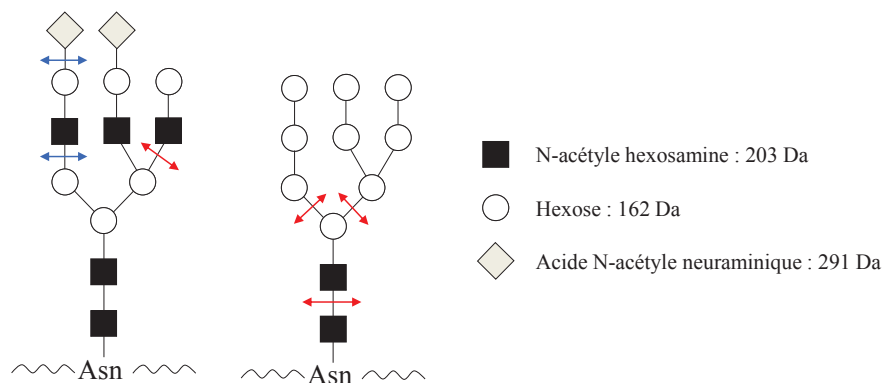


Figure 13. Liaisons devant être fragmentées pour la génération de l'ion m/z 360.

De façon générale, on distingue sur les spectres de fragmentation de N-glycopeptides, 3 types d'ions fragments :

- des ions fragments correspondant à des pertes de sucres successifs sur la chaîne peptidique : ces ions permettent de déterminer l'enchaînement des monosaccharides et ainsi de remonter à la composition de chaque oligosaccharide;
- des ions oxoniums de type Bi et Yi qui sont issus de la fragmentation interne des glycanes. Ces ions donnent des informations supplémentaires sur la composition des glycanes mais surtout ils sont caractéristiques de celui-ci et donc peuvent servir pour identifier rapidement si un spectre MS/MS correspond à la fragmentation d'un glycopeptide ou non;
- des ions correspondant à la fragmentation du peptide non glycosylé.

Par la fragmentation CID, le mode fragmentation le plus courant, les liaisons osidiques des glycanes nettement plus labiles que les liaisons peptidiques se fragmentent préférentiellement au détriment de la séquence d'acides aminés (Figure 15, A). Les spectres obtenus sont en général relativement plus simples à interpréter puisque les ions correspondant à la fragmentation des oligosaccharides sont les plus intenses et les rares fragments des peptides sont moins intenses. Pour obtenir plus d'information sur les peptides, des ions fragments comme par exemple l'ion du peptide portant un GlcNac peut être sélectionné et fragmenté en MS³ (Figure 15, B) dans une trappe ionique (IT), ou alors l'énergie de collision être optimisée pour augmenter la fragmentation de la chaîne peptidique⁷².

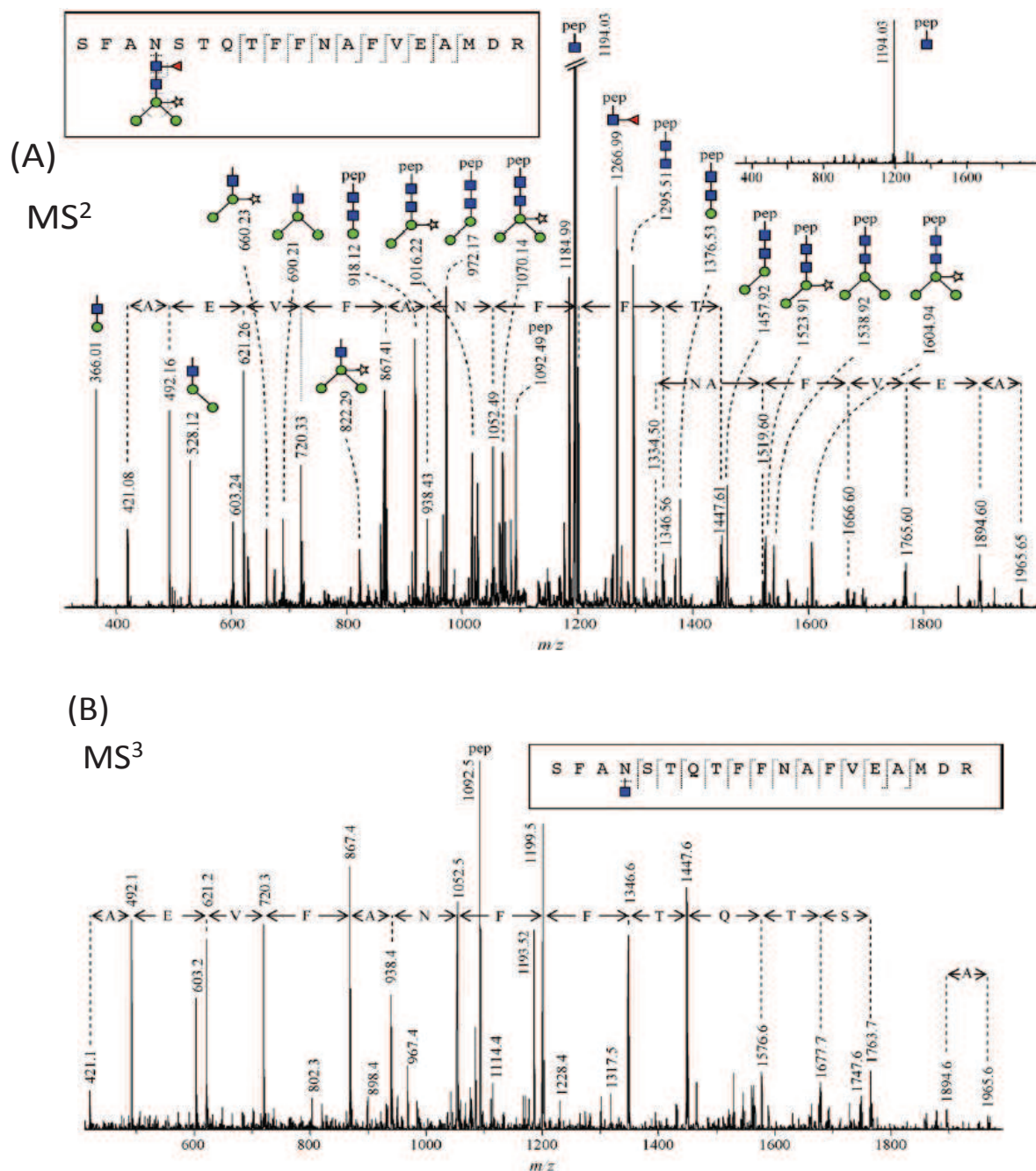


Figure 14. Exemple de spectres de fragmentation CID de glycopeptides obtenus sur la trappe ionique (HCT ultra, Bruker Daltonics): (A) spectre MS/MS de l'ion trichargé de $m/z = 1119$ correspondant au glycopeptide tryptique Ser²⁹⁵- Arg³¹³ de la peroxidase de raifort (horseradish peroxidase HRP); (B) l'ion fragment monochargé $m/z = 1194,03$ a été sélectionné isolé et fragmenté en MS³ pour avoir plus d'information sur la séquences d'acides aminés⁷².

Comme nous venons de le voir, la fragmentation CID, conduit majoritairement à la fragmentation du glycanes et un faible nombre de fragments de la séquence d'acides aminés. Ces fragments ne sont en général

pas assez nombreux pour permettre de reconstituer la séquence du peptide. En complément au mode de fragmentation CID d'autres modes de fragmentation comme l'ETD peuvent être utilisés⁷³.

En mode de fragmentation ETD, la fragmentation de glycopeptides est statistique, c'est-à-dire que l'ETD fragmente aussi bien le peptide que l'oligosaccharide. On obtient ainsi des informations sur le peptide et l'oligosaccharide. Cependant, les spectres ETD des glycopeptides sont assez complexes à interpréter surtout si la séquence du glycopeptide n'est pas connue. Il est donc difficile avec ce mode de fragmentation d'obtenir l'information sur la composition de l'oligosaccharide^{74, 75}, mais néanmoins l'Asn portant les glycanes peut être identifiée (Figure 15).

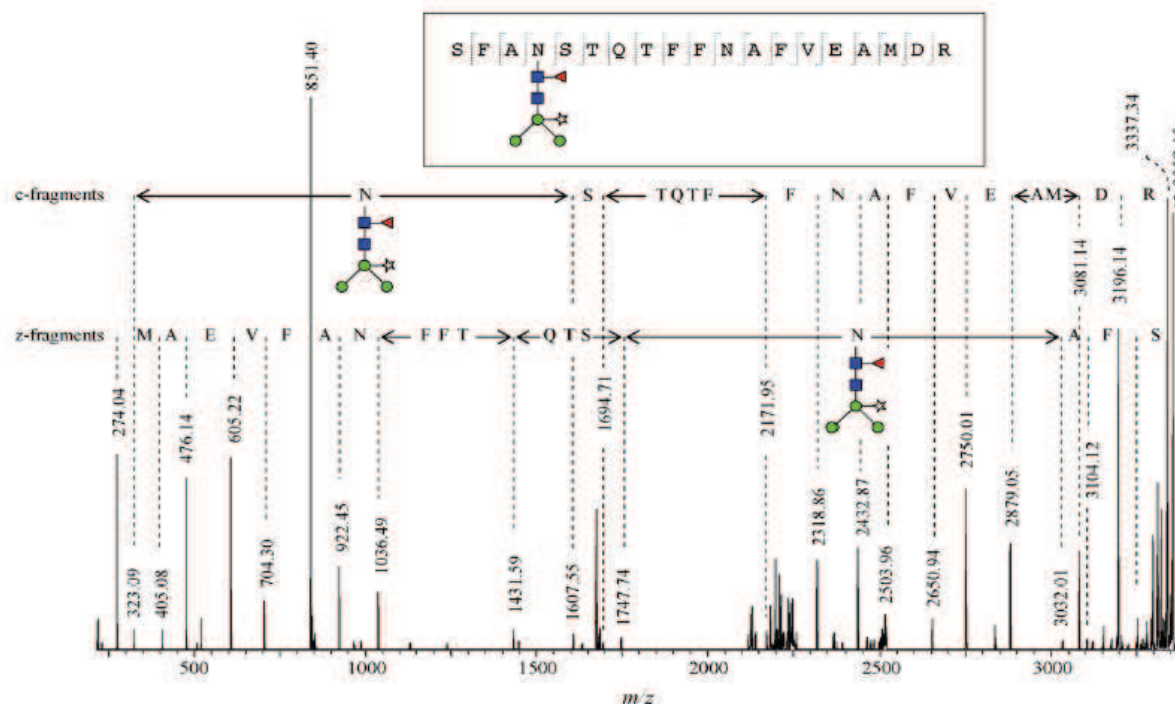


Figure 15. Exemple de spectres de fragmentation ETD de glycopeptides obtenus sur la trappe ionique (HCT ultra, Bruker Daltonics): spectre MS/MS de l'ion trichargé de $m/z = 1119$ correspondant au glycopeptide tryptique Ser295-Arg313 de la peroxidase de raifort (horseradish peroxidase HRP)⁷².

Le mode de fragmentation ETD et CID peuvent dans certains cas être combiné pour remonter à la structure des glycanes et à la séquence des peptides^{74, 76}. Cependant, toute la séquence du peptide ne se fragmente pas et les spectres de fragmentation ETD sont très complexes et l'interprétation doit souvent être faite manuellement.

III. Les outils d'interprétation des données LC-MS/MS des glycopeptides et oligosaccharides

1. Interprétation des données MS et MS/MS des oligosaccharides.

Il existe plusieurs outils d'aide pour l'interprétation des données MS d'oligosaccharides comme par exemple GlycanMass (<http://web.expasy.org/glycanmass/>) qui permettent de calculer la masse théorique des oligosaccharides dérivatisés ou non. Il existe également des outils qui permettent de proposer des structures d'oligosaccharidiques à partir d'une liste de masses (MS) comme c'est le cas pour exemple GlycoMod (<http://web.expasy.org/glycomod/>)^{77, 78}. Cependant l'utilisation de ces outils reste conditionnée par la connaissance préalable du type de sucres contenus dans les oligosaccharides étudiés. Les structures d'oligosaccharides proposées ne sont parfois pas réalistes. En effet les structures d'oligosaccharides qui sont parfois proposées ne sont que des combinaisons linéaires de sucres alors que les oligosaccharides des N-glycoprotéines sont de structures ramifiées. Ce qui a pour conséquence d'augmenter considérablement le nombre des oligosaccharides candidats à examiner. L'identification de structure d'oligosaccharides par simple mesure de masse (MS) va être limitée puisque plusieurs oligosaccharides vont avoir la même masse (oligosaccharides isobariques). Cette recherche peut se faire également dans des banques d'oligosaccharides comme GlycosuiteDB (<http://glycosuitedb.expasy.org>)⁷⁹, GlycomeDB (<http://www.glycome-db.org/>)^{80, 81} ou autres banques pour limiter la recherche à des cas de structure d'oligosaccharides déjà connues. Cependant, cela peut constituer un frein à la découverte de nouvelles structures d'oligosaccharides ou glycoformes chez une espèce ou protéine particulière.

La MS/MS est une solution pour identifier et caractériser la structure d'oligosaccharides, cependant la fragmentation de l'oligosaccharide est dépendante de sa composition en monosaccharides et de son état d'ionisation (adduits de sodium, potassium)⁶⁸. Il existe plusieurs outils d'interprétation de spectres MS/MS des oligosaccharides comme Glycoworkbench en accès libre⁸² ou commercialisés comme SysBioWare⁸³ ou SymGylcan (<http://www.premierbiosoft.com/>)⁸⁴ qui permet d'interpréter automatiquement les spectres MS/MS après indication préalable de glycanes composant les oligosaccharides peptides fragmentés. L'interprétation des données peut donc être facilement biaisée lorsqu'on n'a pas de connaissance suffisante sur les types de glycanes qui sont présents sur les glycoprotéines.

2. Interprétation des données MS/MS de glycopeptides

Bien que plusieurs outils existent pour l'identification et la caractérisation des glycopeptides à partir des données MS/MS, l'automatisation de l'interprétation *de novo* des spectres reste difficile. En effet n'existe pas à ce jour une banque de glycopeptides à même de fournir des données de comparaisons avec les données expérimentales de MS/MS. De plus, les tailles des glycopeptides dans des digestes sont très variées et dépendantes des endoprotéases utilisés et des différentes glycoformes présentes sur chaque site de glycosylation. Mais la véritable difficulté de l'automatisation d'identification et de caractérisation de glycopeptides reste surtout liée à la grande différence de fragmentation de glycopeptides qui sont très dépendants des spectromètres de masse utilisés et du mode de fragmentation MS/MS utilisé. En effet, étant donné que très peu de fragments de peptides sont observés dans les spectres, il est difficile d'établir la séquence en acides aminés de glycopeptides à partir des spectres MS/MS.

Pour ces multiples raisons, beaucoup de logiciel développés à ce jour nécessitent de connaître la séquence en acides aminés des glycoprotéines⁸⁴⁻⁸⁹. Et pour limiter le nombre de faux positifs, l'utilisation de spectromètres de masse de très hautes résolution de type FTICR⁸⁹ est parfois indispensable. Des problèmes liés à la compatibilité avec les formats des listes de masse se posent également dans certains cas, puisque certains logiciels sont développés uniquement pour un constructeur de spectromètres de masse. Il faut noter également que ces logiciels sont souvent développés avec un nombre limité d'endoprotéases classiquement utilisées en protéomique comme la trypsine, l'Asp N, la Glu C et la chymotrypsine, alors que d'autres endoprotéases comme la pronase, protéinase K, pepsine, papaïne, et la subtilisine ont été utilisés avec succès pour la caractérisation de glycopeptides^{30, 72}. Ces outils imposent ainsi une stratégie d'analyse qui n'est pas forcément la mieux adaptée dans un contexte ou un autre. Enfin, certains programmes, semblent prometteurs mais ne sont

utilisables que sur de très petites quantités de données⁵⁷. Ainsi, il n'a y a pas à ce jour, de logiciel permettant d'effectuer une interprétation de spectres MS/MS de glycopeptides présents dans un mélange inconnue de manière indépendante.

Conclusion

La glycosylation est l'une des modifications post-traductionnelles les plus courantes et possède un rôle essentiel dans la fonction biologique des protéines. Cependant, la caractérisation de glycoprotéines reste très complexe en partie à cause de leur hétérogénéité et le manque de matériel standard de référence⁹⁰.

Pour caractériser aussi finement que possible les protéines N-glycosylées, un certain nombre de critères sont à prendre alors en considération comme par exemple, la qualité et la quantité du matériel, la taille de la glycoprotéine, son hétérogénéité, la complexité de l'échantillon et l'information précisément recherchée sur les glycoprotéines. Pour parvenir à cette caractérisation des protéines N-glycosylées, il est donc nécessaire de mettre en place des méthodologies combinant plusieurs techniques et approches analytiques aussi bien au niveau de la préparation d'échantillon, de l'acquisition des données MS et MS/MS, que l'interprétation des résultats. Ces techniques et approches analytiques vont apporter des éléments de caractérisation complémentaires nécessaire pour la caractérisation. Chacune des étapes demande une grande expertise et doit être adaptée au contexte d'étude.

Deux grandes stratégies sont utilisées pour la caractérisation des glycoprotéines. La première stratégie consiste à déglycosyler les glycopeptides obtenus après digestion de la glycoprotéine par une endoprotéase et d'analyser les glycanes libres et les peptides séparément. Les récents progrès réalisés au niveau de la sensibilité et de la résolution de nouvelle génération de spectromètres de masse ont favorisé le développement d'une deuxième stratégie qui consiste à analyser les glycopeptides intacts. Cette stratégie permet ainsi de travailler avec de faible quantité de matériel et de déterminer à la fois la composition de l'oligosaccharide, le site de glycosylation, mais également la micro-hétérogénéité ce qui n'est pas le cas de la première stratégie. Ces deux stratégies ont été combinées dans la Partie Résultats, Chapitre 4, pour tirer profit des avantages de chacune des deux stratégies.

Si l'interprétation de spectres MS/MS de peptides non glycosylés est aujourd'hui entièrement automatisée, il n'en est pas de même pour les spectres de glycopeptides, puisque les quelques logiciels existant à ce jour fonctionnent de manière semi-automatique avec une partie des étapes entièrement manuelle. De plus ces logiciels fonctionnent sur la base d'une connaissance préalable du type de glycosylation, de la glycoprotéine et parfois imposent un mode opératoire bien déterminé (un spectromètre de masse particulier ou une enzyme particulière à utilisée). En outre, les quantités de données prises en charge par ces logiciels sont parfois limitées.

Pour éviter toutes ces contraintes un grand nombre de spectres MS/MS des glycopeptides sont encore entièrement interprétés manuellement (cf. Partie Résultats, Chapitre 4). Et pour faciliter la localisation des spectres MS/MS de glycopeptides on utilise alors les ions diagnostiques formés à partir de la fragmentation partielle des oligosaccharides de glycopeptides (cf. Partie Résultats, Chapitre 4).

Aujourd'hui face à des données de spectrométrie de masse de plus en plus volumineuses et qui portent sur des échantillons pour la plupart très complexes ou sur des espèces parfois très hétérogènes, il apparaît indispensable de développer des méthodologies adaptées à chaque contexte mais aussi des logiciels plus élaborés et universels afin de faciliter l'interprétation de données MS/MS de glycopeptides.

Les projets communs comme MIRAGE lancé en novembre 2011⁹¹, devraient permettre la centralisation des informations de qualité de glycomique comme les données expérimentales LC MS/MS et données biologiques du monde entier et la construction d'une base de données communes à toute la communauté de glycomique à l'instar de Swissprot. De nouvelles directives sur les données MS et MS/MS des glycopeptides et oligosaccharides devraient conduire dans les années à venir au développement des algorithmes d'interprétation des spectres des MS/MS capable de fonctionner de manière autonome et admis par toute la communauté de glycomique.

Références bibliographiques du chapitre 2

1. Bradshaw, R. A.; Medzihradzky, K. F.; Chalkley, R. J. Protein PTMs: post-translational modifications or pesky trouble makers? *Journal of mass spectrometry* **2010**, 45, (10), 1095-7.
2. Mann, M.; Jensen, O. N. Proteomic analysis of post-translational modifications. *Nature biotechnology* **2003**, 21, (3), 255-61.
3. Doerr, A. Making PTMs a priority. *Nature methods* **2012**, 9, (9), 862-3.
4. Zhao, J.; Patwa, T. H.; Pal, M.; Qiu, W.; Lubman, D. M. Analysis of protein glycosylation and phosphorylation using liquid phase separation, protein microarray technology, and mass spectrometry. *Methods in molecular biology* **2009**, 492, 321-51.
5. Nielsen, M. L.; Savitski, M. M.; Zubarev, R. A. Extent of modifications in human proteome samples and their effect on dynamic range of analysis in shotgun proteomics. *Molecular & cellular proteomics* **2006**, 5, (12), 2384-91.
6. Aitken, A. Protein consensus sequence motifs. *Molecular biotechnology* **1999**, 12, (3), 241-53.
7. Davey, N. E.; Edwards, R. J.; Shields, D. C. Computational identification and analysis of protein short linear motifs. *Frontiers in bioscience : a journal and virtual library* **2010**, 15, 801-25.
8. Plewczynski, D.; Tkacz, A.; Wyrwicz, L. S.; Rychlewski, L. AutoMotif server: prediction of single residue post-translational modifications in proteins. *Bioinformatics* **2005**, 21, (10), 2525-7.
9. Rini, J.; Esko, J.; Varki, A. Glycosyltransferases and Glycan-processing Enzymes. *Cold Spring Harbor (NY)*, 2009.
10. Varki, A.; Cummings, R. D.; Esko, J. D.; Freeze, H. H.; Stanley, P.; Marth, J. D.; Bertozzi, C. R.; Hart, G. W.; Etzler, M. E. Symbol nomenclature for glycan representation. *Proteomics* **2009**, 9, (24), 5398-9.
11. Khoury, G. A.; Baliban, R. C.; Floudas, C. A. Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. *Scientific reports* **2011**, 1.
12. Helenius, A.; Aebi, M. Roles of N-linked glycans in the endoplasmic reticulum. *Annual review of biochemistry* **2004**, 73, 1019-49.
13. Vigerust, D. J.; Shepherd, V. L. Virus glycosylation: role in virulence and immune interactions. *Trends in microbiology* **2007**, 15, (5), 211-8.
14. Mitra, N.; Sinha, S.; Ramya, T. N.; Surolia, A. N-linked oligosaccharides as outfitters for glycoprotein folding, form and function. *Trends in biochemical sciences* **2006**, 31, (3), 156-63.
15. Rudd, P. M.; Elliott, T.; Cresswell, P.; Wilson, I. A.; Dwek, R. A. Glycosylation and the immune system. *Science* **2001**, 291, (5512), 2370-6.
16. Varki, A. Essentials of glycobiology. *Cold Spring Harbor, (NY)*, 2009.
17. Varki, A.; Kannagi, R.; Toole, B. P. Glycosylation Changes in Cancer. *Cold Spring Harbor (NY)*, 2009.
18. Varki, A.; Freeze, H. H. Glycans in Acquired Human Diseases. *Cold Spring Harbor (NY)*, 2009.
19. Rosnoblet, C.; Peanne, R.; Legrand, D.; Foulquier, F. Glycosylation disorders of membrane trafficking. *Glycoconjugate journal* **2013**, 30, (1), 23-31.
20. Dube, D. H.; Bertozzi, C. R. Glycans in cancer and inflammation--potential for therapeutics and diagnostics. *Nature reviews. Drug discovery* **2005**, 4, (6), 477-88.
21. Blomme, B.; Van Steenkiste, C.; Callewaert, N.; Van Vlierberghe, H. Alteration of protein glycosylation in liver diseases. *Journal of hepatology* **2009**, 50, (3), 592-603.
22. Bertozzi, C. R.; Freeze, H. H.; Varki, A.; Esko, J. D. Glycans in Biotechnology and the Pharmaceutical Industry. *Cold Spring Harbor (NY)*, 2009.
23. Hanisch, F. G. O-glycosylation of the mucin type. *Biological chemistry* **2001**, 382, (2), 143-9.

24. Flynn, W. G. *Biotechnology and bioengineering. Nova Science Publishers (NK)*, 2008.
25. Satomi, Y.; Shimonishi, Y.; Takao, T., N-glycosylation at Asn(491) in the Asn-Xaa-Cys motif of human transferrin. *FEBS letters* **2004**, 576, (1-2), 51-6.
26. Helenius, A.; Aebi, M. Intracellular functions of N-linked glycans. *Science* **2001**, 291, (5512), 2364-9.
27. Dalpathado, D. S.; Desaire, H. Glycopeptide analysis by mass spectrometry. *The Analyst* **2008**, 133, (6), 731-8.
28. Gomord, V.; Faye, L. Posttranslational modification of therapeutic proteins in plants. *Current opinion in plant biology* **2004**, 7, (2), 171-81.
29. Morelle, W.; Canis, K.; Chirat, F.; Faid, V.; Michalski, J. C. The use of mass spectrometry for the proteomic analysis of glycosylation. *Proteomics* **2006**, 6, (14), 3993-4015.
30. Nwosu, C. C.; Huang, J.; Aldredge, D. L.; Strum, J. S.; Hua, S.; Seipert, R. R.; Lebrilla, C. B. In-gel nonspecific proteolysis for elucidating glycoproteins: a method for targeted protein-specific glycosylation analysis in complex protein mixtures. *Analytical chemistry* **2013**, 85, (2), 956-63.
31. Rudd, P. M.; Colominas, C.; Royle, L.; Murphy, N.; Hart, E.; Merry, A. H.; Hebestreit, H. F.; Dwek, R. A. A high-performance liquid chromatography based strategy for rapid, sensitive sequencing of N-linked oligosaccharide modifications to proteins in sodium dodecyl sulphate polyacrylamide electrophoresis gel bands. *Proteomics* **2001**, 1, (2), 285-94.
32. Pasing, Y.; Sickmann, A.; Lewandrowski, U. N-glycoproteomics: mass spectrometry-based glycosylation site annotation. *Biological chemistry* **2012**, 393, (4), 249-58.
33. Stahl-Zeng, J.; Lange, V.; Ossola, R.; Eckhardt, K.; Krek, W.; Aebersold, R.; Domon, B. High sensitivity detection of plasma proteins by multiple reaction monitoring of N-glycosites. *Molecular & cellular proteomics : MCP* **2007**, 6, (10), 1809-17.
34. Lewandrowski, U.; Zahedi, R. P.; Moebius, J.; Walter, U.; Sickmann, A. Enhanced N-glycosylation site analysis of sialoglycopeptides by strong cation exchange prefractionation applied to platelet plasma membranes. *Molecular & cellular proteomics : MCP* **2007**, 6, (11), 1933-41.
35. Baenziger, J. U.; Fiete, D. Structural determinants of concanavalin A specificity for oligosaccharides. *The Journal of biological chemistry* **1979**, 254, (7), 2400-7.
36. Debray, H.; Decout, D.; Strecker, G.; Spik, G.; Montreuil, J. Specificity of twelve lectins towards oligosaccharides and glycopeptides related to N-glycosylproteins. *European journal of biochemistry* **1981**, 117, (1), 41-55.
37. Kronis, K. A.; Carver, J. P. Specificity of isolectins of wheat germ agglutinin for sialyloligosaccharides: a 360-MHz proton nuclear magnetic resonance binding study. *Biochemistry* **1982**, 21, (13), 3050-7.
38. Zielinska, D. F.; Gnad, F.; Wisniewski, J. R.; Mann, M. Precision mapping of an in vivo N-glycoproteome reveals rigid topological and sequence constraints. *Cell* **2010**, 141, (5), 897-907.
39. Lee, A.; Nakano, M.; Hincapie, M.; Kolarich, D.; Baker, M. S.; Hancock, W. S.; Packer, N. H. The lectin riddle: glycoproteins fractionated from complex mixtures have similar glycomic profiles. *Omics : a journal of integrative biology* **2010**, 14, (4), 487-99.
40. Wollscheid, B.; Bausch-Fluck, D.; Henderson, C.; O'Brien, R.; Bibel, M.; Schiess, R.; Aebersold, R.; Watts, J. D. Mass-spectrometric identification and relative quantification of N-linked cell surface glycoproteins. *Nature biotechnology* **2009**, 27, (4), 378-86.
41. Zeng, X.; Hood, B. L.; Sun, M.; Conrads, T. P.; Day, R. S.; Weissfeld, J. L.; Siegfried, J. M.; Bigbee, W. L. Lung cancer serum biomarker discovery using glycoprotein capture and liquid chromatography mass spectrometry. *Journal of proteome research* **2010**, 9, (12), 6440-9.
42. Zhang, H.; Li, X. J.; Martin, D. B.; Aebersold, R. Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nature biotechnology* **2003**, 21, (6), 660-6.

43. Chelius, D.; Shaler, T. A. Capture of peptides with N-terminal serine and threonine: a sequence-specific chemical method for Peptide mixture simplification. *Bioconjugate chemistry* **2003**, 14, (1), 205-11.
44. Palmisano, G.; Lendal, S. E.; Larsen, M. R. Titanium dioxide enrichment of sialic acid-containing glycopeptides. *Methods in molecular biology* **2011**, 753, 309-22.
45. Engholm-Keller, K.; Larsen, M. R. Titanium dioxide as chemo-affinity chromatographic sorbent of biomolecular compounds--applications in acidic modification-specific proteomics. *Journal of proteomics* **2011**, 75, (2), 317-28.
46. Larsen, M. R.; Thingholm, T. E.; Jensen, O. N.; Roepstorff, P.; Jorgensen, T. J. Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. *Molecular & cellular proteomics : MCP* **2005**, 4, (7), 873-86.
47. Palmisano, G.; Antonacci, D.; Larsen, M. R. Glycoproteomic profile in wine: a 'sweet' molecular renaissance. *J Proteome Res* **2010**, 9, (12), 6148-59.
48. Hagglund, P.; Bunkenborg, J.; Elortza, F.; Jensen, O. N.; Roepstorff, P. A new strategy for identification of N-glycosylated proteins and unambiguous assignment of their glycosylation sites using HILIC enrichment and partial deglycosylation. *Journal of proteome research* **2004**, 3, (3), 556-66.
49. Parker, B. L.; Palmisano, G.; Edwards, A. V.; White, M. Y.; Engholm-Keller, K.; Lee, A.; Scott, N. E.; Kolarich, D.; Hambly, B. D.; Packer, N. H.; Larsen, M. R.; Cordwell, S. J. Quantitative N-linked glycoproteomics of myocardial ischemia and reperfusion injury reveals early remodeling in the extracellular environment. *Molecular & cellular proteomics* **2011**, 10, (8), M110 006833.
50. Ding, W.; Nothaft, H.; Szymanski, C. M.; Kelly, J. Identification and quantification of glycoproteins using ion-pairing normal-phase liquid chromatography and mass spectrometry. *Molecular & cellular proteomics* **2009**, 8, (9), 2170-85.
51. Mysling, S.; Palmisano, G.; Hojrup, P.; Thaysen-Andersen, M. Utilizing ion-pairing hydrophilic interaction chromatography solid phase extraction for efficient glycopeptide enrichment in glycoproteomics. *Analytical chemistry* **2010**, 82, (13), 5598-609.
52. Kay, R.; Barton, C.; Ratcliffe, L.; Matharoo-Ball, B.; Brown, P.; Roberts, J.; Teale, P.; Creaser, C. Enrichment of low molecular weight serum proteins using acetonitrile precipitation for mass spectrometry based proteomic analysis. *Rapid communications in mass spectrometry* **2008**, 22, (20), 3255-60.
53. Alvarez-Manilla, G.; Atwood, J., 3rd; Guo, Y.; Warren, N. L.; Orlando, R.; Pierce, M. Tools for glycoproteomic analysis: size exclusion chromatography facilitates identification of tryptic glycopeptides with N-linked glycosylation sites. *Journal of proteome research* **2006**, 5, (3), 701-8.
54. Larsen, M. R.; Hojrup, P.; Roepstorff, P. Characterization of gel-separated glycoproteins using two-step proteolytic digestion combined with sequential microcolumns and mass spectrometry. *Molecular & cellular proteomics* **2005**, 4, (2), 107-19.
55. Ruhaak, L. R.; Deelder, A. M.; Wuhrer, M. Oligosaccharide analysis by graphitized carbon liquid chromatography-mass spectrometry. *Analytical and bioanalytical chemistry* **2009**, 394, (1), 163-74.
56. Alley, W. R., Jr.; Mechref, Y.; Novotny, M. V. Use of activated graphitized carbon chips for liquid chromatography/mass spectrometric and tandem mass spectrometric analysis of tryptic glycopeptides. *Rapid communications in mass spectrometry* **2009**, 23, (4), 495-505.
57. Ozohanics, O.; Turiak, L.; Drahos, L.; Vekey, K. Comparison of glycopeptide/glycoprotein enrichment techniques. *Rapid communications in mass spectrometry* **2012**, 26, (2), 215-7.
58. Gimenez, E.; Benavente, F.; Barbosa, J.; Sanz-Nebot, V. Towards a reliable molecular mass determination of intact glycoproteins by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid communications in mass spectrometry* **2007**, 21, (16), 2555-63.
59. Wagner-Rousset, E.; Bednarczyk, A.; Bussat, M. C.; Colas, O.; Corvaia, N.; Schaeffer, C.; Van Dorsselaer, A.; Beck, A. The way forward, enhanced characterization of therapeutic antibody glycosylation: comparison of three level mass spectrometry-based strategies. *Journal of chromatography. B* **2008**, 872, (1-2), 23-37.

60. Fryksdale, B. G.; Jedrzejewski, P. T.; Wong, D. L.; Gaertner, A. L.; Miller, B. S. Impact of deglycosylation methods on two-dimensional gel electrophoresis and matrix assisted laser desorption/ionization-time of flight-mass spectrometry for proteomic analysis. *Electrophoresis* **2002**, 23, (14), 2184-93.
61. Morelle, W.; Michalski, J. C. Analysis of protein glycosylation by mass spectrometry. *Nature protocols* **2007**, 2, (7), 1585-602.
62. Palmisano, G.; Melo-Braga, M. N.; Engholm-Keller, K.; Parker, B. L.; Larsen, M. R. Chemical deamidation: a common pitfall in large-scale N-linked glycoproteomic mass spectrometry-based analyses. *Journal of proteome research* **2012**, 11, (3), 1949-57.
63. Tretter, V.; Altmann, F.; Marz, L. Peptide-N4-(N-acetyl-beta-glucosaminyl)asparagine amidase F cannot release glycans with fucose attached alpha 1---3 to the asparagine-linked N-acetylglucosamine residue. *European journal of biochemistry* **1991**, 199, (3), 647-52.
64. Bohle, L. A.; Mathiesen, G.; Vaaje-Kolstad, G.; Eijnsink, V. G. An endo-beta-N-acetylglucosaminidase from *Enterococcus faecalis* V583 responsible for the hydrolysis of high-mannose and hybrid-type N-linked glycans. *FEMS microbiology letters* **2011**, 325, (2), 123-9.
65. Hao, P.; Ren, Y.; Alpert, A. J.; Sze, S. K. Detection, evaluation and minimization of nonenzymatic deamidation in proteomic sample preparation. *Molecular & cellular proteomics* **2011**, 10, (10), O111 009381.
66. Domon, B.; Costello, C. E. Structure elucidation of glycosphingolipids and gangliosides using high-performance tandem mass spectrometry. *Biochemistry* **1988**, 27, (5), 1534-43.
67. Ritamo, I.; Rabina, J.; Natunen, S.; Valmu, L. Nanoscale reversed-phase liquid chromatography-mass spectrometry of permethylated N-glycans. *Analytical and bioanalytical chemistry* **2013**, 405, (8), 2469-80.
68. Harvey, D. J. Collision-induced fragmentation of underivatized N-linked carbohydrates ionized by electrospray. *Journal of mass spectrometry* **2000**, 35, (10), 1178-90.
69. Biemann, K. Sequencing of peptides by tandem mass spectrometry and high-energy collision-induced dissociation. *Methods in enzymology* **1990**, 193, 455-79.
70. Huddleston, M. J.; Bean, M. F.; Carr, S. A. Collisional fragmentation of glycopeptides by electrospray ionization LC/MS and LC/MS/MS: methods for selective detection of glycopeptides in protein digests. *Analytical chemistry* **1993**, 65, (7), 877-84.
71. Peterman, S. M.; Mulholland, J. J. A novel approach for identification and characterization of glycoproteins using a hybrid linear ion trap/FT-ICR mass spectrometer. *Journal of the American Society for Mass Spectrometry* **2006**, 17, (2), 168-79.
72. Wuhler, M.; Catalina, M. I.; Deelder, A. M.; Hokke, C. H. Glycoproteomics based on tandem mass spectrometry of glycopeptides. *Journal of chromatography. B* **2007**, 849, (1-2), 115-28.
73. Scott, N. E.; Parker, B. L.; Connolly, A. M.; Paulech, J.; Edwards, A. V.; Crossett, B.; Falconer, L.; Kolarich, D.; Djordjevic, S. P.; Hojrup, P.; Packer, N. H.; Larsen, M. R.; Cordwell, S. J. Simultaneous glycan-peptide characterization using hydrophilic interaction chromatography and parallel fragmentation by CID, higher energy collisional dissociation, and electron transfer dissociation MS applied to the N-linked glycoproteome of *Campylobacter jejuni*. *Molecular & cellular proteomics* **2011**, 10, (2), M000031-MCP201.
74. Catalina, M. I.; Koeleman, C. A.; Deelder, A. M.; Wuhler, M., Electron transfer dissociation of N-glycopeptides: loss of the entire N-glycosylated asparagine side chain. *Rapid communications in mass spectrometry* **2007**, 21, (6), 1053-61.
75. Ye, H.; Boyne, M. T., 2nd; Buhse, L. F.; Hill, J. Direct approach for qualitative and quantitative characterization of glycoproteins using tandem mass tags and an LTQ Orbitrap XL electron transfer dissociation hybrid mass spectrometer. *Analytical chemistry* **2013**, 85, (3), 1531-9.
76. Alley, W. R., Jr.; Mechref, Y.; Novotny, M. V. Characterization of glycopeptides by combining collision-induced dissociation and electron-transfer dissociation mass spectrometry data. *Rapid communications in mass spectrometry* **2009**, 23, (1), 161-70.

77. Cooper, C. A.; Gasteiger, E.; Packer, N. H. GlycoMod--a software tool for determining glycosylation compositions from mass spectrometric data. *Proteomics* **2001**, 1, (2), 340-9.
78. Cooper, J. D.; Bird, S. M. Predicting incidence of variant Creutzfeldt-Jakob disease from UK dietary exposure to bovine spongiform encephalopathy for the 1940 to 1969 and post-1969 birth cohorts. *International journal of epidemiology* **2003**, 32, (5), 784-91.
79. Cooper, C. A.; Harrison, M. J.; Wilkins, M. R.; Packer, N. H. GlycoSuiteDB: a new curated relational database of glycoprotein glycan structures and their biological sources. *Nucleic acids research* **2001**, 29, (1), 332-5.
80. Ranzinger, R.; Herget, S.; von der Lieth, C. W.; Frank, M. GlycomeDB--a unified database for carbohydrate structures. *Nucleic acids research* **2011**, 39, D373-6.
81. Ranzinger, R.; Herget, S.; Wetter, T.; von der Lieth, C. W. GlycomeDB - integration of open-access carbohydrate structure databases. *BMC Bioinformatics* **2008**, 9, 384.
82. Ceroni, A.; Maass, K.; Geyer, H.; Geyer, R.; Dell, A.; Haslam, S. M. GlycoWorkbench: a tool for the computer-assisted annotation of mass spectra of glycans. *J Proteome Res* **2008**, 7, (4), 1650-9.
83. Vakhrushev, S. Y.; Dadimov, D.; Peter-Katalinic, J. Software platform for high-throughput glycomics. *Anal Chem* **2009**, 81, (9), 3252-60.
84. Apte, A.; Meitei, N. S. Bioinformatics in glycomics: glycan characterization with mass spectrometric data using SimGlycan. *Methods in molecular biology* **2010**, 600, 269-81.
85. Ozohanics, O.; Krenyacz, J.; Ludanyi, K.; Pollreis, F.; Vekey, K.; Drahos, L. GlycoMiner: a new software tool to elucidate glycopeptide composition. *Rapid Commun Mass Spectrom* **2008**, 22, (20), 3245-54.
86. Goldberg, D.; Bern, M.; Parry, S.; Sutton-Smith, M.; Panico, M.; Morris, H. R.; Dell, A. Automated N-glycopeptide identification using a combination of single- and tandem-MS. *Journal of proteome research* **2007**, 6, (10), 3995-4005.
87. Woodin, C. L.; Hua, D.; Maxon, M.; Rebecchi, K. R.; Go, E. P.; Desaire, H. GlycoPep Grader: A Web-Based Utility for Assigning the Composition of N-Linked Glycopeptides. *Anal Chem* **2012**, 84, (11), 4821-9.
88. Deshpande, N.; Jensen, P. H.; Packer, N. H.; Kolarich, D. GlycoSpectrumScan: fishing glycopeptides from MS spectra of protease digests of human colostrum sIgA. *J Proteome Res* **2010**, 9, (2), 1063-75.
89. Irungu, J.; Go, E. P.; Dalpathado, D. S.; Desaire, H. Simplification of mass spectral analysis of acidic glycopeptides using GlycoPep ID. *Analytical chemistry* **2007**, 79, (8), 3065-74.
90. Schiel, J. E.; Au, J.; He, H. J.; Phinney, K. W. LC-MS/MS biopharmaceutical glycoanalysis: identification of desirable reference material characteristics. *Analytical and bioanalytical chemistry* **2012**, 403, (8), 2279-89.
91. Kolarich, D.; Rapp, E.; Struwe, W. B.; Haslam, S. M.; Zaia, J.; McBride, R.; Agravat, S.; Campbell, M. P.; Kato, M.; Ranzinger, R.; Kettner, C.; York, W. S. The Minimum Information Required for A Glycomics Experiment (MIRAGE) Project - Improving the standards for reporting mass spectrometry-based glycoanalytic data. *Molecular & cellular proteomics* **2013**.

PARTIE II :

RESULTATS

Chapitre 1 : Etude du parasite *Toxoplasma gondii* responsable de la toxoplasmose

Chapitre 2 : Etudes protéomique et génomique des changements induits par la maladie d'esca sur le *Vitis vinifera L.cv. Chardonnay*

Chapitre 3 : Etude et identification de nouveaux variants d'hémoglobine humaine

Chapitre 4 : Caractérisation des glycosylations de l'invertase vacuolaire de raisin

Chapitre 1 :

Etude du parasite *Toxoplasma gondii* responsable de la toxoplasmose

Cette étude a été réalisée en collaboration avec le Dr. S. Tomavo de l'Institut Pasteur de Lille.

Ce sujet s'inscrit dans la continuité de la thèse réalisée au laboratoire par Agnès Hovasse¹. Au cours de ce travail de thèse, une première étude de caractérisation d'une glycoprotéine de *T. gondii* impliquée dans l'invasion du parasite chez l'hôte a été menée. En parallèle, une étude protéomique des facteurs de transcriptions impliqués dans l'interconversion de la forme virulente en forme dormante du parasite avait permis de mettre en évidence 39 facteurs de transcription^{2,3}.

Ces différentes études ont ainsi permis aux biologistes de dégager des pistes intéressantes pour ensuite aller "décortiquer" la fonction biologique de chacune de ces protéines. L'étude des différentes protéines parasitaires (TgGAP50, les TgAlba et TgSORTLR) a donc fait l'objet de ma thèse.

I. Contexte de cette étude

1. *Toxoplasma gondii* et la toxoplasmose

Toxoplasma gondii, l'agent pathogène responsable de la toxoplasmose, a été isolé pour la première fois en 1908 chez le rongeur africain *Ctenodactylus gondii*^{4,5}. *T. gondii* est un organisme eucaryote protozoaire et un parasite obligatoire appartenant au phylum des Apicomplexa (sporozoaires) qui inclut d'autres pathogènes comme *Plasmodium falciparum*, responsable de la malaria chez l'homme.

T. gondii est très répandu dans le monde et touche un large éventail d'hôtes, principalement les animaux à sang chaud, ce qui en fait un des parasites les plus étudiés. La prévalence de la toxoplasmose dans la population humaine varie d'un pays à l'autre, en fonction des conditions d'hygiène ou des habitudes alimentaires. Des études épidémiologiques menées entre 1990 et 2001 ont montré que 23% aux USA et 58% en France des populations étudiées étaient atteintes de toxoplasmose^{6,7}. Dans un rapport publié en 2005 par l'Agence Française de Sécurité Sanitaire des Aliments, la séroprévalence de la toxoplasmose en France était estimée entre 50 et 70% de la population adulte⁸. Aujourd'hui, on estime à un tiers la population mondiale infectée par la toxoplasmose.

La toxoplasmose est une maladie asymptomatique et bénigne mais peut être fatale chez les individus immunodéprimés comme cela est le cas pour les personnes atteintes du sida^{9,10}. Il existe plusieurs voies d'infection de la toxoplasmose parmi lesquelles : l'infection par ingestion^{11,12}, l'infection congénitale^{13,14}, l'infection par greffe d'organes^{15,16} et/ou encore celle par transfusion¹⁷. La toxoplasmose congénitale, cérébrale et oculaire sont les cas de toxoplasmose les plus cités chez l'homme¹⁸. La toxoplasmose est à l'origine de nombreux avortements dans les élevages de bétail et a donc un impact économique important¹⁹.

T. gondii a un cycle de vie appelé hétéroxène, qui comporte une phase sexuée qui se déroule chez l'hôte définitif uniquement (le chat et d'autres félinés) et une phase de multiplication asexuée qui se déroule chez les hôtes intermédiaires (les animaux à sang chaud). Les hôtes sont contaminés par ingestion d'aliments ou d'eau souillés par des oocystes. Chez l'homme, l'infection est le plus souvent causée par la consommation de fruits et légumes non épluchés ou par la consommation de viandes crues. La toxoplasmose au cours d'une grossesse, peut se transmettre d'une mère à l'enfant via le placenta (Figure 1).

Le parasite se présente sous différentes formes selon le stade de son évolution au cours de son cycle de vie²⁰. On distingue les 4 stades parasitaires :

- le tachyzoïte, qui désigne la forme virulente, se multiplie très rapidement à l'intérieur d'une vacuole parasitophore dans les cellules des hôtes ;
- le bradyzoïte, qui désigne la forme latente, se multiplie très lentement dans les cellules des hôtes, et est souvent renfermé dans des kystes ;
- le sporozoïte, désigne la forme infectieuse, provient de la sporulation des oocystes libérés dans les matières fécales par l'hôte définitif (chat et d'autres félinés) ;

Comme tout eucaryote, *T. gondii* possède la machinerie nécessaire à la synthèse et au trafic de ses propres protéines : noyau, appareil de Golgi, mitochondrie, réticulum endoplasmique (Figure 2). Cependant, *T. gondii* prolifère obligatoirement dans les cellules de ses hôtes desquels il tire un certain nombre de nutriments essentiels à son développement, et pour lesquelles il ne dispose pas de voie métabolique propre.

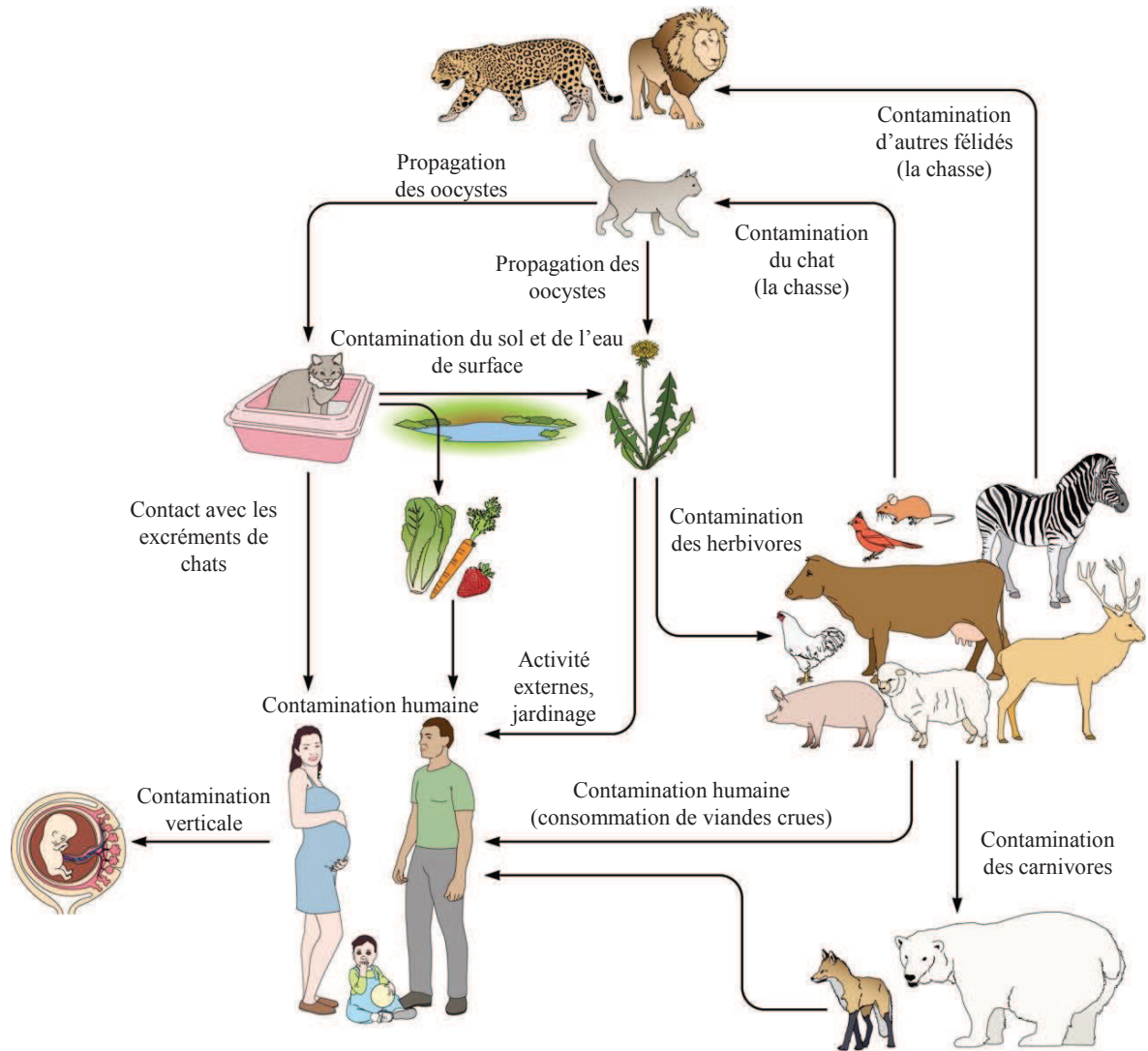


Figure 1: Source de contamination de l'homme par *T. gondii* ²¹.

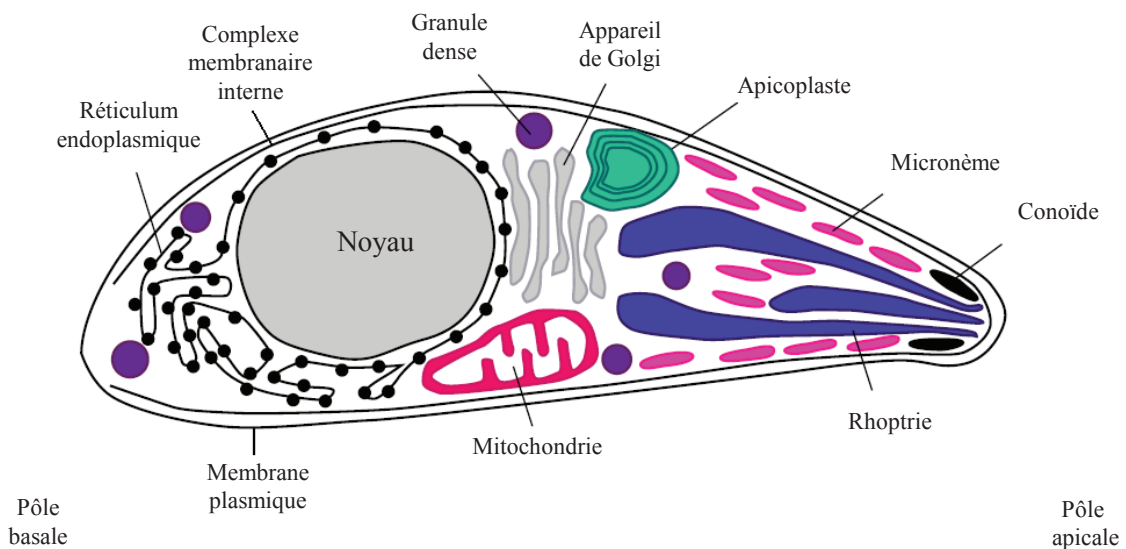


Figure 2. Schéma de l'ultra-structure de la forme tachyzoïte de *Toxoplasma gondii* ²².

Trois organelles sécrétoires, spécifiques au phylum des Apicomplexa, sont présentes dans le cytoplasme de *T. gondii* : les micronèmes, les rhoptries et les granules denses. Ces organelles déversent leur contenu protéique (ex. protéines telles que les rhoptries (ROP) et les micronèmes (MIC)) dans la vacuole parasitophore

autour du parasite à travers la membrane plasmique après le conoïde. Ces organelles sont impliquées dans l'invasion et dans l'interaction des parasites avec les cellules hôtes²³⁻²⁵.

T. gondii comme les autres Apicomplexa, ne possède pas de cils ni flagelles pour sa mobilité. L'invasion active du parasite est basée sur la présence d'un complexe moteur actine-myosine permettant la translocation du pôle apical vers le pôle basal du parasite. Ce mouvement par glissement est dépendant de la polymérisation des filaments d'actines parasitaires avec lesquels interagissent les molécules de myosine. La machinerie moléculaire responsable de ce mouvement porte le nom de "glidosome"²⁶.

La conversion de la forme virulente tachyzoïte à la forme latente bradyzoïte et de la transformation de la vacuole parasitophore en kyste est un phénomène fréquemment observé au cours de la réponse immunitaire des hôtes intermédiaires. Ce phénomène est contrôlé par l'inhibition de l'activité mitochondriale des parasites sous l'influence de protéines du stress du toxoplasme^{27, 28}. Ce processus d'inter-conversion est au centre de la pathogenèse, puisqu'il s'accompagne de l'activation ou de la répression des gènes spécifiques²⁹ et pourrait impliquer des processus de régulation transcriptionnelle encore mal connus.

2. Problématique et objectifs de l'étude

Les incidences graves de la toxoplasmose congénitale chez les nouveaux nés ou chez les patients immunodéprimés comme les sidéens et les taux de prévalence élevés dans les populations, ont conduit des équipes de recherche à se lancer dans la quête de solutions thérapeutiques contre cette maladie. Ainsi des tests de dépistage et de diagnostic de la toxoplasmose utilisant des antigènes dirigés contre les anticorps de *T. gondii* ont été proposés³⁰. La recherche de l'ADN parasitaire, comme par exemple le gène de SAG1 par réaction en chaîne par Polymérase (RT-PCR), a permis de détecter la présence de *T. gondii* chez un model murin³¹. La recherche de vaccins contre la toxoplasmose a conduit à la découverte du vaccin S48 d'abord utilisé chez les ovins, et par la suite chez le chat³². Il n'existe à ce jour aucun de vaccin contre la toxoplasmose développée pour l'homme. Le développement de vaccins contre la toxoplasmose passe par une meilleure connaissance du parasite et de l'interaction hôte-parasite. Il est donc nécessaire de mieux comprendre la biologie du parasite et les mécanismes qui se mettent en place lors de l'infection.

L'étape d'invasion du parasite dans les cellules hôtes est contrôlée par la motilité et la formation de jonction mobile permettant à *T. gondii* de pénétrer à l'intérieur des cellules de mammifères. La présence de N-glycoprotéines chez *T. gondii* et la participation de N-glycosylation, dans le processus d'invasion chez les Apicomplexa ont fait l'objet de controverse³³. De plus, l'absence des gènes codant pour les enzymes de transfert des N-glycanes sur les protéines chez certains Apicomplexa tel que le *Plasmodium falciparum* n'a fait qu'alimenter cette controverse. Néanmoins quelques études ont mis en évidence la présence de glycoprotéines chez *T. gondii*³⁴.

L'étape cruciale de la pathogénicité de ce parasite réside dans sa conversion du stade virulent (tachyzoïte) au stade latent (bradyzoïte) et inversement. Cette inter-conversion de *T. gondii* permet l'enkystement de *T. gondii* et le rend résistant aux attaques de médicaments tels que la pyriméthamine et la sulfadiazine lors des traitements des formes graves de toxoplasmose chez les patients immunodéprimés³⁵. Dans les systèmes immunitaires affaiblis comme le cas des sidéens, le parasite enkysté (bradyzoïte), peut reprendre sa forme virulente tachyzoïte et causer des encéphalites mortelles. Il a bien été mis en évidence que, les gènes exprimés pendant les deux phases (virulente et latente) sont différents²⁸, et que le contrôle de la régulation de la transcription est un facteur clé dans le processus de développement et la virulence³⁶. Cependant les protéines qui contrôlent ce mécanisme de régulation des gènes sont encore mal connues.

Trois organelles sécrétoires apicales : rhoptries, micronèmes et granules denses sont essentiels pour l'invasion, et la survie du parasite dans les cellules hôtes. Les récepteurs qui sont impliqués dans le tri et le transport des protéines présentes dans ces organelles sécrétoires ne sont pas encore connus mais sont au centre du développement des stratégies thérapeutiques.

Ce travail a donc porté sur trois aspects fondamentaux de la biologie du parasite *T. gondii* (la motilité, la virulence et le trafic intracellulaire du parasite), avec pour objectif final d'apporter des éléments de réponse moléculaire et donc de cibles pour développer de nouvelles stratégies thérapeutiques.

L'objectif commun entre ces 3 sous-projets (détaillés dans les trois parties ci-après) est d'apporter des connaissances sur la fonction biologique et sur les mécanismes d'action de plusieurs protéines clés de *T. gondii*. Ceci impliquait donc d'étudier les complexes protéiques dans lesquelles ces différentes protéines sont impliquées et d'identifier leurs protéines partenaires.

En effet, les interactions protéines-protéines sont impliquées dans de nombreuses fonctions essentielles pour le parasite telles que la régulation de l'expression des gènes, les voies de signalisation, le transport intracellulaire des protéines ou la modulation de la spécificité d'un substrat pour une enzyme, etc. La caractérisation de ces complexes est donc essentielle pour une meilleure compréhension des mécanismes moléculaires intracellulaires. Pour étudier ces complexes protéiques, la combinaison de méthodes d'immuno-purification (pour enrichir en complexes protéiques) et l'analyse protéomique par spectrométrie de masse représentent des méthodes de choix^{37, 38}. Ainsi des techniques d'immuno-purification ont été mises en place lors de notre étude pour « piocher » les complexes protéiques concernés par l'intermédiaire d'une protéine appât qui correspond à protéine cible étudiée.

Dans **une première partie**, nous avons abordé l'étude du complexe multi-protéique du "glidéosome" et plus particulièrement, le rôle des glycosylations sur les interactions des protéines partenaires de la N-glycoprotéine impliquée dans la motilité parasitaire, TgGAP50 (*Toxoplasma gondii* Glideosome Associated Protein 50).

Dans **une deuxième partie**, nous avons abordé l'étude de deux protéines identifiées comme spécifiquement liées à certains motifs d'ADN ou d'ARN indispensables à la transcription ou traduction induisant le changement de formes de *T. gondii*. Ces deux protéines nommées TgAlba (*T. gondii* Acetylation Lower Binding Affinity) présentent de fortes homologues avec des protéines d'archaebactéries^{39, 40}. Nos investigations ont porté plus spécifiquement sur l'identification des protéines partenaires de ces deux facteurs de transcription Alba chez *T. gondii*.

Dans **une troisième partie**, nous nous sommes intéressés à un récepteur membranaire putatif nommé TgSORTLR (*Toxoplasma gondii* Sortilin-Like Receptor), homologue de la sortiline humaine qui est décrite notamment pour jouer un rôle dans la maladie d'Alzheimer et le métabolisme du cholestérol^{41, 42}.

Nous concluons le chapitre par l'ensemble des apports de notre travail de protéomique dans l'étude du parasite *T. gondii* et les différentes perspectives.

II. Identification des protéines impliquées dans l'intrusion du parasite chez l'hôte

Ce travail a fait l'objet d'un article publié dans *Molecular & Cellular Proteomics* ⁴³ disponible dans l'annexe 1.

1. Contexte et objectif de cette étude

Une étude protéomique et glycomique menée au laboratoire ², a permis d'identifier 15 protéines dont 6 sont impliquées dans l'invasion de la cellule hôte par le parasite. TgGAP50 encore appelée "Membrane anchor for myosin XIV precursor" était l'une des glycoprotéines identifiées. TgGAP50 fait partie des constituants principaux du "glidéosome". Le "glidéosome" est un complexe macromoléculaire, constitué de protéines adhésives du cytoplasme mais également libérées par les micronèmes dans la région apicale et transloquées vers le pôle basal par l'action des filaments d'actine et de myosine ancrés dans le complexe membranaire interne. TgGAP50 est une protéine membranaire, qui constitue le point essentiel d'ancrage de l'ensemble du mécanisme d'invasion des cellules hôtes et possède trois sites consensus de glycosylation Asn-X-(Ser/Thr): Asn¹⁰¹, Asn¹³⁶, Asn²²⁸. Parmi les trois sites de glycosylation, un seul (Asn¹³⁶) avec deux glycoformes avait pu être identifié ². Grâce à une stratégie spécifiquement développée pour l'analyse de glycopeptides, la caractérisation des 3 sites avec la description de leur micro-hétérogénéité a pu être réalisée ¹.

Ce travail présente l'étude protéomique d'identification des partenaires de TgGAP50 glycosylée et non glycosylée avec pour objectif de déterminer l'impact de la glycosylation de TgGAP50 dans le processus d'invasion du parasite chez l'hôte.

2. Stratégie d'analyse protéomique des complexes protéiques de TgGAP50

a. Description des échantillons

Deux types de parasites différents exprimant d'un côté, le clone de la protéine TgGAP50 sauvage TgGAP50wt HA-FLAG sous sa forme glycosylée, et de l'autre un mutant de TgGAP50mx HA-FLAG non glycosylé, ont été produits en grande quantité (10⁸ parasites pour chaque type de parasite). Le mutant TgGAP50mx HA-FLAG a été produit par mutagenèse dirigée, par remplacement de chacune des asparagines 101, 136 et 228 par des alanines pour empêcher la glycosylation de TgGAP50 chez le mutant. Les parasites exprimant chacune différentes formes de TgGAP50 ont été lysés et les contenus cytoplasmiques ont été purifiés par chromatographie d'affinité (billes FLAG M2 agarose Sigma-Aldrich) spécifique anti-FLAG ³⁷.

Les éluats de chaque purification ont ensuite été concentrés et séparés sur gel SDS-PAGE. Après coloration au bleu de Coomassie, les bandes de gels SDS-PAGE obtenues ont été découpées de façon systématique pour l'analyse nanoLC-MS/MS.

Les protocoles utilisés pour la préparation des échantillons sont détaillés dans la partie expérimentale de la publication (Annexe 1).

b. Analyses nanoLC-MS/MS

La stratégie adoptée pour l'identification des partenaires de TgGAP50 glycosylée et non glycosylée est décrite sur la Figure 3.

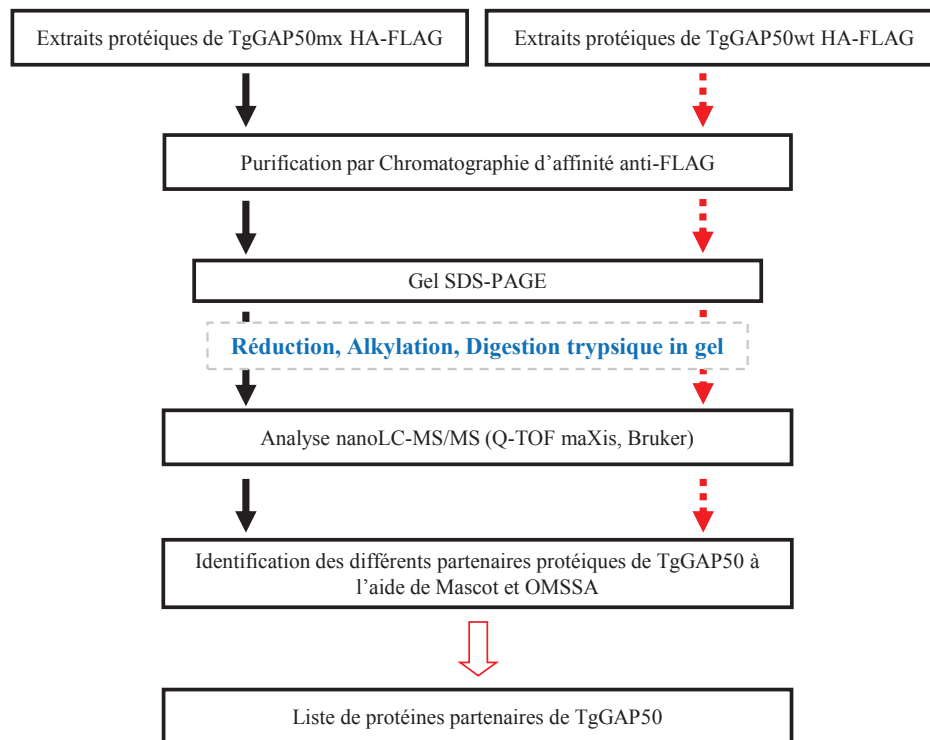


Figure 3. Stratégie globale d'analyse protéomique pour l'identification des partenaires protéiques de TgGAP50 glycosylée et non glycosylée.

La faible quantité de matériel due à la préparation des échantillons, nous a conduit à faire une décomplexification des échantillons sur gel SDS-PAGE, afin de séparer les protéines mineurs des protéines majoritaires suivant leur poids moléculaires. Pour l'analyse nanoLC-MS/MS, notre choix a porté sur un couplage entre la nanoACQUITY Ultra-performance-LC (Waters) avec un spectromètre de masse de type Q-TOF (maXis, Bruker). Ce système nanoACQUITY avec les petites particules de ses nano-colonnes offre une bonne résolution chromatographique avec des pics fins permettant de gagner en sensibilité. Quant au Q-TOF, il possède une vitesse de balayage assez rapide (quelques millisecondes (ms)) pour permettre le séquençage d'un grand nombre de peptides élués et une très bonne précision de masse de l'ordre de 5 ppm en MS.

Pour l'identification des protéines partenaires de TgGAP50 deux algorithmes de recherche (Mascot et Open Mass Spectrometry Search Algorithm OMSSA) ont été utilisés pour augmenter la fiabilité des résultats^{44, 45}. Les paramètres utilisés lors de ces recherches et les critères de validation des résultats sont décrits dans les résultats publiés (Annexe 1).

3. Résultats publiés

Les analyses nanoLC-MS/MS des bandes de gel SDS-PAGE des partenaires de TgGAP50 sauvage glycosylée (protéine native non mutée) ont permis d'identifier une vingtaine de protéines de fonctions biologiques connues et d'autres de fonctions biologiques inconnues "hypothetical protein". Parmi les protéines de fonction connue, nous avons identifié différents composants du "glidéosome" (TgGAP45, TgMLC1 et TgMyoA), l'actine, des enzymes de la glycolyse ainsi que des protéines intervenant dans le trafic intracellulaire. Parmi les protéines inconnues, un certain nombre avait des segments transmembranaires putatifs, et présentaient des sites consensus de N-glycosylation (Annexe 1, Tableau 1). Ces protéines seraient potentiellement des protéines membranaires encore non décrites.

La comparaison des résultats d'identification des partenaires de TgGAP50 sauvage et les résultats d'identification des partenaires TgGAP50 mutante non glycosylée, a permis de montrer l'absence de certaines protéines chez le mutant TgGAP50. Par exemple, TgGAP45, les enzymes de glycolyse et quelques protéines inconnues n'ont pas été identifiées chez le mutant (Annexe 1, Tableau 1). Ces résultats montrent la spécificité de la N-glycosylation de TgGAP50 dans les interactions protéines-protéines du "glidéosome". De plus, grâce à l'alignement de séquences des protéines inconnues identifiées, des protéines orthologues aux protéines des Apicomplexa (*P. falciparum*, *N. caninum*, *C. parvum* et *E. tenella*) ont pu être identifiées par homologie de séquence⁴⁶. En outre des segments transmembranaires putatifs ont pu être identifiés chez certaines de ces

protéines inconnues, et qui sont considérées désormais comme des protéines potentiellement impliquées dans l'invasion du parasite.

Conclusion

Grace à la stratégie d'analyse protéomique des partenaires de TgGAP50 glycosylée et non glycosylée, nous avons identifié les composants du "glidéosome" et des protéines intervenant dans le trafic intracellulaire. La comparaison des résultats obtenus pour le mutant et la protéine TgGAP50 sauvage, a permis d'identifier les protéines du complexe membranaire interne directement impactées par la glycosylation de TgGAP50. En outre, les expériences de mutagenèse dirigée de chacun des sites consensus de glycosylation de TgGAP50 et de co-localisation par microscopie confocale faites par nos collaborateurs biologistes ont permis de montrer que ces sites de glycosylation ne jouent pas tous le même rôle. Par exemple, la mutation de l'asparagine du site 101 ou du site 136 par une alanine ou des trois sites, a conduit à une délocalisation de 75% de TgGAP50 par rapport aux autres protéines du "glidéosome" alors que la mutation du site 228 n'a provoqué qu'une délocalisation de 25%. L'ensemble de ces études démontre le rôle fondamental de la glycosylation dans la formation et le maintien du complexe membranaire impliqué dans l'invasion du parasite dans les cellules hôtes.

Ces études de protéomique ont permis d'orienter le choix des tests biologiques menés par nos collaborateurs biologistes. Par exemple, le suivi de l'expression des gènes des protéines "hypothetical" a montré que certaines d'entre elles avaient le même profil d'expression que TgGAP50⁴⁶. Les expériences de microscopie confocale sur des parasites exprimant les mutants et la protéine TgGAP50 sauvage ont permis de confirmer la co-localisation des TgGAP50 et de certaines protéines identifiées par analyse protéomique.

Annexe 1: Publication

Unusual N-glycan structures required for trafficking *Toxoplasma gondii* GAP50 to the inner membrane complex regulate host cell entry through parasite motility

Fauquenoy, S. ; Hovasse, A. ; Sloves, P. J. ; Morelle, W. ; Dilezitoko Alayi, T. ; Slomianny, C.;
Werkmeister, E. ; Schaeffer, C. ; Van Dorsselaer, A. ; Tomavo, S.

Molecular & cellular proteomics, 2011, 10:9

Unusual *N*-glycan Structures Required for Trafficking *Toxoplasma gondii* GAP50 to the Inner Membrane Complex Regulate Host Cell Entry Through Parasite Motility*[§]

Sylvain Fauquenoy[‡], Agnès Hovasse[§], Pierre-Julien Sloves[‡], Willy Morelle[¶], Tchilabalo Dilezitoko Alayi[§], Christian Slomianny^{||}, Elisabeth Werkmeister[‡], Christine Schaeffer[§], Alain Van Dorsselaer[§], and Stanislas Tomavo^{‡**}

Toxoplasma gondii motility, which is essential for host cell entry, migration through host tissues, and invasion, is a unique form of actin-dependent gliding. It is powered by a motor complex mainly composed of myosin heavy chain A, myosin light chain 1, gliding associated proteins GAP45, and GAP50, the only integral membrane anchor so far described. In the present study, we have combined glycomic and proteomic approaches to demonstrate that all three potential *N*-glycosylated sites of GAP50 are occupied by unusual *N*-glycan structures that are rarely found on mature mammalian glycoproteins. Using site-directed mutagenesis, we show that *N*-glycosylation is a prerequisite for GAP50 transport from the endoplasmic reticulum to the Golgi apparatus and for its subsequent delivery into the inner complex membrane. Assembly of key partners into the gliding complex, and parasite motility are severely impaired in the unglycosylated GAP50 mutants. Furthermore, comparative affinity purification using *N*-glycosylated and unglycosylated GAP50 as bait identified three novel hypothetical proteins including the recently described gliding associated protein GAP40, and we demonstrate that *N*-glycans are required for efficient binding to gliding partners. Collectively, these results provide the first detailed analyses of *T. gondii* *N*-glycosylation functions that are vital for parasite motility and host cell entry. *Molecular & Cellular Proteomics* 10: 10.1074/mcp.M111.008953, 1–17, 2011.

From the [‡]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; [§]Laboratoire de Spectrométrie de Masse Bioorganique, IPHC, CNRS UMR7178, 67087 Strasbourg, France; [¶]Unité de Glycobiologie Structurale et Fonctionnelle, CNRS UMR 8576, IFR 147, Université des Sciences et Technologies de Lille 1, 59655 Villeneuve d'Ascq, France; ^{||}Laboratoire de Physiologie Cellulaire, INSERM U 1003, Université des Sciences et Technologies de Lille 1, 59655 Villeneuve d'Ascq, France

Received February 24, 2011, and in revised form, May 3, 2011

Published, MCP Papers in Press, May 24, 2011, DOI 10.1074/mcp.M111.008953

Apicomplexans are protozoan intracellular parasites that cause many life threatening diseases of humans and animals including malaria (*Plasmodium* spp.), chicken coccidiosis (*Eimeria* spp.), lymphoproliferative diseases of cattle (*Theileria* spp.), as well as some of the opportunistic infections associated with AIDS patients (*Cryptosporidium parvum* and *Toxoplasma gondii*). Among apicomplexan parasites, *T. gondii* is uniquely adapted to infect a wide range of hosts, including virtually all warm blooded animals. Like all other apicomplexan parasites, the ability of *T. gondii* to invade host cells is critical for its survival and its ability to egress from host cells is equally essential to ensure its survival and proliferation through host tissues. Therefore, the apicomplexan parasites require motility to migrate through distinct invertebrates and/or vertebrate hosts and internationalize themselves within targeted host cells. Yet, these parasites lack the classical means of motility such as flagellum, cilia, pseudopodia, and lamellipodia. Instead, they move by a unique form of actin-based locomotion called gliding motility, which is characterized by circular, forward twisting and rolling movements (1, 2).

Gliding motility requires apical organelles such as micronemes for efficient release of micronemal adhesin proteins (MIC2) onto the parasite surface for directional attachment (3). The cytoplasmic tail of extracellular *T. gondii* adhesin MIC2 is linked internally to an anchor, the glycolytic enzyme aldolase, which in turn mediates association with actin filaments underlying the plasma membrane (4) (also depicted in Fig. 9). Ultimately, the unique mechanism of parasite gliding motility is dependent on the interaction of actin filaments and the type XIV myosin (TgMyoA) located between the plasma membrane and its underlying outer flattened membrane cisternae, called the inner membrane complex (IMC)¹ (see Fig. 9). The motor

¹ The abbreviations used are: ER, endoplasmic reticulum; DMEM, Dulbecco's modified Eagle's medium; GAP, gliding-associated protein; HFF, human foreskin fibroblast; HA, hemagglutinin; IP, immunoprecipitation; IMC, inner membrane complex; MyoA, myosin A; MLC, myosin light chain; DHB, 2,5-dihydroxybenzoic acid (DHB, Sigma); PFA, paraformaldehyde.

essential for gliding motility in apicomplexan parasites is composed of a tetrameric complex encompassing MyoA, a myosin tail interacting protein also called myosin light chain 1 (MLC1) (5–7), and the two glideosome-associated proteins GAP45 and GAP50, which are anchored to the outer IMC membrane (8, 9). GAP50 is firmly immobilized in the IMC and considered to act as a fixed anchor for the motor complex (10). To power the forward movement of apicomplexan parasites, myosin pulls the actin filaments and their attached adhesins rearward. This requires the GAP-myosin complex to be anchored to the IMC and the underlying cytoskeleton (9–11), in addition to energy and diverse regulatory events. Most importantly, the gliding machinery (including MyoA, MLC1, GAP45, and GAP50) appears to be highly and exclusively conserved across Apicomplexa species (12). Therefore, the elucidation of molecular mechanisms controlling gliding motility and, ultimately the invasion of and egress from host cells could be valuable for the development of new therapeutic strategies against these major human and animal pathogens. A recent study focusing on TgMLC1 demonstrated that the mechanical activity of Class XIV myosins (TgMyoA) can be modulated by a small-molecule inhibitor of myosin motor activity, which interferes with the post-translational modifications of its associated light chains (13). The regulation of gliding motility can also be controlled at additional stages, including actin polymerization (14) and calcium-mediated secretion of adhesins (15). The final assembly of the gliding motor is controlled by phosphorylation of pre-assembled TgGAP45-MyoA-MLC1 complex (16). In addition, the gliding associating protein GAP45 is modified by acylation, whereas GAP50 is an integral membrane protein of the IMC. Both GAP45 and GAP50 were proposed to anchor MyoA to the IMC (9). The precomplex MyoA-MLC1-GAP45 is anchored to the IMC through association with GAP50 and GAP45, whose acylation by both myristoylation and palmitoylation appears as potentially important translational modifications (17). *T. gondii* is also capable of mobilizing its main source of energy by differential sub-localization of glycolytic enzymes, which are located between the plasma membrane and IMC, instead of being normally present in the parasite cytoplasm (18). This novel location of glycolytic enzymes, which remain pellicle-associated until parasites have completed host cell invasion, probably optimizes ATP delivery required for gliding motility.

Despite the importance of *N*-glycosylation of proteins, a relatively common post-translational modification of eukaryotic proteins translocated into the endoplasmic reticulum, several studies have suggested that this is a rare event in apicomplexan parasites (19, 20). Indeed, the presence of *N*-glycosylation for the gliding associated protein TgGAP50, which was initially suspected by Gaskins *et al.* (9), has only recently been demonstrated by us and others (21, 22). Beyond a very limited number of *N*-glycoproteins discovered in the parasite, the precise functions of *N*-glycans are still unknown in *T. gondii*. Here, we present evidence that the un-

usual *N*-glycan structures on the TgGAP50 protein are crucial for its trafficking through ER/Golgi apparatus to a final loading into the IMC, and for its subsequent interactions with other key gliding partners. Our study provides insight into how *N*-glycosylation regulates the trafficking of GAP50 and its association to other members of the actin-myosin motor complex and also demonstrates that *N*-glycosylation is important for parasite motility and host cell entry.

EXPERIMENTAL PROCEDURES

Culture of Parasites—Human foreskin fibroblasts (HFFs) were maintained in Dulbecco's modified Eagle's medium (DMEM, Invitrogen) supplemented with 10% fetal calf serum (FCS, Invitrogen), 2 mM L-glutamine, Invitrogen) and 10 000 U/ml Penicillin-Streptomycin (Invitrogen). Tachyzoites from *T. gondii* RH 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-mm pore filters followed by centrifugation at 2200 rpm for 15 min.

Cloning of TgGAP50 into *T. gondii* Expressing Vector and Site-Directed Mutagenesis—The gene encoding TgGAP50 protein was obtained in <http://www.toxodb.org> and the following forward and reverse primers, 5'-ccggctgcagATGGCAGGCGCCCCGTC-3' and 5'-ccggctgcagTTTCATGTAGCGAGAGACCGTTC-3', were designed to amplify the complete open reading frame of TgGAP50 using cDNA from tachyzoites of RH strain and Accuprime TaqDNA Polymerase High Fidelity (Invitrogen). The PCR product was purified and cloned in pMAH14 plasmid using PstI restriction site and nucleotide sequence accuracy was checked by DNA sequencing. The pMAH14 vector allows fusion at the C-terminus with the HAFLAG tag and expression of TgGAP50-HAFLAG was under the control of the GRA1 promoter (23). The selectable marker, bleomycin, was amplified from pTub5Bleo plasmid by PCR, sequenced and subcloned via KpnI restriction sites into pTgGAP50-HAFLAG, resulting in a pMAH14-TgGAP50-HAFLAG vector containing the bleomycin-selectable marker (under the control of tubulin promoter) for the cloning of transgenic stable parasites (23). The replacement of individual and all three potential *N*-glycosylation sites (Asn101, 136 and 228) of TgGAP50 to Ala was performed by PCR using QuickChange Site-Directed Mutagenesis Kit (Stratagene). The plasmid pMAH14TgGAP50-HAFLAG encoding the wild-type GAP50 protein was used as template with the following primer pairs, which allow mutation of Asn residues (underlined): 5'-GGCGGAGTTTCCAGCCTGGCGGACACCAGGTGGCAATCC-3', 5'-GGATTGCCACCTGGTGTCCGCCAGGCTGGAAAC-TCCGCC-3' for the first potential *N*-glycosylation site (the mutant was designated pMAH14TgGAP50_N_101A-HAFLAG); 5'-GTGCATGACTGGAGTAGAGCGTACACCTCTGAGGCCCTG-3', 5'-CAGGGCCTCAGAGGTG TACGCTCTACTCCAGTCATCGAC-3' for the second potential *N*-glycosylation site (pMAH14TgGAP50_N_136A-HAFLAG); 5'-TCGTCGTTCCCGTTCTCGGCGGTGACGAGTCCG CCGCTGG-3', 5'-CCAGGCGGACTCGTCACCGCCGAGAACGGGAACGACGA-3' for the third potential *N*-glycosylation site (pMAH14TgGAP50N_228A-HAFLAG). The following double mutants were also generated: pMAH14TgGAP50N_101_136-AA-HAFLAG, pMAH14TgGAP50 N_101_228-AA-HAFLAG, pMAH14TgGAP50N_136_228-AA-HAFLAG. Finally, the triple mutant pMAH14TgGAP50_N_101_136_228_AAA-HAFLAG plasmid was made by successive PCR and site-directed mutation that sequentially replaced all three Asn to Ala. In addition, the GRA1 promoter was replaced by the putative TgGAP50 promoter using genomic DNA corresponding to 2.2 kb chosen upstream of the ATG of the gene coding GAP50. The following forward and reverse primers were used to amplify the promoter region of TgGAP50 and PCR product was cloned between BamHI and SmaI restriction sites:

5'-ccggccgggTTGGAGTTGGCCGAGAGCAG-3' and 5'-ccgggag-ctcGCAAAGCATTGGTAAGCACGTGC-3'.

Transient and Stable *T. gondii* Transfections and Immunofluorescence Assay—Transient transfection was performed using 100 µg of circular plasmids corresponding to wild type and mutated TgGAP50 into 10⁷ tachyzoites of the RH strain. After transfection, the parasites were loaded onto confluent monolayers of HFF and 24 h later the intracellular tachyzoites were fixed and processed for immunofluorescence assay. Stable *T. gondii* transformants were obtained by two selections of transfected tachyzoites in DMEM supplemented with 10 mM Hepes and 50 µg/ml phleomycin (InvivoGen) for 4 h at 37 °C or 5 µg/ml phleomycin for 10 h at 37 °C and the resistant parasites were cloned by limiting dilution. Intracellular transfected tachyzoites were fixed for 30 min in 4% paraformaldehyde in phosphate-buffered saline (PBS), permeabilized and saturated in PBS containing 0.2% Triton X-100 and 0.1 M glycine at room temperature for 15 min. Samples were blocked with 5% FCS in the same buffer and mouse monoclonal anti-HA.11 antibodies (Covance), diluted at 1:1000, were added to 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 colocalization assays, the rabbit antibodies specific to IMC1 (generous gift of Dr G Ward, University of Vermont, USA), anti-GAP45, anti-MLC1 (kindly provided by Dr D Soldati, University of Geneva, Switzerland) 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 whereas confocal imaging was performed with a LSM710 microscope (Zeiss) and a Plan Aplanachromat objective (Plan-Aplanachromat 63x/1.40 Oil DIC M27, Zeiss). The associated software (Zen 2008) enabled the adjustment of acquisition parameters. Overlay images were assembled with Image J (National Institutes of Health).

Gene Knockin for Epitope-tagged Protein Expression in *T. gondii*—The plasmid for the gene knockin of the hypothetical proteins was made with the pLIC-HA-DHFR plasmid. A DNA fragment of 2165-bp upstream of the stop codon from the TGME49_049850 genomic sequence was cloned in the PacI restriction site of the pLIC-HA-DHFR plasmid with the ligation independent cloning techniques, using the following primers: 5'-TACTTCCAATCCAATTTAATGCGAG-AAGGAACA CCAATCGGAATCTG-3' and 5'-TCCTCCACTTCCAA-TTTTAGCGCTCGAATGGGCTTCGT CGTCA-3'. For the cloning of the 4386-bp upstream of the stop codon from the TGME49_020950 genomic sequence, the following primers were used: 5'-TACTTCCAATCCAATTTAATGCA AAGTGTGGTTGTCGACATTTTCGATTTTC-3' and 5'-TCCTCCACTTCCAAATTTAGCCTCG GACTACAAGCCAGT-GAATAC-3'. For the cloning of the 2700-bp upstream of the stop codon from the TGME49_058060 genomic sequence, the primers used were: 5'-TACTTCCAATCCAAT TTAATGCGTGCGCGCCTT-TTTCAGC-3' and 5'-TCCTCCACTTCCAAATTTAGCTTTGCTTTCG-ACCAAGTCCA-3'. The plasmid corresponding to TGME49_049850 was digested with *Bss*HIII restriction enzyme whereas TGME49_020950 plasmid was cut with *Bss*HIII, and TGME49_058060 plasmid with *Bst*BI. Ten micrograms of each digested plasmid were transfected in 5 × 10⁶ tachyzoites from Ku80 strain and stable parasites were obtained after a selection with 1 µM pyrimethamine. Emerging resistant parasites were loaded onto a new confluent monolayer of HFF cells without pyrimethamine and cloned by limiting dilution.

Affinity Purification Using Anti-FLAG Chromatography—For pilot experiments, immunopurification of the different TgGAP50-HAFLAG tagged proteins was performed using anti-FLAG M2 Affinity beads (Sigma-Aldrich) and 10⁸ stable transformants were lysed in Tris-buffered saline (TBS) (50 mM Tris.HCl, pH 7.4 containing 1% Triton X-100, 150 mM NaCl, 1 mM EDTA) for 1h at 4 °C. After centrifugation, lysates were cleared at 14,000 rpm at 30 min and incubated overnight

at 4 °C with 50 µl of anti-FLAG M2 beads under agitation. After five washes in TBS, bound proteins were eluted with 2X SDS Laemmli buffer and subjected to SDS-PAGE followed by Western blots. A large scale purification of TgGAP50 partners was performed with lysate from 5 × 10¹⁰ stable transformants expressing the wild-type TgGAP50-HAFLAG and incubated with 1 ml of anti-FLAG beads. After five washes with TBS, the bound proteins were eluted with 0.1 M glycine.HCl pH 2.5 and neutralized with 0.1 M Tris.HCl pH 9.4. Following SDS-PAGE and Coomassie blue staining, the band corresponding to GAP50-HAFLAG was excised and processed for glycomic and proteomic analyses. Comparative proteomic analyses were also performed using lysates from equal numbers (10¹⁰) of parasites from stable wild type expressing triple *N*-glycosylated GAP50-HAFLAG tachyzoites and completely *N*-deglycosylated GAP50-HAFLAG parasites. In this case, the affinity chromatography was performed under higher salt stringency using TBS containing 0.5 M NaCl and bound proteins were also eluted with 0.1 M glycine.HCl pH 2.5 and neutralized with 0.1 M Tris.HCl pH 9.4. The eluted proteins were dialyzed and concentrated before SDS-PAGE analysis followed by Coomassie blue staining.

Protein Identification and *N*-glycopeptide Analysis by Mass Spectrometry—Reduction, alkylation and tryptic in-gel digestion were performed as previously described (22). For *N*-glycopeptide analysis, the band corresponding to TgGAP50-HAFLAG was also treated overnight with 1.25 µg of chymotrypsin (Roche, Mannheim, Germany) at room temperature. Both digests were analyzed by C18 reversed phase nanoHPLC on a nanoACQUITY Ultra-Performance-LC system coupled to a Q-TOF maXis (Bruker Daltonics, Bremen, Germany) mass spectrometer. The complete system was fully controlled by Hystar 3.2 (Bruker Daltonics). For tandem MS experiments, the system was operated in the data-dependant mode with 3 tandem MS (MS/MS) scans.

For *N*-glycopeptides analysis, oxonium ions for glycopeptide detection were generated using alternate collision energy in the collision cell (10/28 eV). Glycopeptides were highlighted by searching MS data for the presence of the 204.09 *m/z* HexNAc and the 366.14 *m/z* Hex(HexNAc) and confirmed by fragment ions generated by CID from liquid chromatography (LC)-MS/MS data. The glycopeptide sequences were validated by manual assignment of peptide fragment ions as previously described (24). Mass data collected during nano-LC-MS/MS analysis were processed, converted into .mgf files using Data Analysis 4.0 software (Bruker Daltonics). MS and MS/MS data were analyzed using the Mascot 2.2.0 algorithm (Matrix Science, London, UK) and OMSSA (Open Mass Spectrometry Search Algorithm) (25). For peptide identification, an in-house generated protein database composed of protein sequences of *Toxoplasma gondii* downloaded from NCBI (February 17, 2010) concatenated with reversed copies of all sequences (total 51258 entries) was used. Searches were performed with a mass tolerance of 5 ppm in MS mode and 0.02 Da in MS/MS mode. Enzyme specified was trypsin and one missed cleavage per peptide was allowed. Variable modifications were taken into account such as carbamidomethylation of cysteine residues and oxidation of methionine residues. Neither protein molecular weight nor isoelectric point constrains were applied. The Mascot and OMSSA results were loaded into the Scaffold software (Proteome Software, Portland, OR). To minimize false positive identifications, results were subjected to very stringent filtering criteria as follows: 1) for the identification of proteins with two peptides or more, a Mascot ion score must be higher than 20 and had to be minimum 10 below the 95% Mascot significance threshold ("Identity score") or an OMSSA E-value above -log(e3) was required; 2) in the case of single peptide hits, the Mascot score of the unique peptide must be higher than 25 and had a minimal "difference score" of 5 than the "Identity score" and the OMSSA E-value must be greater than 6.

The target-decoy database search allowed us to control and estimate the false positive identification rate of our study (26, 27). Thus, the final catalogue of proteins presents an estimated false positive rate below 0.5%.

Determination of *N*-glycan Structures Using Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry (MALDI-TOF-MS)—The Coomassie-stained bands were excised from the gel, destained and the protein was then reduced and alkylated with iodoacetamide as previously described (28). After PNGase F digestion, *N*-glycans were eluted from the gel pieces and desalted on minicolumns with 10 mg of nonporous graphitized carbon followed by mass spectrometry. MALDI-TOF MS experiments were carried out on Voyager Elite DE-STR Pro instrument (PersSeptive Biosystem, Framingham, MA) equipped with a pulsed nitrogen laser (337 nm) and a gridless delayed extraction ion source. The spectrometer was operated in positive reflectron mode by delayed extraction with an accelerating voltage of 20 kV and a pulse delay time of 200 nsec and a grid voltage of 66%. All spectra shown represent accumulated spectra obtained by 500 laser shots. Sample was prepared by mixing 1 μ l aliquot (5 to 10 picomoles) with 1 μ l of matrix solution, on the MALDI sample plate. The matrix solution was prepared by saturating methanol-water (1:1) with DHB (10 mg/ml). The on-target α -mannosidase digestion was performed on desalted *N*-glycans dissolved with water at 5 to 10 picomoles/ μ l and incubated with Jack bean α -mannosidase (EC 3.2.1.24, Sigma Chemicals). One μ l of the sample was spotted on the MALDI sample plate and 1 μ l of 10 mM ammonium acetate (pH 4.5), and 35 mU of α -mannosidase were added. The MALDI plate was then placed in a crystallization beaker containing water, at 37 °C for 6 h. The enzymatic reaction was terminated by addition of 1 μ l of a matrix solution (DHB) and the sample was analyzed by MALDI-TOF-MS. Ribonuclease B (Sigma Chemicals, St. Louis, MO) was used as control *N*-glycoprotein.

Biochemical Analyses of Glycoform Heterogeneity Using LC-MS—Five microliters of chymotrypsin GAP50 digest were analyzed by nanoLC-MS using the nanoACQUITY Ultra-Performance-LC system (UPLC, Waters, Milford, MA) coupled to a SYNAPT High Definition Mass Spectrometry quadrupole time-of-flight tandem mass spectrometer (Waters, Milford, MA) equipped with a nano-electrospray source. Extraction of the current ions corresponding to the different glycoforms, identified by preliminary MS/MS experiments, was performed (m/z 1150.0, 1231.0, 1245.5, 1326.5, 1407.5, 1261.0, 1342.1, 1423.1, \pm 0.2 Da) and for each glycosylation site, MS spectra of glycopeptides elution zone were combined.

In Vitro Gliding Assay, Host Cell Invasion and Proliferation Assay—The wild-type parasites, ectopic GAP50-HAFLAG expressers and completely deglycosylated GAP50-HAFLAG parasites, treated or untreated with 5–10 μ g/ml of tunicamycin, were released 30 h post-infection and filtered in DMEM containing 10 mM Hepes and 1 mM EGTA. After centrifugation at 1000 rpm for 10 min, the parasite pellet was resuspended in the same medium and the parasites were allowed to move for 15 min at 37 °C on slides previously coated overnight with 50% FCS prepared in PBS. Slides were immediately fixed with 4% PFA and processed for immunofluorescence assays using antibodies specific to the major surface antigen SAG1 in PBS. Trails of gliding parasites were visualized with a Zeiss Axiophot microscope. These treated or untreated wild type, ectopic expresser and completely deglycosylated GAP50 expressing parasites were also resuspended with DMEM medium containing 10% FCS, 2 mM *L*-glutamine and 10 000 U/ml Penicillin-Streptomycin, loaded onto confluent monolayer HFF cells and grown for 24 h at 37 °C. The intracellular parasites were fixed and stained with monoclonal antibody specific to SAG1 for immunofluorescence as above and intracellular tachyzoites were counted under Zeiss Axiophot microscopy.

RESULTS

Ectopic Expression of the Integral Membrane Anchor GAP50 of the *T. gondii* Gliding Motor—We and others have previously shown that GAP50, the membrane anchor for the gliding motor complex in *T. gondii*, contains potential *N*-linked oligosaccharides (9, 21, and 22). We have previously investigated the *N*-glycosylation status of *T. gondii* by exploring the presence of *N*-glycosylated peptides isolated from trypsin-digested total detergent extract proteins followed by extensive glycomic and proteomic analyses. This led to the identification of one *N*-glycopeptide of GAP50, which carries immature oligomannosidic glycans (22). To gain more insight into protein *N*-glycosylation functions in *T. gondii*, we decided to determine the fine structures of *N*-glycan chains of all three potential *N*-glycosylation sites: Asn residues at position 101, 136, and 228. Toward this goal, we engineered stable *T. gondii* transgenic parasites, ectopically expressing HAFLAG-tagged GAP50 protein, using both endogenous GAP50 and heterologous dense granule *GRA1* gene promoters. We chose the heterologous *GRA1* promoter because the weak expression driven by GAP50 promoter yields an insufficient amount of protein for further classical glycomic (gas chromatography coupled to mass spectrometry) and proteomic analyses. To ascertain if, like the wild-type GAP50 protein, the ectopically expressed GAP50-HAFLAG protein (under *GRA1* promoter control) localizes in the IMC, we first investigated the localization of GAP50-HAFLAG protein using specific anti-HA antibodies and indirect immunofluorescence visualized by confocal microscopy. This confirmed that the ectopically expressed GAP50-HAFLAG protein (green) was strongly detected in the inner membrane complex, which was stained with monoclonal antibodies specific to IMC1 (red), a major component of the subpellicular network that tightly apposes the cytoplasmic face of the IMC (29) ([supplemental Fig. S1, panel A](#)). The GAP50-HAFLAG protein was shown to be in association with the newly formed IMC membranes of daughter parasites ([supplemental Fig. S1, panel A](#)). Proteomic analyses using a large scale affinity purification of the HAFLAG-tagged GAP50 and its binding partners under a low salt stringency (see details under “Experimental Procedures”) showed after separation by SDS-PAGE and Coomassie blue staining ([supplemental Fig. S1, panel B](#)), all major known gliding partners, such as myosin A (MyoA), myosin-light chain (MLC1), glideosome associated proteins GAP45 and GAP50 ([supplemental Table S1](#)). Under these low salt stringency conditions, other proteins identified by proteomics were enolase, fructose-1, 6-bisphosphatase, glyceraldehyde-3-phosphate dehydrogenase, hexokinase, lactate dehydrogenase, phosphofructokinase and numerous hypothetical proteins ([supplemental Table S1](#) and [supplemental Fig. S1, panel B](#)). Although, the pull down of the glycolytic enzymes is also consistent with their dual role in both glycolysis and specific relocation from the cytoplasm to IMC before host cell egress

TABLE I

Comparison of proteins that bind to N-glycosylated and unglycosylated GAP50-HAFLAG using nanoLC-MS/MS. The affinity purification was performed under high stringency conditions and protein identity was determined using a protein database of *T. gondii* downloaded from NCBI nr databases. The corresponding identity in *T. gondii* databases (<http://www.toxodb.org>) and the different genome sequences of *T. gondii* strains RH, GT1, VEG, and ME49 used for protein identification are shown. We identified three novel putative GAP50 binding proteins including GAP40, which has been independently reported and validated as a novel glideosome component that interacts with GAP45 (34)

Protein name and function	Protein accession numbers	Protein molecular weight (Da)	N-glycosylated GAP50		Unglycosylated GAP50		Number of potential N-glycosylation site	Number of transmembrane domains
			Number of unique peptides/Percentage sequence coverage	Percentage sequence coverage	Number of unique peptides/Percentage sequence coverage	Percentage sequence coverage		
Glideosome and regulatory proteins								
Actin	gii1703160	41891	3	12%				
Actin depolymerizing factor, putative ^c	gii221481151	12904	2	21%				
Gliding-associated protein 45, GAP45 ^a	gii30840677	27316	2	13%				
Membrane anchor for myosin XIV precursor, GAP50 ^a	gii46948064	46590	18	47%	11	36%	3	1
Myosin A, putative ^c	gii221483418	93319	33	56%	11	15%		
Myosin light chain TgMLC1 ^a	gii16580144	24107	8	65%	2	11%		
Glideosome multiprotein complex								
Actin	gii1703160	41891	7	24%	8	27%		
Gliding-associated protein 45, GAP45 ^a	gii30840677	27316	3	20%				
Membrane anchor for myosin XIV precursor, GAP50 ^a	gii46948064	46590	11	37%	8	26%	3	1
Myosin A, putative ^c	gii221483418	93319	51	73%	15	22%		
Myosin light chain TgMLC1 ^a	gii16580144	24107	4	25%				
Unknown functions								
Conserved hypothetical protein ^{c,d} TGME49_020950	gii221480609	47891	4	11%	5	12%	2	1
Conserved hypothetical protein ^c TGME49_058060	gii221487120	23334	2	10%				
Hypothetical protein, conserved ^e TGME49_049850	gii221484942	43063	2	8%			1	9

^a *Toxoplasma gondii*.

^b *Toxoplasma gondii* RH.

^c *Toxoplasma gondii* GT1.

^d *Toxoplasma gondii* VEG.

^e *Toxoplasma gondii* ME49.

to presumably provide ATP to the actin-myosin motor, as reported by Pomel *et al.* (18), most of these enzymes and proteins were not identified when higher salt stringency conditions were used (Table I). This suggests that these abundant glycolytic enzymes and hypothetical proteins could also be nonspecific contaminants. However, none of these glideosome components, glycolytic enzymes and hypothetical proteins was identified when nonimmune purified IgGs and unrelated IgGs specific to *T. gondii* sortilin-like receptor (22 and our unpublished data) were used as negative controls for the pull down experiments. In these cases, only few cytoplasmic organelles such dense granules, micronemes, rhoptries, and ribosomal proteins were detected and these nonspecific and irrelevant components were therefore removed from the proteomic data. In addition, Western blots performed with the affinity purified materials and different specific polyclonal antibodies confirmed the presence of all known glideosome proteins (Fig. 6, lanes GAP50 WT), thereby demonstrating that the ectopic HAFLAG-GAP50 fusion protein was full-length, had been properly tagged and retained its ability to bind to gliding partners.

The Integral Membrane Anchor GAP50 of T. gondii Gliding Motor Contains Unusual N-glycan Structures—The GAP50

protein was excised and purified from SDS-PAGE described in supplemental Fig. S1, panel B, and the nature of its N-glycans was determined after reduction, carboxyamido-ethylation, digestion with PNGase F, purification on a non-porous graphitized carbon column followed by MALDI-TOF-MS analyses. The predominant molecular ion corresponding to Hex₆HexNAc₂ + Na⁺ (*m/z* 1419) and two additional weaker ions, Hex₇HexNAc₂ + Na⁺ (*m/z* 1581), and Hex₈HexNAc₂ + Na⁺ (*m/z* 1743) were identified (Fig. 1A). To discriminate the oligomannosylated type N-glycans from the glucose-containing truncated glycans, the glycans were submitted to on-plate α -mannosidase digestion followed by MALDI-TOF-MS analysis (Fig. 1B). After this treatment, two new molecular ions were observed at *m/z* 771 (Hex₂HexNAc₂) and 933 (Hex₃HexNAc₂). The intensity of ion at *m/z* 1419 (Hex₆HexNAc₂) was reduced, and this strongly suggests the existence of an oligomannosylated type N-glycans with six mannose residues. However, a significant α -mannose resistant signal was also present (see enlarged insert in Fig. 1B). These data suggest that other isomeric structures, which probably correspond to terminal glucose-containing N-glycans that are unaffected by α -mannosidase, are also present

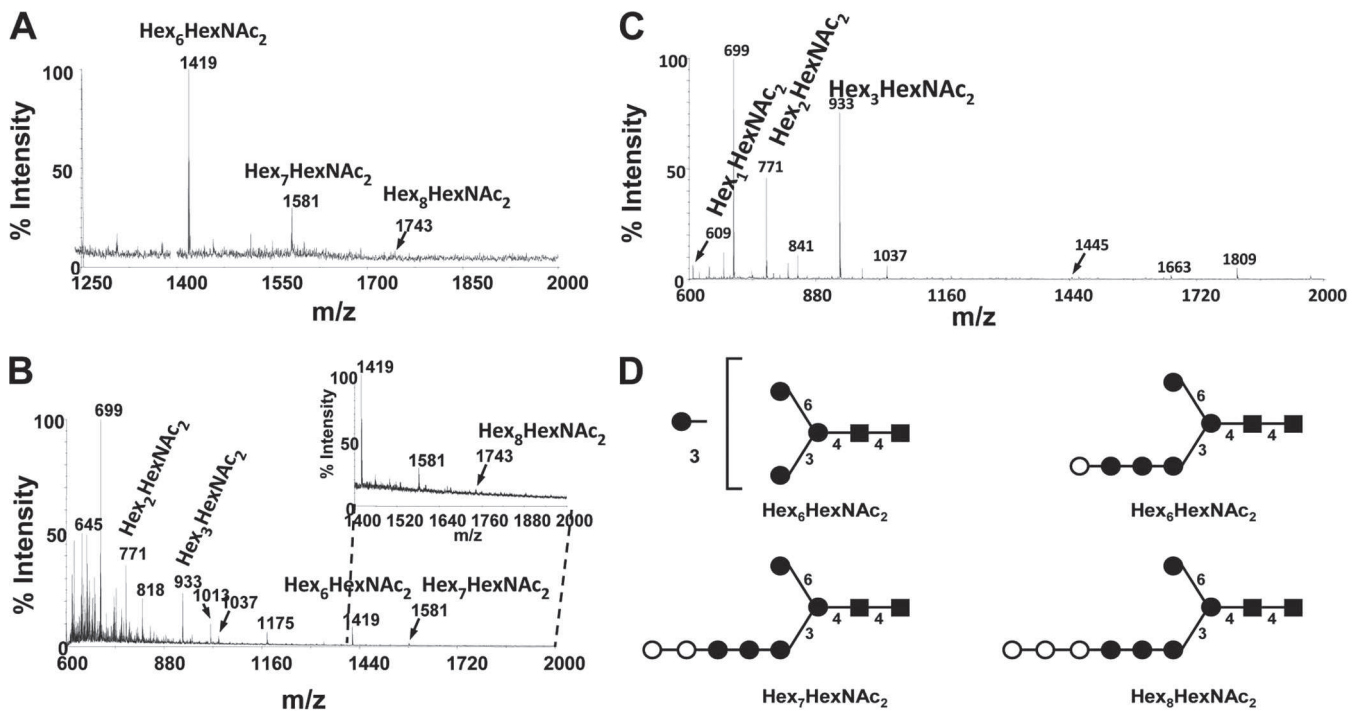


FIG. 1. *T. gondii* GAP50 contains unusual oligomannose-rich N-glycan with terminal glucose residues. MALDI-TOF-MS spectra of the in-gel PNGase F released N-glycans from *T. gondii* GAP50-HAFLAG before (A) and after α -mannosidase digestion (B). Purified GAP50 (as described in supplemental Fig. S1B) was excised, de-stained and the protein was reduced and alkylated with iodoacetamide. After in-gel deglycosylation by treatment with PNGase F, the extracted glycans were purified on nonporous graphitized carbon solid phase extraction cartridge and characterized by MALDI-TOF-MS analysis, as $[M+Na]^+$ pseudomolecular ions, before (A) and after on-plate α -mannosidase digestion (B). Mass spectra were recorded in positive ion reflective mode. C, MALDI-TOF-MS spectrum of the PNGase F released N-glycans from ribonuclease B after on-plate α -mannosidase digestion was used as positive control. D, Proposed structures of N-glycans on *T. gondii* GAP50. The data do not exclude the presence of N-glycans with seven or eight mannose residues. Symbols were as follows: ■, N-acetylglucosamine; ●, mannose; ○, glucose.

on GAP50. We also checked whether α -mannose digestion under our experimental conditions is optimal on a positive control using on-plate α -mannosidase digestion applied to free N-glycans of bovine ribonuclease B that contains well-characterized oligomannosylated N-glycans composed of Hex₅HexNAc₂, Hex₆HexNAc₂, Hex₇HexNAc₂, Hex₈HexNAc₂, and Hex₉HexNAc₂. Similarly, two major molecular ions were also observed at *m/z* 771 (Hex₂HexNAc₂) and 933 (Hex₃HexNAc₂) and a minor one at *m/z* 609 (Hex₁HexNAc₂) (Fig. 1C). The ions at *m/z* 1257 (Hex₅HexNAc₂), 1419 (Hex₆HexNAc₂), 1581 (Hex₇HexNAc₂), 1743 (Hex₈HexNAc₂), and 1905 (Hex₉HexNAc₂) disappeared after the α -mannosidase treatment, which demonstrates that these oligomannosylated N-glycans of ribonuclease B were efficiently and similarly digested by α -mannosidase as for GAP50. We therefore conclude that biochemical properties of N-glycans at *m/z* 1581 (Hex₇HexNAc₂) and 1743 (Hex₈HexNAc₂), shown to be unaffected by α -mannosidase, are consistent with the assignment of terminal glucose-containing N-oligomannosidic chains to these ions (Fig. 1D). In agreement with this statement, no shift in electrophoretic migration was observed after SDS-PAGE and Western blotting when GAP50 was incubated with endo H, an exoglycosidase known to specifically cleave

off oligomannosylated glycans from N-glycoproteins, confirming the occupancy of GAP50 N-glycans by terminal glucose or other sugars (21). In contrast, we demonstrated that GAP50 was sensitive to PNGase F, resulting in a mobility shift (Fig. 5C). Thus, we propose that N-glycan structures composed of oligomannosylated type N-glycans (Hex₆HexNAc₂) and glucose-containing truncated glycans (Hex₆₋₈HexNAc₂) are both linked to *T. gondii* GAP50 protein even though we cannot rule out the presence of additional minor N-glycans with seven mannose (*m/z* 1581) or eight mannose residues (*m/z* 1743) (Fig. 1D).

Positioning N-glycan Structures on Asn-X-Ser/Thr Sequons of *T. gondii* GAP50 and Glycoform Heterogeneity—To determine which N-glycan structure is precisely carried by each individual N-glycosylation site in the GAP50 glycoprotein, the in-gel purified protein was separately subjected to trypsin and chymotrypsin digestions. Both digests were first analyzed by LC-MS under conditions which generate glycan-specific fragment ions (*m/z* 204.09 from HexNAc⁺ and *m/z* 366.14 from HexHexNAc⁺) indicating the location of the glycopeptides in the chromatogram. Based on the glycopeptide fragmentation pattern, high-level confidence assignment for each glycosylated site was identified by LC-MS/MS (Fig. 2, panels A–C).

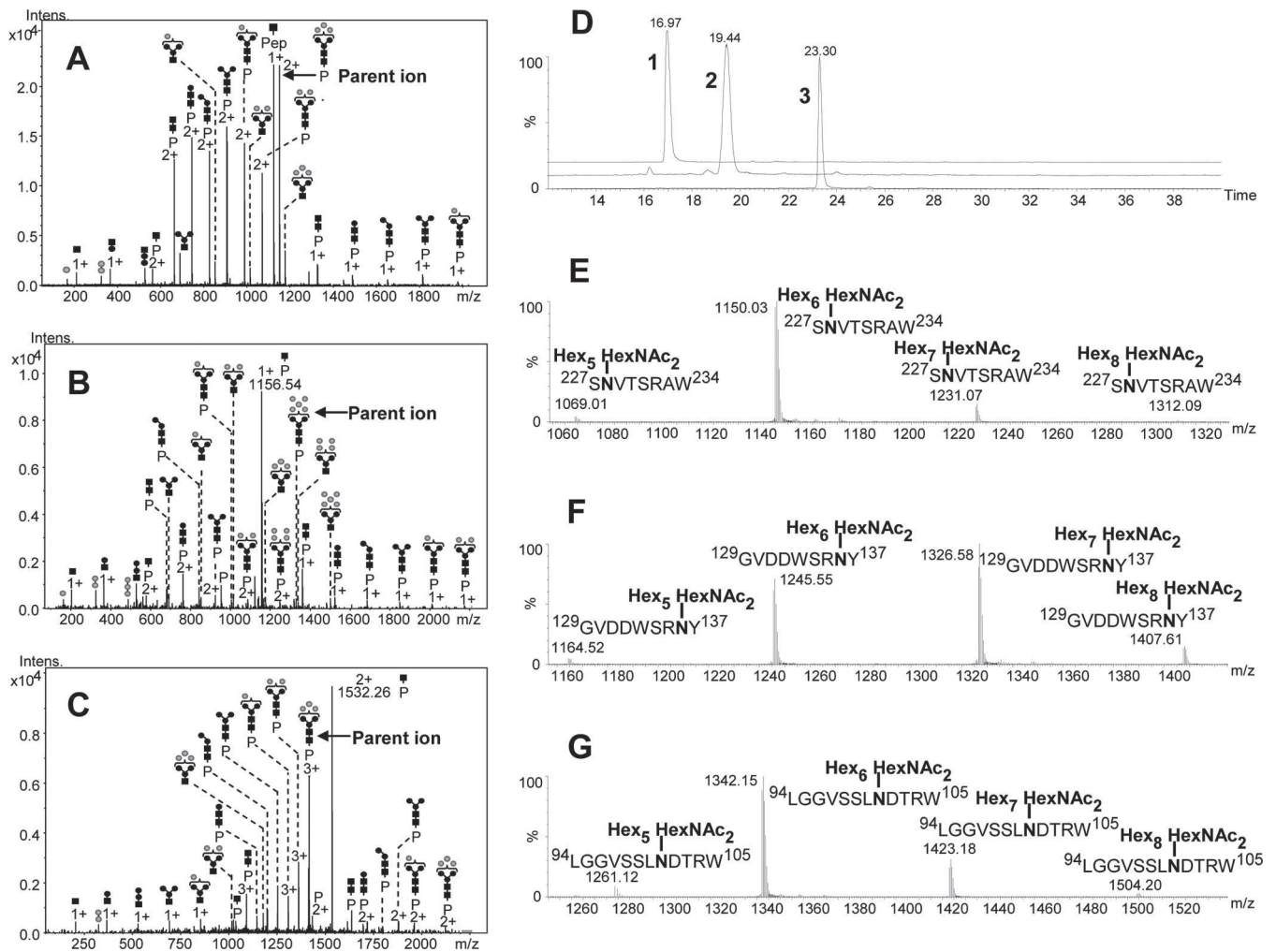


FIG. 2. Deciphering N-glycan structures and positioning on GAP50 protein. A, GAP50 glycopeptides were identified using nanoLC-MS/MS. The MS/MS spectrum from the parent ion $m/z = 1150.48$ corresponds to the chymotryptic peptide $^{227}\text{SNVTSRAW}^{234}$ with the N-glycosylation Hex₆(HexNAc)₂ in position Asn228. B, The MS/MS spectrum from parent ion at $m/z = 1329.03$ corresponds to the tryptic peptide $^{136}\text{NYTSEALR}^{143}$ with the N-glycosylation Hex₆(HexNAc)₂ in position Asn136. C, The MS/MS spectrum from parent ion at $m/z = 1414.31$ corresponds to the tryptic peptide $^{77}\text{VAANEHISFIASPGSNFLGGVSSLNDR}^{104}$ with the N-glycosylation Hex₆(HexNAc)₂ in position Asn101. Symbols were as follows: p, peptide; ■, N-acetylglucosamine; ●, mannose; ●, hexose. (D–G) Chymotryptic digestion of GAP50-HAFLAG profile using nanoLC-MS. D, Extracted ion chromatograms (m/z 1150.0, 1231.0 for peak 1; 1245.5, 1326.5, 1407.5 for peak 2; 1261.0, 1342.1, 1423.1, for peak 3; ± 0.2 Da) indicated the elution zone of glycopeptides assigned by MS/MS. E, Combined mass spectrum of the glycopeptides containing Asn228 eluting at 17 min (peak 1). F, Combined mass spectrum of the glycopeptides containing Asn136 eluting at 19 min (peak 2). G, Combined mass spectrum of the glycopeptides containing Asn101 eluting at 23 min (peak 3).

In the MS/MS spectrum of the doubly protonated parent ion at m/z 1150.48 (Fig. 2A), the most intense fragment ion at m/z 1123.55 (1+) corresponds to peptide sequence $^{227}\text{SNVTSRAW}^{234}$ with one HexNAc. This interpretation is based on the previous observation indicating that often the MS/MS spectrum of doubly charged ion of a glycopeptide is dominated by a fragment ion corresponding to the peptide carrying a single HexNAc (30). This allows the identification of peptide carrying the N-glycopeptide at Asn228 because only one chymotryptic peptide is expected at mass 919.46 Da. The corresponding spectrum displayed a series of single protonated ion fragments higher than molecular ion at m/z 1150.48 and a series of lower double protonated ion fragments, most

of them corresponding to successive sugar neutral losses. The other fragment ions are oxonium ions which correspond to internal glycan fragmentations. Taken together, these data demonstrate the presence of Hex₆HexNAc₂ structure at position Asn228 of GAP50. In addition, MS/MS spectra of molecular ions m/z 1329.03 (2+) and m/z 1414.31 (3+) displayed intense ion fragments respectively at m/z 1156.54 (1+) and m/z 1532.26 (2+), which represent the two other N-glycosylated sites of the tryptic peptides $^{136}\text{NYTSEALR}^{143}$ and $^{77}\text{VAANEHISFIASPGSNFLGGVSSLNDR}^{104}$ carrying one HexNAc on Asn136 and Asn101, respectively (Figs. 2B and 2C). These MS/MS spectra also revealed the structures of the N-linked glycans, Hex₈HexNAc₂ at position Asn136 and

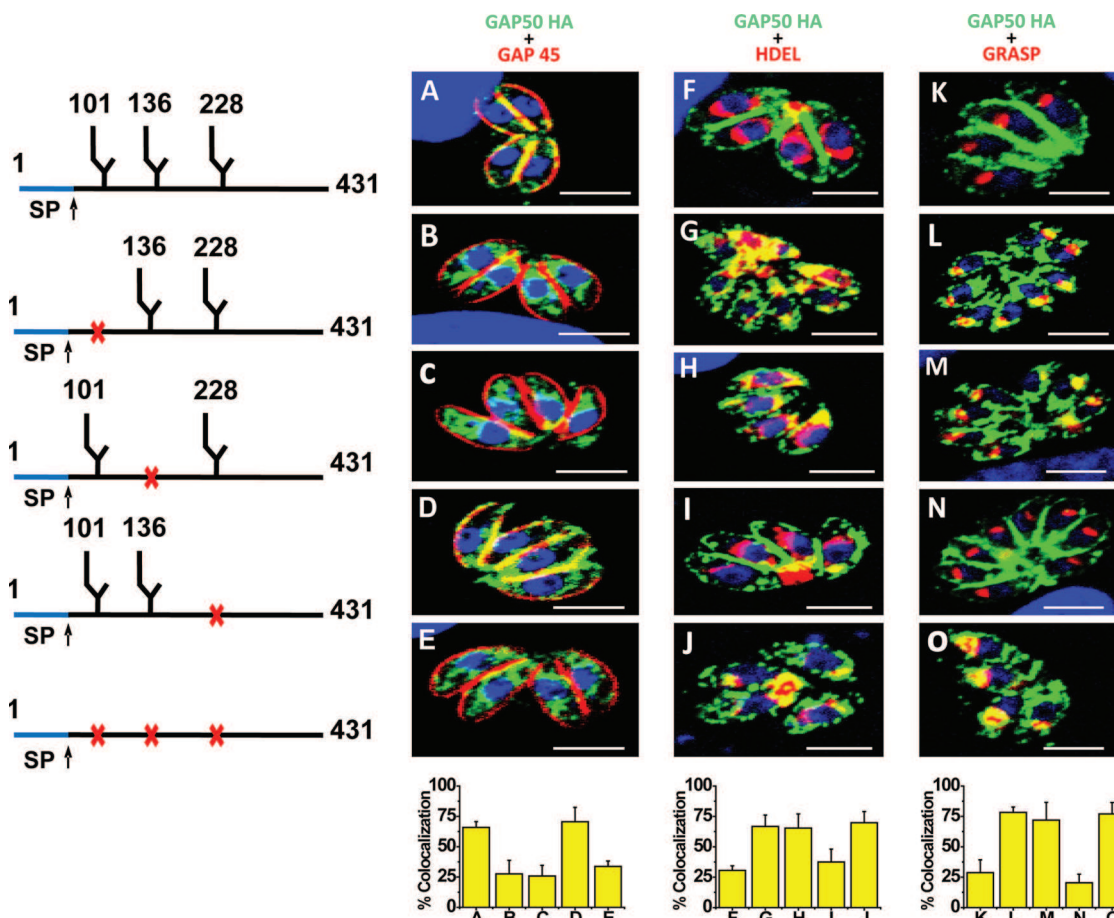


FIG. 3. Mutations of N-glycosylation sites on GAP50 cause retention of the deglycosylated protein in the endoplasmic reticulum (ER) and Golgi apparatus. Schematic representations showing the GAP50 gene sequences corresponding to the three potential N-glycosylated sites in the wild-type GAP50-HAFLAG and the different mutants used for transient transfection and IFA are shown on left. (A–E) Location of GAP50-HAFLAG, GAP50N_101-A-HAFLAG, GAP50N_136-A-HAFLAG and GAP50N_101_136_228-AAA-HAFLAG proteins (green) in intracellular tachyzoites from wild-type RH strain using anti-HA and antibodies specific to another glideosome partner GAP45 (red). (F–J) Location of GAP50-HAFLAG, GAP50N_101-A-HAFLAG, GAP50N_136-A-HAFLAG and GAP50N_101_136_228-AAA-HAFLAG proteins (green) in transgenic tachyzoites expressing the ER marker HDEL (HDEL-Red Florescent Protein). (K–O) Location of GAP50-HAFLAG, GAP50N_101-A-HAFLAG, GAP50N_136-A-HAFLAG and GAP50N_101_136_228-AAA-HAFLAG proteins (green) in transgenic tachyzoites expressing the Golgi marker GRASP (GRASP-Red Fluorescent Protein). The intracellular parasites shown in the merged pictures were transiently transfected and analyzed 24 h postinfection. Scale bar, 5 μ m. Colocalization values are measured with the ImageJ software (NIH) and expressed as histograms below each IFA column and the letters A–E, F–J, and K–O correspond to the corresponding IFA panels. Results are from two experiments each measuring the colocalization within at least four to eight vacuoles each with four or eight transfected parasites. The ratio (% colocalized area) was calculated using Red and Green Area Fraction/Red Area Fraction. Bars represent S.E. from two independent experiments ($n = 2$ and $p < 0.05$).

Hex₆HexNAc₂ at position Asn101, respectively. It should be noted that this MS-based approach led to the identification of 8 distinct glycopeptides assigned to the three glycosylation sites as follows: Hex_{5–7} HexNAc₂ (3 N-glycan chains) at position Asn101, Hex_{6–8} HexNAc₂ (three N-glycan chains) at position Asn136 and Hex_{6–7} HexNAc₂ (two N-glycan chains) at position Asn228. We showed that there was also heterogeneity of oligosaccharides at each three N-glycosylation sites. A better picture of the relative abundance of the different glycoforms present on each of the three N-glycosylated sites was given by an LC-MS experiment (Figs. 2D–2G). The N-glycan Hex₆ (HexNAc)₂ is the most predominant sugar on

Asn228 and Asn101 whereas Hex₆ (HexNAc)₂ and Hex₇ (HexNAc)₂ are present in equal amount on the N-glycosylation site Asn136. Together, these data provide insights into the structures, length, heterogeneity and topology of N-glycans present on each Asn-X-Ser/Thr sequon of *T. gondii* GAP50.

N-glycosylation is Required for Proper Targeting of GAP50 into the Inner Membrane Complex of T. gondii—Previous studies have reported that N-glycosylation inhibition using tunicamycin can affect parasite gliding (21, 22). We speculated that the effect of tunicamycin on N-glycans of GAP50 may occur either by direct inhibition of N-glycosylation syn-

thesis or through indirect and nonspecific effect, as tunicamycin-treatment of mammalian cells can also result in global decrease of cellular protein synthesis (31). In the light of these two possibilities, we first decided to determine whether unglycosylated GAP50 can be used *in vivo* for proper targeting into IMC and gliding motor assembly using reverse genetic approaches instead of biochemical assays with N-glycosylation inhibitors. To address this issue, *T. gondii* GAP50 mutants lacking one, two or three of the normally occupied N-glycosylation sites were constructed by site directed mutagenesis (replacement of Asn of the sequon to Ala) and the corresponding expression vectors, under the control of homologous GAP50 promoter, were transfected into the parasites. The localization of mutated GAP50 proteins tagged with HAFLAG was analyzed by indirect immunofluorescence and confocal microscopy. A schematic depicting the position and number of sites mutated in GAP50 is shown on the left corner of Fig. 3. Under the control of GAP50 promoter, the transgenic wild-type GAP50-HAFLAG protein and GAP45 partially colocalized in the IMC, as expected (Fig. 3A). It should be noticed that the signal of the ectopic GAP50-HAFLAG (green), which is transiently expressed after transfection appeared only in few areas of the inner membrane complex that are estimated to cover ~70% (histogram A, yellow) of areas occupied by the endogenous GAP45 signal that is used as a marker delineating the whole IMC surface (red, Fig. 3A). These data also confirm the previous studies that indicate that the endogenous GAP50 has lower turnover rates (10). Consistent with these results, the stably expressing GAP50-HAFLAG protein showed fluorescence signal that occupied the whole IMC surface (supplemental Fig. S1, panel A). The mutation of the first Asn at the position 101 in GAP50, however, revealed a confocal fluorescence signal restricted to the parasite cytoplasm with a significant decrease of co-localization (about 25%) with GAP45 (Fig. 3B, histogram B). It is worth noting that despite the presence of the two other sequons that were demonstrated by glycomic approaches and enzymatic digestions to be readily occupied by N-glycans (Figs. 2 and 5B), a single replacement of the first N-glycosylation site Asn101 by Ala prevents an efficient targeting of GAP50 into the IMC. Moreover, the single mutation of the second N-glycosylation site Asn136 behaves similar to the first Asn101 mutant, with the mutated GAP50 protein detected mostly in the parasite cytoplasm (Fig. 3C). About 25% of co-localization with GAP45 (Fig. 3C, histogram C) was observed despite the presence of the two other sites Asn101 and Asn228 occupied by N-glycans (Fig. 5B). In contrast, the mutation of the third N-glycosylation site Asn228 reveals no changes in the classical IMC localization, about 75% of this mutated GAP50 colocalized with GAP45 in the IMC (Fig. 3D, histogram D), similarly to the wild-type GAP50 (Fig. 3A, histogram A). The mutants lacking all three N-glycosylated sites show behavior similar to the single first or second Asn-x-Ser/Thr sequon with a stronger fluorescence signal surrounding the nucleus and more

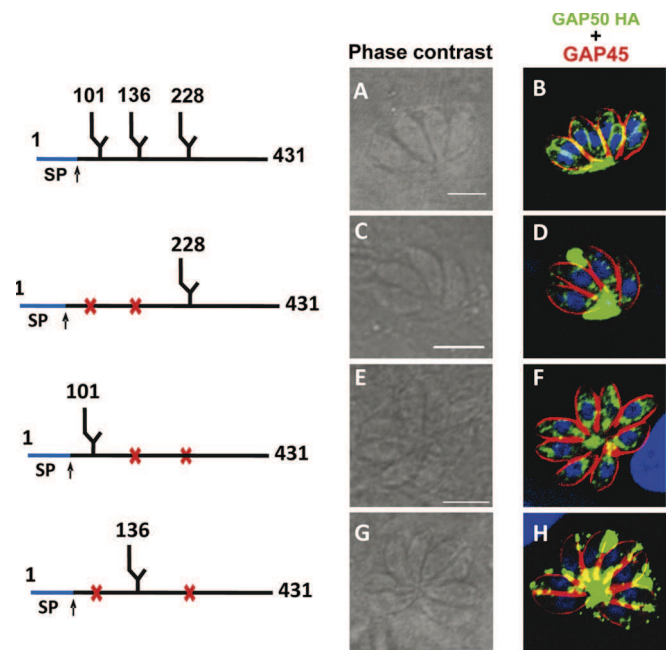


Fig. 4. Ectopic expression of double mutant GAP50 prevents targeting of the deglycosylated GAP50 protein into *T. gondii* inner membrane complex. Schematic representations showing the GAP50 gene sequences corresponding to the three potential N-glycosylated sites in the wild-type GAP50-HAFLAG and the different mutants used for transient transfection and IFA are shown on left. Location of GAP50-HAFLAG (A–B), GAP50N_101_136-AA-HAFLAG (C–D), GAP50N_136_228-HAFLAG (E–F) and GAP50N_101_228-AA-HAFLAG (G–H) proteins in intracellular tachyzoites from wild-type RH strain using anti-HA and anti-GAP45. These intracellular parasites were transiently transfected and analyzed 24 h post-infection. Scale bar, 5 μ m.

intra-cytoplasmic vesicles (Fig. 3E). In this case, about 25–30% of deglycosylated GAP50 colocalized with GAP45 (Fig. 3E, histogram E). Taken together, these results strongly suggest that mutation of either first, second or all three N-glycosylation sites may cause the deglycosylated GAP50 protein to remain in the perinuclear compartment defining *T. gondii* endoplasmic reticulum and Golgi apparatus.

The Deglycosylation of GAP50 Impaired Its Intracellular Trafficking in T. gondii—We next focused on the intracellular trafficking of deglycosylated GAP50 to know if the different mutated GAP50 proteins reside in specific sub-cellular compartments of *T. gondii*. Therefore, we transfected the expression vectors corresponding to wild type and each deglycosylated GAP50 protein into the transgenic parasite line, which stably expressed a marker of the endoplasmic reticulum, named HDEL, tagged to red fluorescent protein (RFP) (32). The confocal images demonstrate that mutants at the first (Fig. 3G), second (Fig. 3H), and all three (Fig. 3J) N-glycosylated sites displayed fluorescence signals (green) that colocalized with HDEL-RFP (red), suggesting that significant levels of deglycosylated GAP50 (about 70% of colocalization, histograms G, H and J) are retained in the parasite ER. In

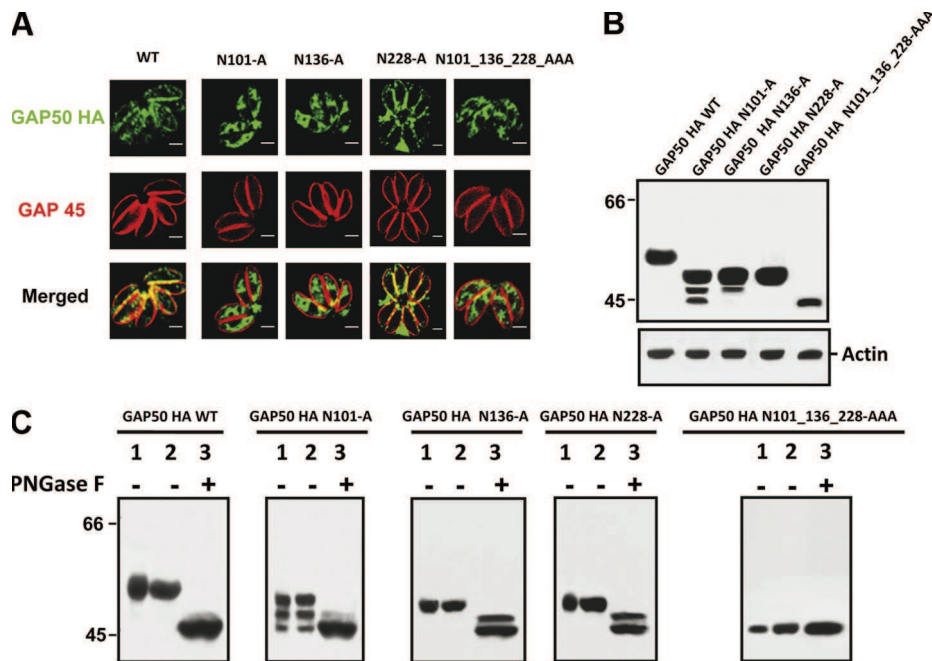


FIG. 5. Biochemical analysis of N-glycan site occupancy of wild type and mutated TgGAP50. A, Both mislocalization of deglycosylated GAP50N_101-A-HAFLAG, GAP50N_136-A-HAFLAG and GAP50N_101_136_228-AAA-HAFLAG mutants and classical location of wild-type GAP50-HAFLAG and deglycosylated GAP50N_228-A-HAFLAG in the IMC were confirmed in stably transformed tachyzoites. B, Western blots of lysates from the stable parasites expressing wild-type GAP50-HAFLAG (GAP50 WT) or mutants lacking one potential N-glycosylation site (GAP50N_101-A, GAP50N_136-A, GAP50N_228-A) or all three N-glycosylation site (GAP50N_101_136_228-AAA) proteins. The samples were reduced before electrophoresis. The absence of one or all three N-glycans was confirmed by the mobility shift of partial or fully deglycosylated GAP50 relative to wild-type GAP50-HAFLAG. The housekeeping actin was used as a loading control. C, Analysis of N-glycan site occupancy in wild-type GAP50-HAFLAG (GAP50 WT) or mutants lacking one potential N-glycosylation site (GAP50N_101-A, GAP50N_136-A, GAP50N_228-A) or all three N-glycosylation site (GAP50N_101_136_228-AAA) after PNGase F digestion. Tachyzoite lysates were treated with PNGase F (lanes 3) or with buffer alone (lanes 2) at 37 °C for 4 h and compared with untreated control (lanes 1). The samples were reduced before electrophoresis followed by Western blots as above. After PNGase F treatments, all single deglycosylated mutants (GAP50N_101-A, GAP50N_136-A, GAP50N_228-A) collapse into one major and faster mobility band that co-migrates with the triple deglycosylated mutant (GAP50N101_136_228-AAA), itself insensitive to PNGase F, as expected (right panel, lane 3).

contrast, the wild-type GAP50 protein (Fig. 3F) and the third N-glycan mutant (Fig. 3I) were mostly found in the IMC, thereby only weakly detectable in the ER (histograms F and I). In addition, when GAP50 protein mutants at the first, second, and all three N-glycosylated sites were transfected in the transgenic parasites, which stably expressed in *T. gondii* Golgi apparatus, the RFP-tagged Golgi marker GRASP (33, red signal), the fluorescence signal of these deglycosylated GAP50 (green) significantly colocalized (about 75% of colocalization, histograms L, M and O) with GRASP protein (Figs. 3L, 3M, and 3O, respectively). In the case of the wild type and third N-glycan GAP50 mutant, again the deglycosylated and wild-type proteins, which are correctly transported to the IMC cannot be detected in the Golgi apparatus (Fig. 3K and 3N). In this case, only 15–25% of colocalization was observed (histograms K and N). In addition, we also created three independent double mutants GAP50N101_N136AA (Fig. 4C and 4D), GAP50N136_N228AA (Fig. 4E and 4F) and GAP50N101_N228AA (Fig. 4G and 4H), which should lead to simultaneous deglycosylation of GAP50 protein at two Asn sites. Consistently, these double mutations prevent the trans-

port and insertion of their respective double mutated GAP50 protein into the IMC (Fig. 4), and these double deglycosylated transgenic GAP50 retained in the ER and Golgi apparatus (supplemental Fig. S2), similar to the single deglycosylated GAP50 (Fig. 3). Taken together, our data strongly suggest that the two first N-glycosylation sites are more critical for proper trafficking of GAP50 than the third N-glycosylated site.

N-glycosylation is Critical for Association of GAP50 with Other Components of T. gondii Glideosome—To investigate whether the lack of N-glycans on GAP50 might have a direct consequence in the formation of multi-complex proteins assembled as glideosome components, stable transgenic parasites expressing GAP50 mutants and wild-type proteins tagged to HAFLAG were generated (Fig. 5A). We confirmed that in the stable transgenic parasite lines, which were cloned, the deglycosylated GAP50 at the first (GAP50N_101-A), the second (GAP50N_136-A) and at all three N-glycosylated sites GAP50N_101_136_228-AAA, the mutated GAP50 was not localized in the IMC whereas GAP50 proteins from stably transformed wild-type HAFLAG-GAP50 and mutant at third N-glycosylation site GAP50N_228-A-HAFLAG are readily

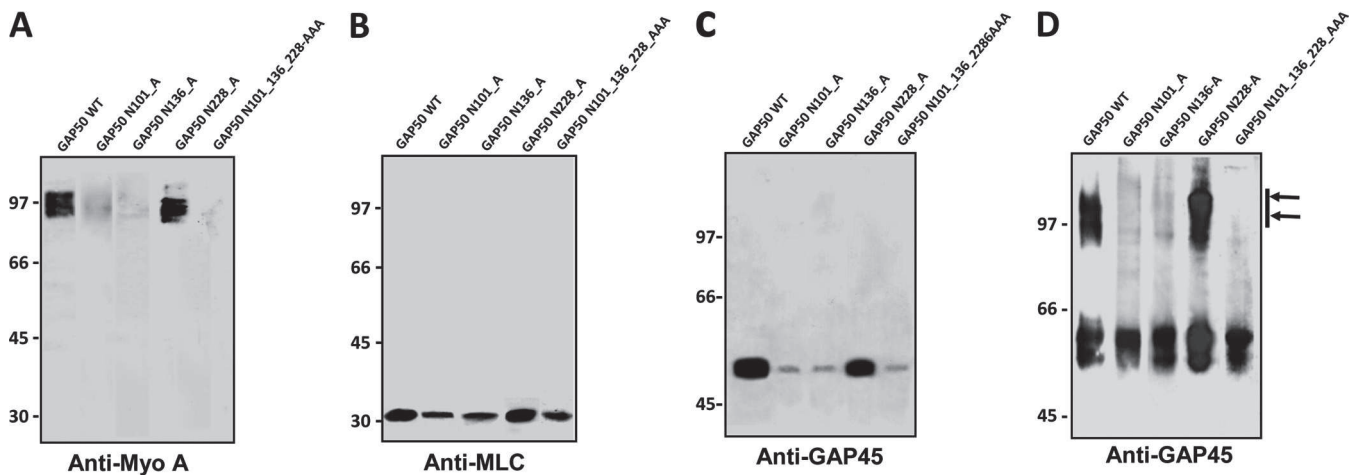


FIG. 6. Deglycosylation of GAP50 prevents its association with other components of *T. gondii* glideosome. (A–D) Parasites expressing wild-type GAP50-HAFLAG (GAP50 WT) or mutants lacking one potential *N*-glycosylation site (GAP50N_101-A, GAP50N_136-A, GAP50N_228-A) or all three *N*-glycosylation sites (GAP50N_101_136_228-AAA) were subjected to immunoprecipitation (IP) with anti-FLAG beads. Under non-reducing electrophoretic conditions, these Western blots (revealed with polyclonal antibodies specific to MyoA and MLC1) showed that the mutations of the first two *N*-glycan sites reduced or completely abolished MyoA (A) and MLC1 (B) pull-down. C, Cross-linking parasites expressing wild-type GAP50-HAFLAG (GAP50 WT) or mutants lacking one potential *N*-glycosylation site (GAP50N_101-A, GAP50N_136-A, GAP50N_228-A) or all three *N*-glycosylation sites (GAP50N_101_136_228-AAA) with DSP followed by IP and Western blots under reduced electrophoretic conditions demonstrated that the partial deglycosylated GAP50 in GAP50N_101-A, GAP50N_136-A and in the complete unglycosylated GAP50N_101_136_228-AAA transgenic prevent interactions with key glideosome partners *in vivo*. Note that the lack of *N*-glycan in the third site Asn (GAP50N_228-A) has no consequence on partner binding, instead this partial deglycosylated GAP50 behaves similar to wild-type GAP50. D, Western blots of parasites expressing wild-type GAP50-HAFLAG (GAP50 WT) or mutants lacking one potential *N*-glycosylation site (GAP50N_101-A, GAP50N_136-A, GAP50N_228-A) or all three *N*-glycosylation sites (GAP50N_101_136_228-AAA) after IP with anti-FLAG beads and staining with polyclonal antibodies specific to GAP45. Note the presence of GAP45 assembled with other glideosome components as large macromolecular complexes under nonreduced electrophoretic conditions (black arrows). Note the presence of a band of 55 kDa in the immunoprecipitated materials derived from the wild type and deglycosylated mutants tested with the anti-GAP45 antibodies. This band, which could represent a cross-reacting protein, was not further investigated.

present in the IMC (Fig. 5A). In addition, we have also checked the *N*-glycosylation status of all four stable mutant lines expressing transgenic GAP50 protein using crude protein extracts and Western blots. Fig. 5B shows that all three GAP50 mutants lacking respectively the first (GAP50N_101-A), second (GAP50N_136-A) and third (GAP50N_228-A) of the normally occupied *N*-glycosylated sites shifted similarly and migrate faster than the wild-type GAP50 protein (GAP50 WT). These electrophoretic mobility shifts and co-migration patterns were consistent with the molecular weight estimated at 52 kDa of the single mutated proteins lacking only one *N*-glycans because the triple GAP50 mutant (lacking all three *N*-glycans) displayed faster mobility shift with a molecular mass estimated at 45 kDa (Fig. 5B, lane GAP50HA N101_136_228-AAA) relative to the wild-type protein with a molecular mass of 55 kDa (lane GAP50 HA WT). Consistent with this conclusion, all three single mutants treated with PNGase F, which removed the two remaining *N*-glycans, lead to fully deglycosylated GAP50 species that co-migrate with the genetically engineered triple GAP50 mutant (Fig. 5C). Moreover, digestion of the wild-type GAP50 protein with PNGase F confirmed the electrophoretic mobility behavior of triple deglycosylated GAP50 mutants. Altogether, these data demonstrated that effective transfer of the two other *N*-gly-

cans occurs efficiently at some extends on each single mutant protein even though we noticed faster migrating bands in the first two mutants (Fig. 5B and 5C, lanes GAP50_N101-A and GAP50N_136-A) that perhaps more likely suggested that mutation of the first or second glycosylation site may also affect glycosylation of the other sites. Consistent with this notion, the fastest migrating species in the first mutant comigrates with the fully deglycosylated protein (Fig. 5B and 5C, lanes GAP50_N101-A).

Next, we examined the potential of deglycosylated GAP50 proteins to interact with glideosome partners. The stably transgenic expressing wild type and deglycosylated mutants were lysed with nonionic detergent (Triton X-100) and immunoprecipitated using anti-FLAG beads. Fig. 6 shows that the mutations of the first (GAP50N_101-A), second (GAP50N_136-A) or all three *N*-glycosylation sites (GAP50N_101_136_228-AAA), dramatically reduced the ability to co-immunoprecipitate the key gliding partner, TgMyoA (Fig. 6A). A significant reduction in the level of Myosin A-associated light chain 1 (MLC1, another actin-myosin motor) pulled down by the three deglycosylated GAP50 proteins was also noticed (Fig. 6B). To ensure that the wild type and mutated HAFLAG-tagged GAP50 proteins, used to co-immunoprecipitate partners, would reflect the association of glideo-

some components *in vivo*, living transgenic parasites were cross-linked with the reduction sensitive, cell-permeable cross-linker dithiobis succinimidyl propionate (DSP), which has a bond length of 7.7 Ångström. After cross-linking, parasites were lysed with Triton X-100 followed by immunoprecipitation using anti-FLAG column. The binding protein complexes were reduced by dithiothreitol before SDS-PAGE and Western blots confirmed that the mutations of the first, second or all three *N*-glycosylation sites prevent the ability of these mutated proteins to co-immunoprecipitate GAP45, one of its important partners involved in glideosome formation (Fig. 6C). As expected, the mutation of the third *N*-glycosylated site, which has no critical effect on trafficking and IMC localization, readily immunoprecipitated GAP45 in a manner similar to the wild-type GAP50 (Fig. 6C, lanes GAP50WT and GAP50N_228-A). Taken together, these data are in good agreement with the two-step assembly of the glideosome: first, the transmembrane type I GAP50 protein is inserted in the nascent IMC during daughter parasite formation and, second, GAP45, TgMyoA and TgMLC are preformed cytoplasmic complexes, which associate to GAP50, already present in the IMC (9). Therefore, we interpreted that the residual binding of MyoA, MLC and GAP45 to deglycosylated GAP50 is likely because of the detergent lysis of transgenic parasites that disrupts the ER/Golgi compartment where deglycosylated GAP50 was stuck, thereby bringing GAP50 to the preformed cytoplasmic GAP45-MyoA-MLC complexes to close vicinity for binding. Thus, the single GAP50 mutant at the first and second *N*-glycosylation site and the triple GAP50 mutant protein, which result in deglycosylation of GAP50 proteins, preventing proper trafficking to IMC, also implies the absence of association with the other glideosome partners.

Glideosome Complexes are Disrupted by the Absence of N-glycans on GAP50—We discovered that, in contrast to previous studies (9), a mild detergent extraction (1% Triton X-100) followed by SDS-PAGE analysis under unreduced conditions (without β -mercaptoethanol or dithiothreitol) in SDS-containing buffers was insufficient to completely dissociate all of the glideosome proteins, even after sample boiling at 100 °C, suggesting that the glideosome proteins can form a large stable and lower migrating complexes of about 125 kDa (Fig. 6D, lanes WT and GAP50N228-A, black arrows). When this lower migrating protein complex (>120 kDa) was excised after SDS-PAGE (supplemental Fig. S1, panel B) and subjected to in-gel trypsin digestion followed by proteomic analyses, we identified numerous peptides from all known glideosome components such as GAP50, GAP45, MyoA, MLC, and actin (Table I). The presence of all glideosome components was confirmed by proteomic analyses under lower (supplemental Table S1) and higher salt stringency (Table I), and by Western blots using specific antibodies. These data provide new evidence of the existence of a glideosome multiprotein complex containing GAP50, GAP45, MyoA, MLC1, and actin, which cannot be dissociated under mild

detergent, boiling and unreduced electrophoretic conditions. In light of the fact that myosin and actin are known to multimerize, it remains to estimate whether all proteins are in near stoichiometric amounts in this glideosome multiprotein complex. However, the tiny amounts of proteins that can be isolated from the glideosome multiprotein complex did not allow us to perform such quantitative proteomic experiments.

To further investigate the implications of *N*-glycosylation in partner binding by GAP50, we have compared wild type, single and triple deglycosylated GAP50 proteins in pull down assays using anti-FLAG beads. In contrast to wild-type GAP50 and the GAP50N_228-A mutant, we found that the deglycosylated GAP50N_101-A, GAP50N_136-A and fully deglycosylated GAP50N_101_136_228-AAA mutants were not able to form significant levels of the faster migrating band (Fig. 6D, lanes GAP50N_101-A, GAP50N_136-A and GAP50N_101_136_228-AAA), similar to the absence of MyoA and GAP45 binding (Fig. 6A and 6C).

Among the glideosome proteins identified under high salt stringency in Table I is also GAP40, a polytopic protein of the IMC, which has been recently validated as an additional glideosome component (34). Our comparative proteomic analyses also indicated that GAP40 (hypothetical protein TGME49_049850) is present in both monomeric and large macromolecular complexes pulled down with wild-type GAP50, whereas the triple deglycosylated mutant failed to immunoprecipitate GAP40 (Table I). As for GAP40, two other novel hypothetical proteins (TGME49_020950 and TGME49_058060) were also pulled down when GAP50 is *N*-glycosylated (Table I). To get more insight into the possible roles of these three novel hypothetical proteins that are pulled down by GAP50-HAFLAG, we generated stable transgenic parasites that ectopically expressed these novel proteins using knockin promoter targeting strategy that allowed the expression of HA tagged version of these genes in their corresponding loci in the parasite genome. We confirmed that the hypothetical protein coded by TGME49_049850, which has been recently described as a novel gliding associated protein GAP40 (34) colocalized with two IMC markers, GAP45 and IMC1 (Fig. 7A). In contrast, the two other hypothetical proteins, which were respectively coded by TGME49_020950 and TGME49_058060, were detected in dense granules and also in the parasitophorous vacuole, suggesting that these two other hypothetical proteins are unrelated components that localized in the dense granules. Therefore, these two later hypothetical proteins were not further investigated. Because, we found that GAP40 protein can be only pulled down by *N*-glycosylated GAP50, the role of *N*-glycans in the association between GAP40 and other gliding partners was investigated. When the parasites expressing GAP40-HA were incubated with the *N*-glycosylation inhibitor tunicamycin, neither GAP45 nor MLC1 protein can be immunoprecipitated by GAP40-HA (Fig. 7B, lanes 4). As a positive control, GAP45 and MLC1 can be specifically immunoprecipitated by

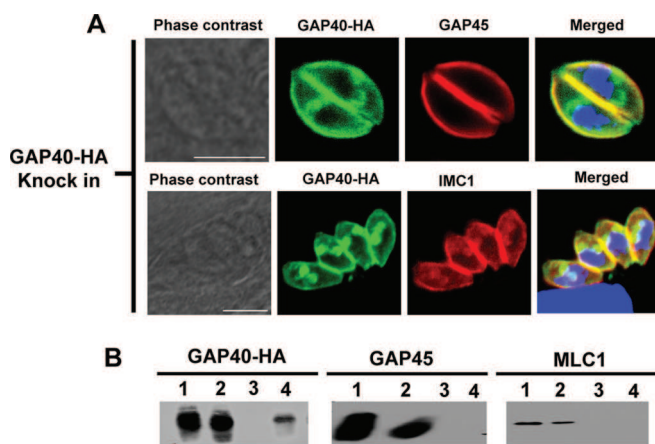


FIG. 7. Gene knock-in for HA-tagged GAP40 expression in *T. gondii*. *A*, Tachyzoites expressing the GAP40-HA protein (green) under the control of its own promoter targeted by gene knock-in strategy. A perfect colocalization between GAP40-HA fluorescence signal and that of GAP45 (red, upper panels) and IMC1 (red, lower panels), two markers of the inner membrane complex was observed. Note the presence of GAP40 in the newly formed daughter parasites (intracellular green dots), indicating that GAP40 is inserted as early as GAP50 in the IMC of the daughter parasites. *B*, Western blots after immunoprecipitation of GAP40-HA using detergent lysates from stable transgenic parasites and antibodies specific to HA. In addition to GAP40-HA, two gliding partners GAP45 and MLC1 were shown to be specifically immunoprecipitated. Lanes 1, total protein extracts from transgenic GAP40-HA expressers; lanes 2, lysates were immunoprecipitated with anti-HA; lanes 3, lysates were immunoprecipitated with negative sera; lane 4, lysates from transgenic GAP40 expressers treated with tunicamycin were immunoprecipitated with mouse anti-HA. Western blots were performed using rabbit anti-HA, anti-GAP45 and anti-MLC1, respectively.

GAP40-HA in the absence of tunicamycin (lanes 2). Collectively, these data support the notion that *N*-glycosylation is important for the assembly of the glideosome components in the parasite's pellicle.

N-glycosylation of GAP50 is Pertinent to Parasite Motility and Host Cell Invasion—We and others have previously reported that tunicamycin, an inhibitor which prevents the addition of *N*-glycans to the Asn residues of *N*-glycoproteins, also acts in a peculiar manner in *T. gondii* (21, 22) compared with other eukaryotic cells. Although no obvious effects can be seen during the first round of intracellular development of tachyzoites inside the host cells, the tunicamycin-treated parasites were considerably reduced in motility and host cell invasion when they ruptured from the cell and are engaged in a second infection cycle. We observed that tunicamycin induces important alterations in the biogenesis of *T. gondii* IMC and the parasite motile apparatus are impaired. To determine whether the effect of tunicamycin observed specifically involved *N*-glycosylation of GAP50, we first wanted to knockout the gene encoding wild-type GAP50 in the transgenic parasites already containing a second mutant copy of the gene, which ectopically expressing either *N*-glycosylated or unglycosylated GAP50 protein. As previously reported (9), our at-

tempts to knock-out GAP50 gene, despite the presence of the second gene copy, have also been unsuccessful. Therefore, we decided to investigate how the wild type and mutant parasites ectopically expressing wild type and unglycosylated GAP50-HAFLAG protein behave in the presence of tunicamycin. It should be noted that we chose to study a transgenic line where fluorescence and Western blot signals of ectopic GAP50-HAFLAG were roughly comparable to endogenous signals (9). In the transgenic parasites, the sum of GAP50 levels is made up of both ectopically expressed GAP50-HAFLAG plus endogenous GAP50 and, as a consequence, the overall GAP50 protein levels are higher than in wild-type parasites. We then monitored the gliding of drug-treated and untreated parasites using immunofluorescence staining to detect the trails left on slides by motile parasites (Fig. 8). When intracellular parasites expressing GAP50-HAFLAG protein ectopically in addition to wild-type GAP50 protein were incubated with tunicamycin, the freshly released ectopic expressers were able to display linear trail motility that was longer (Fig. 8D) than those of the normal wild type (Fig. 8B and triple deglycosylated GAP50 mutant (Fig. 8F), also treated by the same concentration of tunicamycin. We estimated that the parasites expressing two copies of *N*-glycosylated GAP50 gene trailed three times longer than the two other parasite lines (Fig. 8G). In the absence of tunicamycin, these ectopically expressing GAP50 parasites displayed a normal circular gliding (Fig. 8C), a behavior typical of either wild type (Fig. 8A) or fully deglycosylated GAP50-expressing parasites (Fig. 8E). Interestingly, we notice that the circular trail, another characteristic of parasite motility, is completely abolished in all parasite lines in the presence of tunicamycin (Fig. 8, compare left panels A, C, and E to right panels B, D, and F). To establish whether tunicamycin could inhibit the host cell invasion of the three parasite lines, we monitored host cell re-infection after the first cycle of intracellular parasite treatment. Fig. 8H shows that the capability of the ectopic GAP50-expressers to display longer trails or motility in the presence of tunicamycin is consistent with these transgenic parasites being able to invade new host cells more efficiently (53%) than wild-type parasites (26%) and unglycosylated GAP50 mutants (29%). This suggests that tunicamycin does not completely shut down glycosylation under our experimental conditions, and the over-expression of GAP50 probably allows obtaining a larger share of the available *N*-glycans than the wild-type GAP50. We conclude that *N*-glycosylation plays key roles in the trafficking of gliding components to the IMC, in the assembly of functional glideosome that is required for parasite motility and host cell invasion.

DISCUSSION

In this study, we demonstrate that *N*-linked glycosylation plays two important roles in *T. gondii* GAP50. First, *N*-linked glycans are essential for the correct intracellular trafficking of

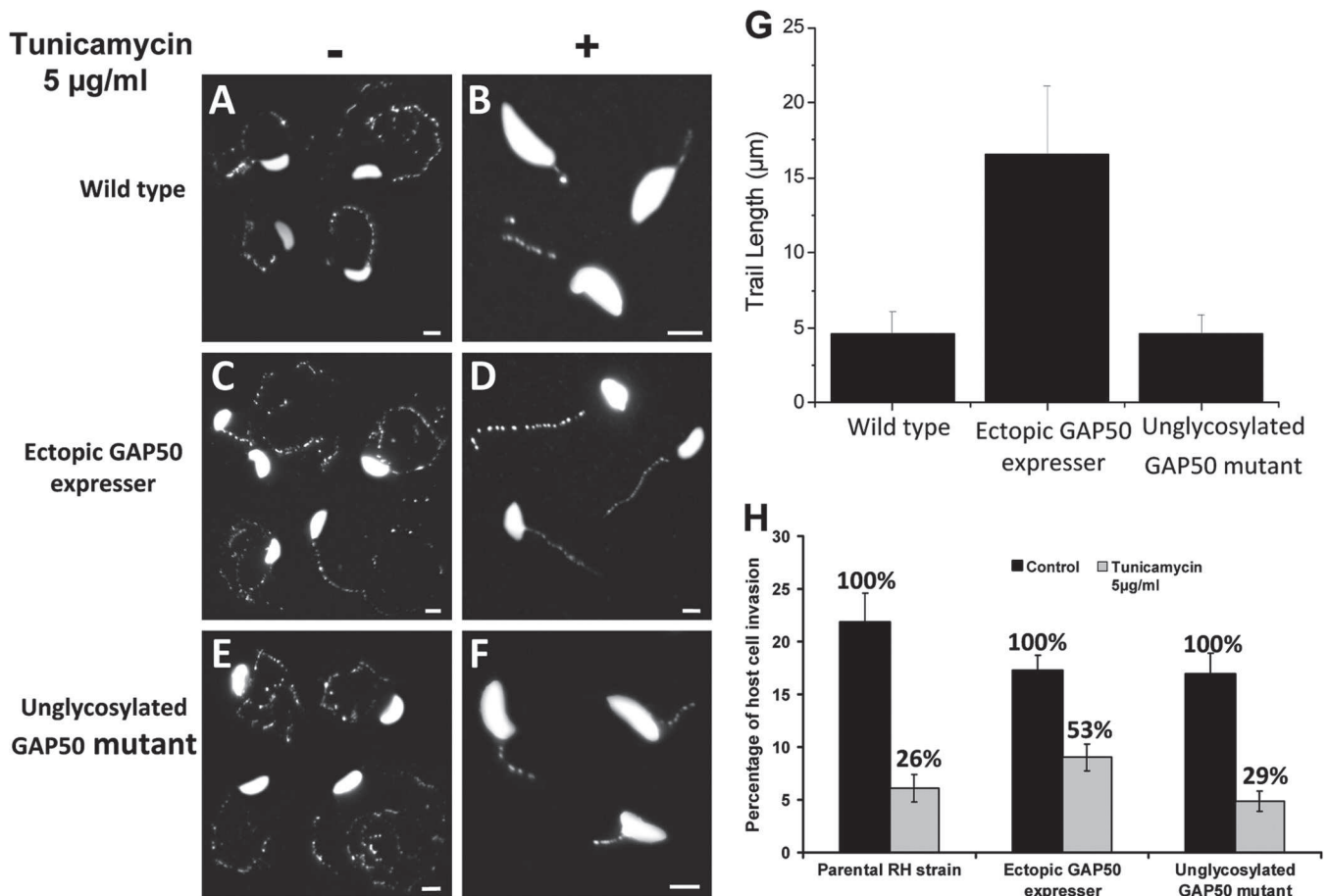


FIG. 8. Assessment of gliding motility by trail deposition and host cell invasion. (A, C, and E) Tachyzoites from wild-type RH strain, ectopic GAP50-HAFLAG expressers and triple mutant (GAP50N101_136_228-AAA-HAFLAG) lines were treated with DMSO as a control. (B, D, and F) The three transgenic lines treated with tunicamycin were released and allowed to glide on serum-coated slides and visualized by staining with anti-SAG1 antibodies. Scale bar, 5 µm. G, Quantification of the length of linear trails of tachyzoites from wild-type RH strain, ectopic GAP50-HAFLAG expressers and triple mutant (GAP50N101_136_228-AAA-HAFLAG) parasite lines treated with tunicamycin during the first cycle of host cell infection as above. Results are mean ± S.E. of at least three experiments ($n = 3$ and $p < 0.005$). H, Effect of tunicamycin on host cell entry by tachyzoites from wild-type RH strain, ectopic GAP50-HAFLAG expressers and triple mutant (GAP50N101_136_228-AAA-HAFLAG). The percentage of invaded host cell was calculated by comparing that of the untreated (DMSO control) and tunicamycin-treated parasites, which were allowed to invade new HFF, fixed and stained with anti-SAG1 for IFA. Bars represent S.E. from three independent experiments ($n = 2$ and $p < 0.001$).

GAP50 and for its insertion into the IMC, a structure that powers parasite motility, migration through host tissues, host cell invasion and egress. Second, the interactions between components of the gliding machinery can be influenced by the N-glycosylation status of GAP50. Although the third N-glycosylation site is apparently not crucial, both the first or second N-glycosylation sites has to carry at least one N-linked sugar for GAP50 to be properly trafficked to the IMC. When drafting this paper, we were aware about the discovery of GAP40, a novel glideosome component recently validated by Frénil *et al.* (34), thereby defining this motor machinery as being composed of at least five proteins (MyoA-MLC1-GAP45-GAP50-GAP40). We independently identified GAP40 during our study as a hypothetical protein (TGME49_049850), which is present in both monomeric protein and is associated with the glideosome multiprotein complex pulled down using higher salt

stringency. We showed that removal of N-glycans from GAP50 results in dissociation of this glideosome multicomplex complex and losses of several key glideosome components, including GAP40.

Interestingly, we also found that the two first key N-glycosylation sites for GAP50 trafficking in *T. gondii* are well-conserved in two other related apicomplexan parasites, *Neospora caninum* and *Eimeria tenella*, whereas the third N-glycosylation site, apparently not essential, is absent in *E. tenella* (supplemental Fig. S3). Furthermore, the two first putative N-glycosylation sites are also present as divergent sequons in GAP50 proteins of *Plasmodia* species. Despite, the apparent lack of numerous genes coding glycosyl-transferases involved in the assembly of complex N-glycan structures, biochemical studies recently indicated that *Plasmodium falciparum* also synthesizes N-glycans that are shorter than that of

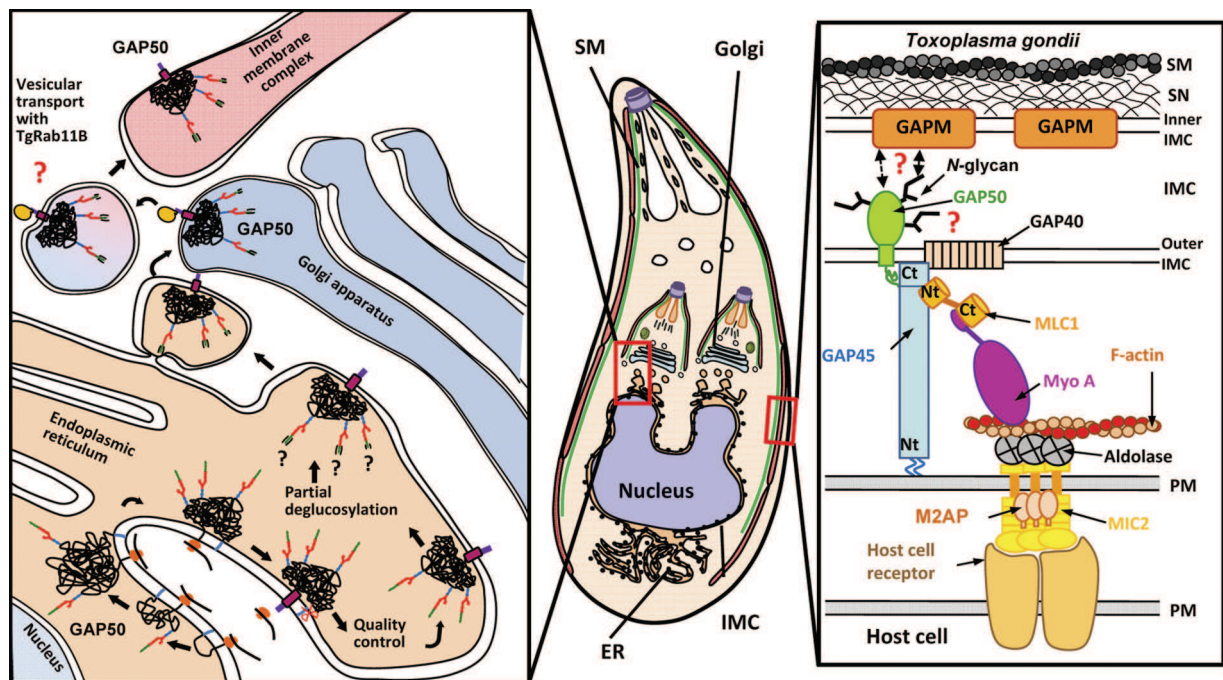


FIG. 9. **Proposed model of N-glycan functions of *T. gondii* GAP50 in trafficking and binding key glideosome partners.** According to the currently accepted model, which has been recently revisited and improved by Frénal *et al.* (34), our data indicate that N-glycosylation may be a key post-translational modification for GAP50 transport through the classical ER-Golgi secretory pathway (*left panel*) until its final destination, the IMC, where it will associate with other glideosome partners (*right panel*). The N-glycans on GAP50 are not only essential for proper trafficking but probably also important for direct or indirect interactions with the other glideosome partners GAP45, GAP40, MLC1, MyoA, and Actin.

mammalian and *T. gondii* counterparts (35, 36). Interestingly, recent studies have also demonstrated that GAP50 of *P. falciparum* is located in the endoplasmic reticulum in early stage parasites (37). However, it remains to be determined if the shorter N-glycan structures could also play a direct role for trafficking *P. falciparum* GAP50 to the inner membrane complex.

Our present study provides for first time insights into how N-glycosylation could regulate correct protein folding and trafficking of N-glycoproteins in a model of apicomplexan parasites. It is tempting to hypothesize that in *T. gondii*, the correct protein trafficking may be ensured by the unusual N-glycan chains, which are present on the three potential glycosylated sites of GAP50. These N-glycan chains are composed of immature and truncated oligomannose-rich sugars, which in some cases contain abnormally uncleaved terminal glucoses that are rarely present on N-glycoproteins of higher eukaryotes. In addition, our studies also provide evidence on the degree of N-glycan heterogeneity and site occupancy on GAP50. The absence of these N-glycans on the first two sites of GAP50 held the mutated protein in the ER/Golgi. Thus, the deglycosylated GAP50 cannot reach the IMC and participate in glideosome formation.

Previous studies have recently proposed models of how the glideosome is inserted in the IMC and this is schematically depicted in Fig. 9, updated according to the potential of

N-glycosylation functions reported herein. This model suggests that GAP50 functions as the protein that anchors the glideosome *via* its transmembrane domain, with the much larger N-terminal part of the protein being located in the lumen of the IMC, except the last C-terminal six amino acids that is involved in binding to GAP45, the novel GAP40 and the motor MyoA-MLC1-Actin. Even though the simplest interpretation of N-glycans functions on GAP50 protein could be that unglycosylated GAP50 is stuck in the trafficking ER and Golgi pathway, thereby preventing its interactions with other glideosome partners, we also demonstrated that deglycosylated GAP50 cannot physically bind to partners. This supports the notion that glycan-glycan or glycan-protein interactions may be involved in the glideosome formation and parasite's motility. Furthermore, we show in this study that N-glycosylation status of GAP50 is important for parasite gliding and host cell invasion.

Because the three N-glycans of GAP50 protein are located inside the lumen of the IMC, we hypothesize that these N-glycan chains may impact glideosome assembly and functions at the opposite side of the outer IMC membrane through glycan-glycan and/or glycan-protein interactions (Fig. 9). One of these GAP50 interacting partners that could cross the outer IMC is the novel GAP40, which contains nine spanning transmembrane domains and one potential N-glycosylation site. However, the exact topology of

GAP40, its precise *N*-glycosylation status and contribution to the glideosome's functions will await further experimental investigations.

The conservation of the motor complex across apicomplexan parasites is further supported by the recent demonstration that gliding motility of these pathogens to invade host cells can be inhibited in *T. gondii* (38), *Cryptosporidium* (39), *Theileria* (40), and *Babesia* (41). Together with the presence of glideosome in *Plasmodium falciparum* (11, 12), we conclude that a generic motor complex is used across the phylum Apicomplexa and has been adapted to the variety of post-translational modifications that can control glideosome functions, such as phosphorylation (16), acylation (17) and *N*-glycosylation (21, 22, this study). A detailed understanding of the biological functions of factors that modulate parasite motility, such as the unusual *N*-glycan structures identified in *T. gondii*, and the conservation of this motor complex with numerous hypothetical proteins exclusively found in apicomplexan parasites highlights potential chemotherapeutic targets that may be applicable for a wide variety of these important human and livestock pathogens.

Acknowledgements—We thank Dr D Roos (University of Pennsylvania) for providing HDEL-RFP and GRASP-RFP parasite lines, Dr D Soldati for rabbit polyclonal antibodies specific to MLC1, MyoA and GAP45, Dr G Ward for monoclonal antibodies specific to IMC1. We are grateful to Drs M Gissot, R Walker and SG Ball for critical reading of the manuscript.

* This work was supported by the CNRS, INSERM and Pasteur Institute of Lille. SF was supported by a fellowship from the French Ministry of Research. The Mass Spectrometry facility used in this study for released glycan studies was funded by the European Community (FEDER), the “Region Nord-Pas de Calais,” France, the CNRS and the University of Sciences and Technology of Lille. Proteomic and glycopeptides studies were supported by the CNRS, the GIS IBISA, the “Agence Nationale de la Recherche” (ANR) the “Fondation de la Recherche Médicale” (FRM) and the “Region Alsace”.

☐ This article contains supplemental Figs. S1 to S3 and Table S1.

** To whom correspondence should be addressed: 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. E-mail: Stan.Tomavo@pasteur-lille.fr.

REFERENCES

- Sibley, L. D., Håkansson, S., and Carruthers, V. B. (1998) Gliding motility: an efficient mechanism for cell penetration. *Curr. Biol.* **8**, R12–14
- Håkansson, S., Morisaki, H., Heuser, J., and Sibley, L. D. (1999) Time-lapse video microscopy of gliding motility in *Toxoplasma gondii* reveals a novel, biphasic mechanism of cell locomotion. *Mol. Biol. Cell* **10**, 3539–3547
- Huynh, M. H., and Carruthers, V. B. (2006) *Toxoplasma* MIC2 is a major determinant of invasion and virulence. *Plos Pathol.* **2**, e84
- 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
- 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
- Meissner, M., Schlüter, D., and Soldati, D. (2002) Role of *Toxoplasma gondii* myosin A in powering parasite gliding and host cell invasion. *Science* **298**, 837–840
- Opitz, C., and Soldati, D. (2002) The glideosome: a dynamic complex powering gliding motion and host cell invasion by *Toxoplasma gondii*. *Mol. Microbiol.* **45**, 597–604
- 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
- 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
- Johnson, T. M., Rajfur, Z., Jacobson, K., and Beckers, C. J. (2007) Immobilization of the type XIV myosin complex in *Toxoplasma gondii*. *Mol. Biol. Cell* **18**, 3039–3046
- Bullen, H. E., Tonkin, C. J., O'Donnell, R. A., Tham, W. H., Papenfuss, A. T., Gould, S., Cowman, A. F., Crabb, B. S., and Gilson, P. R. (2009) A novel family of apicomplexan glideosome-associated proteins with an inner membrane-anchoring role. *J. Biol. Chem.* **284**, 25353–25363
- Baum, J., Richard, D., Healer, J., Rug, M., Krnjajski, Z., Gilberger, T. W., Green, J. L., Holder, A. A., and Cowman, A. F. (2006) A conserved molecular motor drives cell invasion and gliding motility across malaria life cycle stages and other apicomplexan parasites. *J. Biol. Chem.* **281**, 5197–5208
- Heaslip, A. T., Leung, J. M., Carey, K. L., Catti, F., Warshaw, D. M., Westwood, N. J., Ballif, B. A., and Ward, G. E. (2009) A small-molecule inhibitor of *T. gondii* motility induces the posttranslational modification of myosin light chain-1 and inhibits myosin motor activity. *PLoS Pathol.* **6**, e1000720
- Wetzel, D. M., Håkansson, S., Hu, K., Roos, D., and Sibley, L. D. (2003) Actin filament polymerization regulates gliding motility by apicomplexan parasites. *Mol. Biol. Cell* **14**, 396–406
- Wetzel, D. M., Chen, L. A., Ruiz, F. A., Moreno, S. N., and Sibley, L. D. (2004) Calcium-mediated protein secretion potentiates motility in *Toxoplasma gondii*. *J. Cell Sci.* **117**, 5739–5748
- Gilk, S. D., Gaskins, E., Ward, G. E., and Beckers, C. J. (2009) GAP45 phosphorylation controls assembly of the *Toxoplasma* myosin XIV complex. *Eukaryot Cell* **8**, 190–196
- Rees-Channer, R. R., Martin, S. R., Green, J. L., Bowyer, P. W., Grainger, M., Molloy, J. E., and Holder, A. A. (2006) Dual acylation of the 45 kDa gliding-associated protein (GAP45) in *Plasmodium falciparum* merozoites. *Mol. Biochem. Parasitol.* **149**, 113–116
- Pomel, S., Luk, F. C., and Beckers, C. J. (2008) Host cell egress and invasion induce marked relocations of glycolytic enzymes in *Toxoplasma gondii* tachyzoites. *PLoS Pathol.* **4**, e1000188
- 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
- Bushkin, G. G., Ratner, D. M., Cui, J., Banerjee, S., Duraisingh, M. T., Jennings, C. V., Dvorin, J. D., Gubbels, M. J., Robertson, S. D., Steffen, M., O'Keefe, B. R., Robbins, P. W., and Samuelson, J. (2010) Suggestive evidence for Darwinian Selection against asparagine-linked glycans of *Plasmodium falciparum* and *Toxoplasma gondii*. *Eukaryot Cell* **9**, 228–241
- Luk, F. C., Johnson, T. M., and Beckers, C. J. (2008) *N*-linked glycosylation of proteins in the protozoan parasite *Toxoplasma gondii*. *Mol. Biochem. Parasitol.* **157**, 169–178
- Fauquenoy, S., Morelle, W., Hovasse, A., Bednarczyk, A., Slomianny, C., Schaeffer, C., Van Dorsselaer, 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
- Olguin-Lamas, A., Madec, E., Hovasse, A., Werkmeister, E., Callebaut, I., Slomianny, C., Delhaye, S., Mouveaux, T., Schaeffer-Reiss, C., Van Dorsselaer, A., and Tomavo, S. (2011) A novel *Toxoplasma gondii* nuclear factor TgNF3 is a dynamic chromatin-associated component, modulator of nucleolar architecture and parasite virulence. *PLoS Pathogens* **7**, e1001328
- Castro, A. J., Bednarczyk, A., Schaeffer-Reiss, C., Rodríguez-García, M. I., Van Dorsselaer, A., and de Dios, Alché, J. (2010) Screening of Ole e 1 polymorphism among olive cultivars by peptide mapping and *N*-glycopeptide analysis. *Proteomics* **10**, 953–962
- Geer, L. Y., Markey, S. P., Kowalak, J. A., Wagner, L., Xu, M., Maynard, D. M., Yang, X., Shi, W., and Bryant, S. H. (2004) Open mass spectrometry search algorithm. *J. Proteome Res.* **3**, 958–964

26. Peng, J., Elias, J. E., Thoreen, C. C., Licklider, L. J., and Gygi, S. P. (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
27. Elias, J. E., and Gygi, S. P. (2007) Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Meth.* **4**, 207–214
28. Morelle, W., and Michalski, J. C. (2007) Analysis of protein glycosylation by mass spectrometry. *Nat. Protoc.* **2**, 1585–1602
29. Mann, T., and Beckers, C. (2001) Characterization of the subpellicular network, a filamentous membrane skeletal component in the parasite *Toxoplasma gondii*. *Mol. Biochem. Parasitol.* **115**, 257–268
30. Wuhrer, M., Catalina, M. I., Deelder, A. M., and Hokke, C. H. (2007) Glycoproteomics based on tandem mass spectrometry of glycopeptides. *J. Chromatogr. B* **849**, 115–128
31. Elbein, A. D. (1981) The tunicamycin: useful tools for studies on glycoproteins. *Trends Biochem. Sci.* **6**, 219–221
32. Nishi, M., Hu, K., Murray, J. M., and Roos, D. S. (2008) Organellar dynamics during the cell cycle of *Toxoplasma gondii*. *J. Cell Sci.* **121**, 1559–1568
33. Pelletier, L., Stern, C. A., Pypaert, M., Sheff, D., Ngó, H. M., Roper, N., He, C. Y., Hu, K., Toomre, D., Coppens, I., Roos, D. S., Joiner, K. A., and Warren, G. (2002) Golgi biogenesis in *Toxoplasma gondii*. *Nature* **418**, 548–552
34. Fréchal, K., Polonais, V., Marq, J. B., Stratmann, R., Limenitakis, J., and Soldati-Favre, D. (2010) Functional dissection of the apicomplexan glideosome molecular architecture. *Cell Host Microbe* **8**, 343–357
35. Banerjee, S., Vishwanath, P., Cui, J., Kelleher, D. J., Gilmore, R., Robbins, P. W., and Samuelson, J. (2007) The evolution of N-glycan-dependent endoplasmic reticulum quality control factors for glycoprotein folding and degradation. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 11676–11681
36. Bushkin, G. G., Ratner, D. M., Cui, J., Banerjee, S., Duraisingh, M. T., Jennings, C. V., Dvorin, J. D., Gubbels, M. J., Robertson, S. D., Steffen, M., O’Keefe, B. R., Robbins, P. W., and Samuelson, J. (2010) Suggestive evidence for Darwinian Selection against asparagine-linked glycans of *Plasmodium falciparum* and *Toxoplasma gondii*. *Eukaryot Cell* **9**, 228–241
37. Yeoman, J. A., Hanssen, E., Maier, A. G., Klonis, N., Maco, B., Baum, J., Turnbull, L., Whitchurch, C. B., Dixon, M. W., and Tilley, L. (2011) Tracking glideosome-associated protein-50 reveals the development and organization of the inner membrane complex of *P. falciparum*. *Eukaryot Cell* **4**, 556–564
38. 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
39. Wetzel, D. M., Schmidt, J., Kuhlenschmidt, M. S., Dubey, J. P., and Sibley, L. D. (2005) Gliding motility leads to active cellular invasion by *Cryptosporidium parvum* sporozoites. *Infect. Immun.* **73**, 5379–5387
40. Carreno, R. A., Martin, D. S., and Barta, J. R. (1999) *Cryptosporidium* is more closely related to the gregarines than to coccidia as shown by phylogenetic analysis of apicomplexan parasites inferred using small-subunit ribosomal RNA gene sequences. *Parasitol. Res.* **85**, 899–904
41. Shaw, M. K. (1999) *Theileria parva*: sporozoite entry into bovine lymphocytes is not dependent on the parasite cytoskeleton. *Exp. Parasitol.* **92**, 24–31

III. Identification des facteurs de transcription qui gouvernent le mécanisme de différenciation de *T. gondii* de la forme virulente en forme dormante

Ce travail a fait l'objet d'un article dans le Journal Molecular Biology (Annexe 2).

1. Contexte et objectif de cette étude

Des études récentes ont permis la découverte d'un large répertoire de "small RNA" encore jamais décrit chez le tachyzoïte. Or les "small RNA" sont largement reconnus chez d'autres organismes pour aider à la régulation des facteurs de transcription et le maintien des télomères. Cette découverte a montré que le mécanisme de régulation de l'expression des gènes de *T. gondii* largement contrôlé par la transcription reste encore mal connu⁴⁷.

Dans le but d'identifier les facteurs de transcription responsables de la différenciation chez *T. gondii*, nous avons réalisé une étude protéomique qui s'inscrit dans le cadre d'un projet ANR MIE (2009-2012). Nous nous sommes attachés à identifier les protéines spécifiquement liées à certains motifs d'ADN ou ARN qui sont indispensables à la transcription ou traduction induisant le changement de formes de *T. gondii*. Dans une étude préalable réalisée au laboratoire³, 39 protéines nucléaires avaient été identifiées, dont 21 "hypothetical proteins" qui ne possédaient pas de fonction biologique décrite et parmi lesquelles plus de la moitié n'avait pas d'homologie avec des facteurs nucléaires connus.

De cette liste de protéines identifiées, 2 protéines nommées TgAlba1 et TgAlba2 (*T. gondii* Acetylation Lowers Binding Affinity) présentant de forte homologie avec des protéines d'archaebactéries^{39, 40} se sont révélées particulièrement intéressantes. Chez les archaebactéries, les protéines Alba sont les constituants majeurs de la chromatine qui modulent la transcription⁴⁸. L'interaction de ces protéines Alba avec les acides nucléiques dépend considérablement de leur degré d'acétylation⁴⁸ et de leur capacité à former des dimères⁴⁹. Chez *Trypanosoma brucei* par exemple, les protéines Alba contrôlent l'initiation de la transcription⁵⁰ et la différenciation⁵¹.

Afin de mieux comprendre les différents rôles du complexe TgAlba dans la transcription et la traduction des protéines, nous avons réalisé une étude protéomique des protéines co-purifiées avec les protéines TgAlba1 ou TgAlba2 grâce à la technique du TAP-tag (Tandem affinity Purification tag).

2. Stratégie d'analyse protéomique des complexes protéiques des TgAlba

a. Description des échantillons

Les parasites ont été génétiquement modifiés pour exprimer deux formes différentes de la protéine TgAlba soit sous la forme TgAlba1-TAP-tag soit sous la forme TgAlba2-TAP-tag. Les parasites ont été lysés pour extraire les protéines de leur cytoplasme. Les différents complexes protéiques des protéines TgAlba ont été purifiés par chromatographie d'affinité spécifique au tag et séparés sur gel SDS-PAGE. Les bandes de gel obtenues ont été découpées de façon systématique tous les 5 mm pour l'analyse par nanoLC-MS/MS. Les protocoles utilisés pour la préparation des échantillons sont détaillés dans la partie expérimentale de la publication de ces résultats (Annexe 2).

b. Analyse nanoLC-MS/MS

La stratégie adoptée pour l'identification des partenaires des protéines TgAlba est décrite sur la Figure 4.

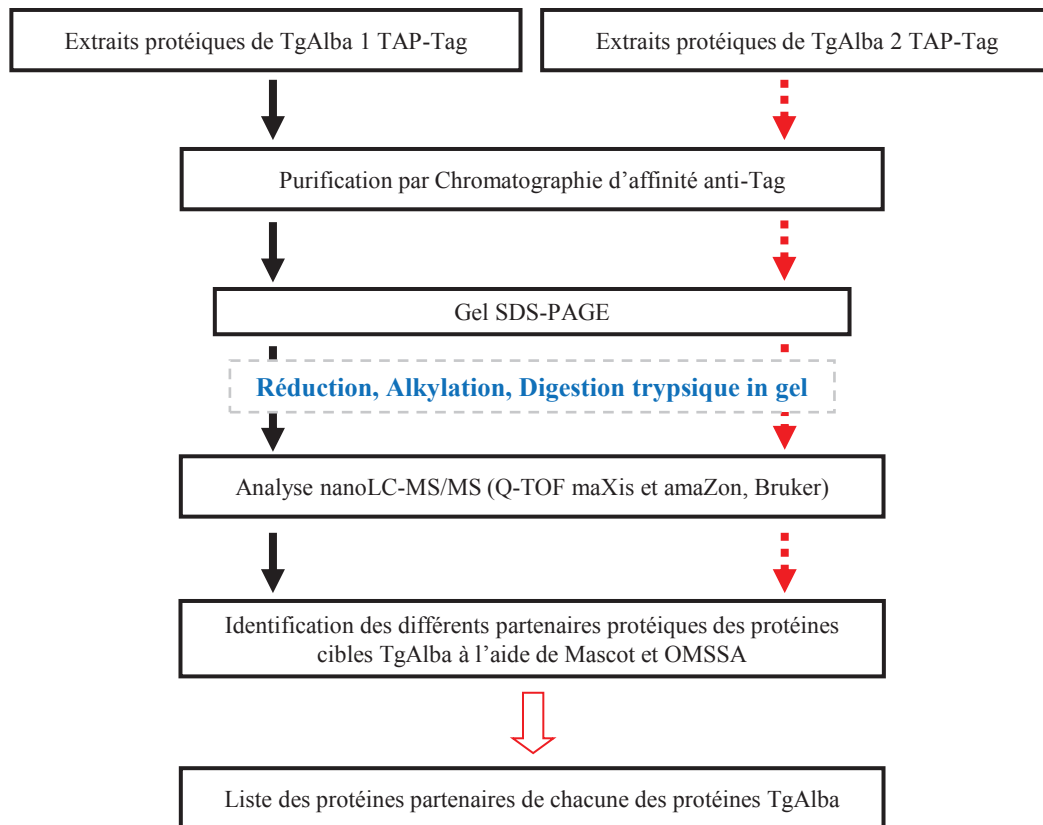


Figure 4. Stratégie globale d'analyse protéomique pour l'identification des partenaires protéiques des 2 protéines cibles TgAlba1 et TgAlba2.

Les analyses nanoLC-MS/MS ont été réalisées, soit sur une nanoACQUITY Ultra-performance-LC (Waters) couplée à un spectromètre de masse de type Q-TOF (maXis, Bruker) soit un système nanoLC-Chip/MS (Agilent) couplé à une trappe ionique amaZon (Bruker) dans le but de bénéficier des différents avantages de chacun de ces systèmes. Les paramètres instrumentaux utilisés sont détaillés dans la partie expérimentale des résultats soumis pour publication (Annexe 2). Les identifications des partenaires des protéines TgAlba ont été faites avec deux algorithmes de recherche (Mascot et OMSSA) pour augmenter la fiabilité des résultats. Une banque extraite de ToxoDB (www.toxodb.org) a été utilisée.

Les paramètres utilisés lors de cette recherche et les critères de validation des résultats de la recherche d'identification sont décrits dans les résultats soumis pour publication (Annexe 2).

3. Résultats publiés

L'analyse protéomique des partenaires des protéines TgAlba a permis de confirmer l'interaction entre les deux protéines TgAlba1 et TgAlba2, avec notamment l'identification des 40 protéines partenaires communes entre les deux protéines TgAlba.

De plus, TgAlba2 a été co-purifiée avec de nombreuses protéines impliquées dans la régulation de la traduction, comme les sous-unités du facteur d'initiation de la traduction (eIF3). L'ensemble des sous-unités du complexe HSP-chaperonin impliqué dans le repliement co-traductionnel des protéines, des sous-unités du protéasome et les protéines ribosomales 40S et 60S ont été identifiées. De plus, parmi les partenaires putatifs de TgAlba 2, nous avons identifié 10 des 12 sous-unités eIF3 codées dans le génome de *T. gondii* et 9 des 16 des sous unités de régulation du protéasome 26S.

Une protéine homologue à la famille SsossB, qui est décrite chez les archéobactéries comme partenaire de l'ADN monobrin⁵², que nous avons appelée TgSsossB a été identifiée. Les protéines partenaires de TgSsossB, ont été co-purifiées avec TgSsossB par la technique de TAG-Tag, et analysées par nanoLC-MS/MS avec la même approche méthodologique.

La comparaison des partenaires identifiés entre les protéines TgAlba et TgSsossB ont montré que 24 protéines (Tableau 1, Annexe 2) étaient communes entre les trois facteurs de transcription (Figure 5).

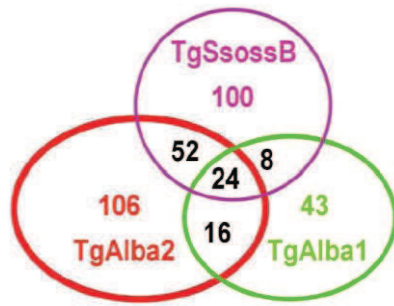


Figure 5. Comparaison des protéines partenaires des TgAlba et TgSsossB identifiées par analyse protéomique. Les chiffres représentent les nombres de protéines identifiées.

Aucune acétylation sur les protéines TgAlba comme il est décrit chez les homologues Alba des archéobactéries n'a pu être identifiée. La méthodologie développée dans ce travail n'a pas été spécifiquement développée pour la détection de protéines acétylées, ce qui pourrait expliquer ce résultat. En effet, l'étude spécifique des PTMs (Modifications Post-Traductionnelles) comme les acétylations nécessitent la purification de protéines mais aussi des techniques d'enrichissement de peptides acétylés par immuno-affinité avec les anticorps anti Acétylation-K⁵³.

Conclusion

Nous avons adapté une stratégie d'analyse de complexes protéiques disponibles en faibles quantités et permettant de limiter la contamination des échantillons par la kératine. Ainsi grâce à la stratégie d'analyse protéomique ciblant les complexes protéiques des protéines TgAlba purifiées par TAP-tag, nous avons pu mettre en évidence le double rôle de régulation de la transcription et la traduction de ces protéines TgAlba.

Les résultats d'analyse protéomique ont permis de confirmer l'interaction entre TgAlba1 et TgAlba2. De plus, TgAlba2 est co-purifiée avec de nombreuses protéines impliquées dans la régulation de la traduction comme les sous-unités du facteur d'initiation de la traduction eIF3. L'ensemble des sous-unités du complexe HSP-chaperonin impliqué dans le repliement co-translationnel des protéines ainsi que des sous-unités du protéasome ont été de même identifiées. Il est important de noter qu'eIF3, le complexe chaperonin et le protéasome sont connus pour interagir chez la levure⁵⁴. De plus, certaines combinaisons de sous-unités du complexe eIF3 permettent la traduction spécifique de certains ARNm⁵⁵. Ces données indiquent que le complexe TgAlba serait aussi impliqué dans la régulation de la traduction spécifique de certains transcrits via son interaction avec eIF3.

Parmi les protéines identifiées, certaines ont été sélectionnées et suivies par microscopie confocale par nos collaborateurs biologistes, ce qui a permis de les localiser dans le cytoplasme du parasite. La co-localisation de ces partenaires avec les protéines TgAlba et TgSsossB a confirmé leur rôle dans les complexes protéiques étudiés.

L'ensemble des résultats de ces études de protéomique et de biologie moléculaire a permis d'émettre une hypothèse selon laquelle, la traduction des transcrits de gènes spécifiques de *T. gondii* pourrait être la fonction potentielle de ces complexes multi-protéiques. Nous avons ainsi pu établir que le complexe TgAlba joue un rôle important dans la régulation de la virulence du *T. gondii*.

Ce travail de protéomique va se poursuivre au laboratoire par l'étude des 19 autres partenaires putatifs de l'ADN précédemment identifiés³, avec une approche d'identification et quantification relative sans marquage dite "label free".

Cette étude pourra aider à découvrir le lien entre la régulation traductionnelle de l'expression des gènes, les mécanismes de la virulence et la différenciation chez *T. gondii*.

Annexe 2: Publication

***Toxoplasma gondii* Alba proteins are involved in translational control of gene expression**

Gissot M.; Walker R. ; Delhay S. ; **Alayi T D** ; Huot L. ; Hot D.; Callebaut I. ; Schaeffer-Reiss C. ;
Van Dorsselaer A. ; Tomavo S.
Journal Molecular Biology, (2013)

Toxoplasma gondii Alba Proteins Are Involved in Translational Control of Gene Expression

Mathieu Gissot¹, Robert Walker¹, Stephane Delhay¹, Tchilabalo Dilezitoko Alayi², Ludovic Huot¹, David Hot¹, Isabelle Callebaut³, Christine Schaeffer-Reiss², Alain Van Dorsselaer² and Stanislas Tomavo¹

1 - Center for Infection and Immunity of Lille, CNRS UMR 8204, INSERM 1019, Université Lille Nord de France, Institut Pasteur de Lille, Lille, France

2 - Laboratoire de Spectrométrie de Masse Bioorganique, IPHC, CNRS UMR 7178, Université de Strasbourg, 67087 Strasbourg, France

3 - Institut de Minéralogie et de Physique des Milieux Condensés, CNRS UMR 7590, Université Pierre et Marie Curie, 75005 Paris, France

Correspondence to Stanislas Tomavo: Stan.Tomavo@pasteur-lille.fr

<http://dx.doi.org/10.1016/j.jmb.2013.01.039>

Edited by R. L. Gonzalez

Abstract

Molecular mechanisms controlling gene expression in apicomplexan parasites remain poorly understood. Here, we report the characterization of two *Toxoplasma gondii* homologs of the ancient archeal Alba proteins named TgAlba1 and TgAlba2. The targeted disruption of TgAlba1 and TgAlba2 genes in both virulent type I and avirulent type II strains of *T. gondii* reveals that TgAlba proteins may have an important role in regulating stress response. We found that although the steady-state level of the *Tgalba2* transcript is increased in the $\Delta Tgalba1$ null mutant parasites, the cognate TgAlba2 protein is undetectable, suggesting that TgAlba1 is required for translation of TgAlba2. Using a tandem affinity purification tag strategy combined with proteomic analyses, we provide evidence that many factors known to be involved in the translation machinery are co-purified with TgAlba1 and TgAlba2. We further performed RNA pull-down and microarray analyses to show that TgAlba1 and TgAlba2 bind to more than 30 RNAs including their own transcripts. Moreover, we demonstrate that the tight translational regulation of the *Tgalba2* endogenous transcript relies on the presence of both its 3' untranslated region and that of the TgAlba1 protein. Thus, our findings on TgAlba1 and TgAlba2 are consistent with a role in gene-specific translation.

© 2013 Elsevier Ltd. All rights reserved.

Introduction

Toxoplasma gondii belongs to the apicomplexan phylum of eukaryotic parasites. *T. gondii* is responsible for encephalitis in immunocompromised individuals and birth defects in the offspring of infected mothers. The genetic tractability of *T. gondii* makes it a useful model for the study of other apicomplexan parasites.¹ The life cycle of *T. gondii* is complex with multiple differentiation steps that are critical to parasite survival in intermediate and final hosts.¹ It has been described that the gene expression profiles change dramatically during cell cycle and differentiation.^{2,3} Nonetheless, the molecular mech-

anisms that control gene expression are still poorly understood in *T. gondii* and other apicomplexan parasites. The identification of numerous enzymes that modify or remodel chromatin and the discovery of a specific set of chromatin marks at the promoters of expressed genes suggest an important role for chromatin structure in *T. gondii* gene regulation.⁴⁻⁶ However, the identification of a novel family of plant-like transcription factors together with DNA motifs required for the transcription of genes indicates that the control of transcription may also be important for the regulation of gene expression in apicomplexan parasites.^{7,8} For instance, it has been described that *Plasmodium* sexual development relies on

transcriptional and translational control of specific genes, indicating that both mechanisms are important for protein expression.^{9,10} Little is known about translational control of gene expression in *T. gondii* although several proteins deduced from the parasite genome are potential factors for translational regulation. Furthermore, discrepancies between the transcriptome and the proteome have been reported for *T. gondii* tachyzoites, suggesting that translational control is likely pertinent to gene expression in the parasite.¹¹ For instance, it has been described that even though the transcript level corresponding to *bsr4* gene was shown to be equally abundant in both rapidly replicating tachyzoites and dormant bradyzoites, the *bsr4* protein is up-regulated only in bradyzoites.¹² It seems that *T. gondii* is able to shut down the global translation of proteins after stress-induced phosphorylation of the eukaryotic initiation factor eif2a.¹³ It has been postulated that this mechanism of translational control through phosphorylation of eif2a may be particularly important for extracellular survival and parasite differentiation.¹³ In addition, the recent discovery of a large repertoire of small RNA species in tachyzoites indicates that translational control may be underestimated as a mechanism that regulates gene expression in *T. gondii*.¹⁴

In an effort to discover proteins that are potentially involved in the regulation of gene expression in *T. gondii*, two proteins containing a domain similar to Alba proteins of archaeobacteria were identified.¹⁵ Proteins containing an Alba domain have been first described in the crenarchaeal genus but are also present in eukaryotes including apicomplexan parasites.¹⁶ In archaea, Alba proteins are major components of chromatin and mediate transcriptional repression through interactions with nucleic acids. This interaction requires lysine acetylation for protein dimerization.^{17,18} However, Alba proteins in archaea are also known to bind RNA *in vivo*.¹⁹ Crystal structures of Alba proteins revealed that their fold is similar to prokaryotic translation initiation factor 3 and other RNA binding proteins.¹⁶ In *Plasmodium berghei*, three proteins containing an Alba domain were co-purified with a translation repression complex that is crucial for the sexual cycle in mosquito.²⁰ Similarly, Alba proteins have also been reported to be involved in the control of translation initiation and differentiation in *Trypanosoma brucei*.^{21,22} In contrast, *Plasmodium falciparum* Alba proteins have been shown to be able to bind both DNA and RNA.^{23,24} Altogether, these studies reveal that Alba proteins may have evolved to differentially bind DNA and RNA. In this report, we provide evidence that *T. gondii* Alba proteins define a new eukaryotic RNA binding protein family that is involved in the translational regulation of specific transcripts and may also have an important role in parasite differentiation.

Results

Molecular characterization of TgAlba proteins

T. gondii Alba proteins were originally identified in a screen for candidate factors involved in the regulation of gene expression.¹⁵ TgAlba2 (TGME49_018820) is a small 148-amino-acid protein while TgAlba1 (TGME49_021380) defines a longer protein (247 amino acids) containing the Alba domain at its N-terminus (Fig. S1). TgAlba2 belongs to the Rpp20/Pop7 subfamily, and TgAlba1, which possesses a long carboxy-terminal tail of ~100 amino acids composed of imperfect RGG repeats, belongs to the Rpp25/Pop6 subfamily.

Recombinant glutathione *S*-transferase fusion proteins were produced, purified, and used to raise polyclonal antibodies to better characterize TgAlba1 and TgAlba2 proteins. Using cytoplasmic and nuclear protein extracts from *T. gondii* tachyzoites and Western blots, we showed that two proteins of 15 and 35 kDa were recognized by the antibodies specific to TgAlba2 and TgAlba1, respectively (Fig. 1a). As expected, the glycolytic enzyme lactate dehydrogenase LDH1^{25,26} was mostly detected in the cytoplasmic fraction of the parasite, and the *T. gondii* nuclear factor TgNF3¹⁵ was mainly found in the nuclear fraction. In intracellular parasites, TgAlba1 and TgAlba2 were mainly detected in the cytoplasmic fractions (Fig. 1a, left and upper panels). In contrast, TgAlba1 and TgAlba2 are significantly present in the nuclear fraction of extracellular parasites (Fig. 1a, left and lower panels), suggesting that these proteins are enriched in the two subcellular compartments according to the extracellular or intracellular status of the parasites. We confirmed the nuclear/cytoplasmic localization for TgAlba1 and TgAlba2 by indirect immunofluorescence assays (IFAs) (Fig. 1b and c), with a predominant cytoplasmic *versus* nuclear signal in intracellular parasites. We noticed that TgAlba1 and TgAlba2 proteins showed a clear perinuclear labeling with granular signals around the nuclei of extracellular parasites (Fig. 1c).

It has been previously described that RNA granules were present in extracellular parasites upon induction by various stresses.²⁷ We found that both TgAlba1 and TgAlba2 proteins co-localized with the RNA granules (Fig. 1d and e). Interestingly, upon induction with high salt buffer, TgAlba1 (Fig. 1d) and TgAlba2 (Fig. 1e) signals are concentrated in discrete foci in extracellular parasites. We conclude that TgAlba proteins co-localize with RNA granules in extracellular parasites of *T. gondii* under stress conditions.

TgAlba1 affects response to alkaline stress *in vitro*

In order to characterize the biological function of these proteins, we performed a knock-out of each

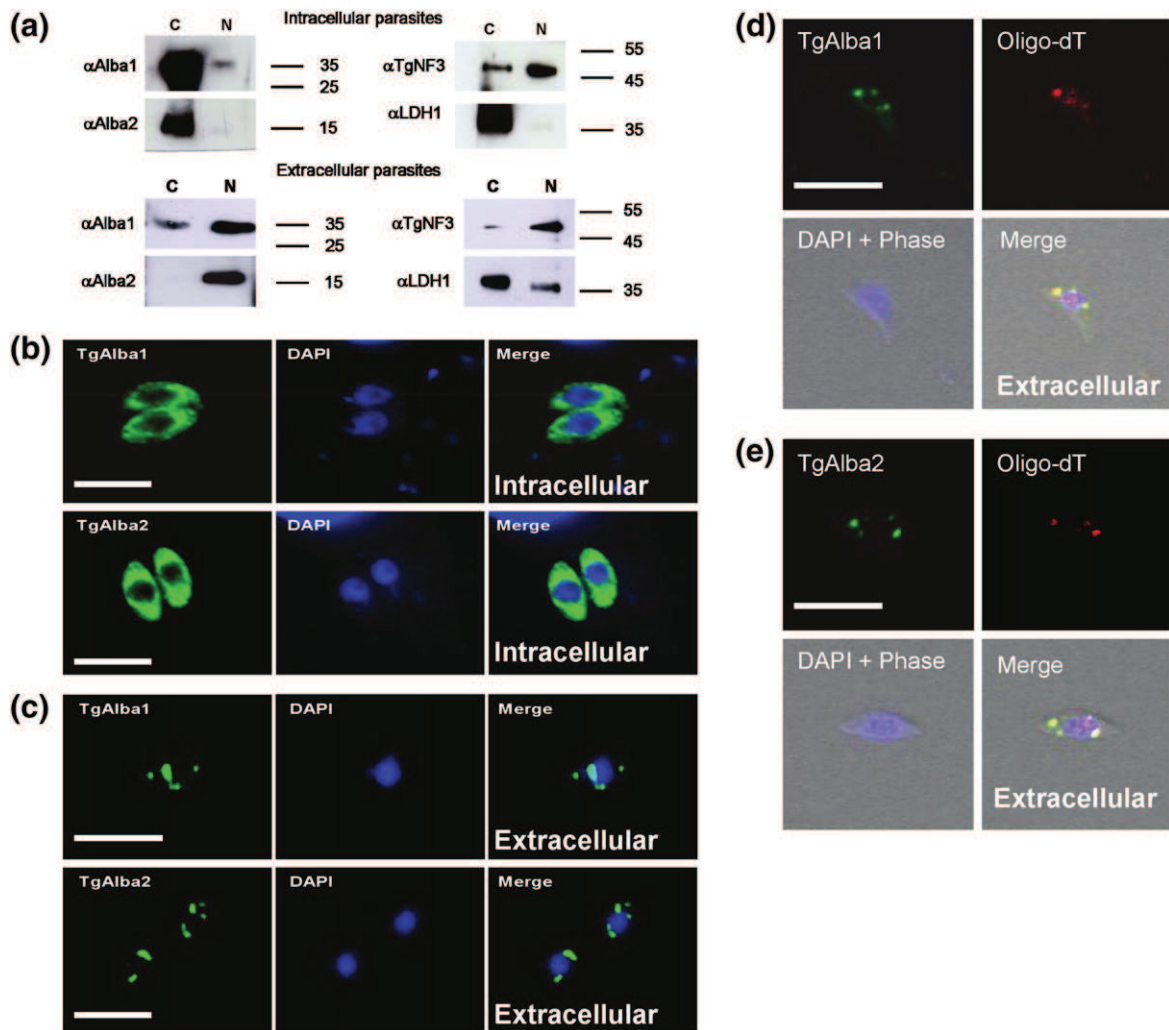


Fig. 1. Characterization of *T. gondii* Alba proteins. (a) Western blots of nuclear and cytoplasmic extracts of RH Δ Ku80 parasites. *Upper panel:* Cytoplasmic and nuclear fractions were harvested from intracellular parasites and subjected to Western blot using the anti-TgAlba1 and anti-TgAlba2 antibodies that revealed two single bands, respectively. Molecular mass markers are indicated on the right. The same cytoplasmic and nuclear fractions were subjected to Western blots using the anti-TgNF3 (nuclear marker) and anti-LDH1 (cytoplasmic marker) antibodies used as positive controls for protein distribution after subcellular fractionation. *Lower panel:* Cytoplasmic and nuclear fractions were harvested from extracellular parasites and subjected to Western blot using the anti-TgAlba1 and anti-TgAlba2 antibodies. Molecular mass markers are displayed on the right (in kilodaltons). The same cytoplasmic and nuclear fractions were subjected to Western blots using the anti-TgNF3 (nuclear marker) and anti-LDH1 (cytoplasmic marker) antibodies as described above. (b) TgAlba1 and TgAlba2 localization in intracellular parasites. IFA on paraformaldehyde-fixed intracellular parasites of the RH Δ Ku80 strain using the anti-TgAlba1 or TgAlba2 antibodies and staining of the nuclear DNA with DAPI. TgAlba1 or TgAlba2 (green) has a cytoplasmic and perinuclear localization. DAPI (blue) was used to stain the nuclear DNA. The scale bar represents 5 μ m. (c) TgAlba1 and TgAlba2 localization in extracellular parasites. IFA on paraformaldehyde-fixed extracellular parasites of the RH Δ Ku80 strain using the anti-TgAlba1 or TgAlba2 antibodies and staining of the nuclear DNA with DAPI. TgAlba1 or TgAlba2 (green) has a perinuclear localization. DAPI (blue) was used to stain the nuclear DNA. The scale bar represents 5 μ m. (d) Co-localization of RNA granules with the TgAlba1 protein in extracellular parasites. RNA FISH was performed on extracellular parasites after 1 h incubation in high [KCl] buffer. Extracellular parasites of the Pru Δ Ku80 strain were hybridized with Cy3-labeled oligo-dT primers (red), subjected to IFA with the anti-TgAlba1 antibody (green) and nuclear DNA staining with DAPI (blue). The DAPI image is merged with the phase contrast image. RNA granules and the TgAlba1 protein are partially co-localized. The scale bar represents 5 μ m. (e) Co-localization of RNA granules with the TgAlba2 protein in extracellular parasites. RNA FISH was performed on extracellular parasites after 1 h incubation in high [KCl] buffer. Extracellular parasites of the Pru Δ Ku80 strain were hybridized with Cy3-labeled oligo-dT primers (red), subjected to IFA with the anti-TgAlba2 antibody (green) and nuclear DNA staining with DAPI (blue). The DAPI image is merged with the phase contrast image. RNA granules and the TgAlba2 protein are shown to be partially co-distributed. The scale bar represents 5 μ m.

gene in the Δ Ku80 RH (type I) and the Δ Ku80 Pru (type II) genetic backgrounds. In addition, we complemented the Pru Δ Ku80 Δ Tgalba1 strain by inserting an HA-tagged copy of the Tgalba1 gene into the *uprt* locus. It was shown previously that disruption of the *uprt* locus in the Pru Δ Ku80 strain did not affect virulence or differentiation.²⁸ The intracellular parasites of these knock-out mutants displayed similar growth rates at normal pH 7.0 (Fig. S2a) and at alkaline pH 8.2 culture conditions (Fig. S2b) in infected human foreskin fibroblast (HFF) cells. In addition, there is no difference in their ability to kill infected mice (Fig. S2c). In contrast, when the extracellular parasites of these Δ Ku80 RH TgAlba1 and TgAlba2 mutants were incubated with culture media at alkaline pH 8.2, we found that their ability to invade and grow in HFF cells was significantly impaired, as assessed by *in vitro* plaque assays (Fig. S2d). These data suggest that the Δ Ku80 RH

Δ Tgalba1 and Δ Ku80 RH Δ Tgalba2 strains are less able to recover from alkaline pH stress than the parental parasites.

Alkaline pH is known to induce bradyzoite differentiation but the type I RH Δ Ku80 strain is a poor model to examine this aspect of the *T. gondii* life cycle. We therefore measured the ability of the type II Pru Δ Ku80 Δ Tgalba1 and Δ Tgalba2 strains to grow and produce bradyzoite markers *in vitro* (Fig. 2). First, we measured the ability of the parental, Pru Δ Ku80 Δ Tgalba1, Pru Δ Ku80 Δ Tgalba1 Δ uprt::Tgalba1-HA (Pru Δ Alba1:Comp), and Δ Tgalba2 intracellular parasites to grow under alkaline pH conditions (Fig. 2a). While parasites grew at a similar pace under normal conditions (Fig. S3), the Pru Δ Ku80 Δ Tgalba1 strain was able to replicate much faster than the parental, Pru Δ Ku80 Δ Tgalba1 Δ uprt::Tgalba1-HA, or Δ Tgalba2 strains under alkaline pH stress (Fig. 2a). We performed

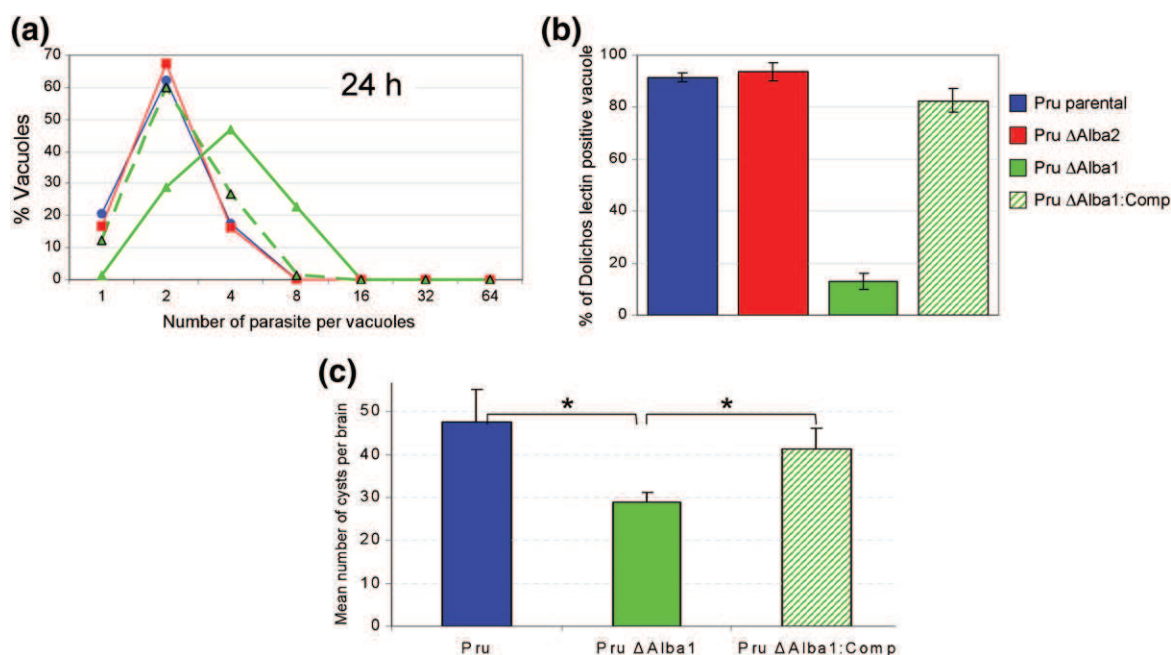


Fig. 2. The Pru Δ Ku80 Δ Tgalba1 strain is unresponsive to alkaline pH stress. (a) Growth assay under alkaline stress condition at 24 h. Parasites were incubated on coverslips in normal culture conditions for 3 h and then subjected to alkaline pH stress (pH 8.2) for 24 h. The number of parasites per vacuoles was scored for a minimum of 200 vacuoles. The number of vacuoles of any given size (number of parasites per vacuole) is expressed as the percentage of the total number of vacuoles. Parasites of Pru Δ Ku80 parental, Δ Tgalba1, Δ Tgalba1 Δ UPRT::Tgalba1-HA complemented (labeled Δ Tgalba1:Comp), and Δ Tgalba2 strains were fixed and stained after 24 h treatment. The Pru Δ Ku80 parental (blue continuous line), Δ Tgalba2 (red continuous line), and Δ Tgalba1 Δ UPRT::Tgalba1-HA complemented (green broken line, labeled Δ Tgalba1:Comp) strains grew slower in response to pH stress than the Δ Tgalba1 (green continuous line) strain. A representative experiment is shown. (b) Cyst wall detection with *D. biflorus* lectin after 48 h alkaline pH treatment. Parasites were grown under alkaline pH conditions for 48 h and subjected to labeling of the cyst wall using the *D. biflorus* lectin. The percentage of vacuoles with positive lectin staining is shown. The Pru Δ Ku80 parental (blue), Δ Tgalba2 (red), and the complemented Δ Tgalba1 Δ UPRT::Tgalba1-HA (striped green) strains have similar number of positive vacuoles while the Δ Tgalba1 (green) positive vacuoles are greatly reduced. Data are represented as mean \pm SD. (c) Number of cysts in mouse brains. Parasites from the Pru Δ Ku80 parental (blue), Δ Tgalba1 (green), and the complemented Δ Tgalba1 Δ UPRT::Tgalba1-HA (striped green) strains were injected into Balb/c mice. After 5 weeks, the brains were collected and the number of cysts per brain was scored. Data are represented as mean \pm SD.

RNA FISH to check if RNA granule formation was affected in the $\Delta Tgalba1$ and $\Delta Tgalba2$ strains and found that there was no significant difference in the number of RNA granules (Fig. S3). Because alkaline stress is a potent inducer of *in vitro* bradyzoite differentiation, we measured the ability of these strains to produce the cyst wall using the *Dolichos biflorus* lectin known to specifically stain this material. After 48 h of alkaline treatment, around 90% of the parental and Pru $\Delta Tgalba2$ vacuoles are labeled by the lectin whereas only 20% are labeled in the Pru $\Delta Tgalba1$ strain (Fig. 2b). These data indicate that the Pru $\Delta Tgalba1$ parasites did not respond properly to alkaline stress. This conclusion is supported by the fact that labeling of *D. biflorus* lectin was restored to 80% with the complemented strain (Fig. 2b). Finally, we showed that Pru $\Delta Tgalba1$ produced fewer cysts than the parental and the complemented strain in the brains of mice that were chronically infected for 1 month (Fig. 2c).

Proteins co-purified with TgAlba1 and TgAlba2 are potentially involved in translational regulation

To gain further insight into the function of the putative multi-complex proteins involving TgAlba1 and TgAlba2, we genetically modified $\Delta Ku80$ RH parasites to express the TgAlba proteins with the C-terminal tandem affinity purification (TAP) tag under the control of their endogenous promoters and at their own loci. These genetically engineered strains, TgAlba1-TAP and TgAlba2-TAP, were used to purify and identify the protein complex associated with each protein. The affinity-purified proteins were analyzed by SDS-PAGE followed by silver staining and mass spectrometry (MS) (Fig. S4). We confirmed the co-purification of numerous proteins. Among these proteins, we estimated that 40 proteins were commonly associated to TgAlba1 and TgAlba2 (Tables S1 and S2, respectively; also see Fig. 3c). Strikingly, most of these proteins are related to translational regulation of gene expression. A TgAlba interacting protein containing a single-strand DNA-binding domain, found in the SsossB archeal protein,²⁹ was also identified and named TgSsossB (TGME49_048810). We engineered a strain expressing a TAP-tagged TgSsossB protein and confirmed the physical interaction between TgAlba proteins and TgSsossB protein by affinity chromatography (Fig. 3a). It should be noted that TgAlba1 and TgAlba2 were not co-purified with TgSsossB when the parental strain was used as a negative control (Fig. 3a). Using specific anti-TgAlba1 antibodies and Western blots, we found that the TgAlba1 protein that co-immunoprecipitated by TgSsossB ran slower than the major TgAlba1 protein pulled down by TgAlba2-TAP (see

arrows). However, it seems that a faint protein band corresponding to the slower running TgAlba1 protein can also be detected in the co-immunoprecipitate of TgAlba2-TAP (star and dashed arrow). These data suggest that this novel TgAlba1 form is likely enriched in the TgSsossB-interacting proteins (Fig. 3a). These data also suggest that TgAlba1 may be modified post-translationally and that this modified version of the protein is mainly enriched in the TgSsossB co-purified proteins. However, the nature of this putative post-translational modification remains to be determined.

The localization of the HA-tagged version of TgSsossB was also similar to that of TgAlba1 and TgAlba2 in both intracellular and extracellular parasites (Fig. 3b). This co-distribution indicates that TgSsossB might form a complex with TgAlba proteins. Finally, we analyzed the other proteins that co-purified with TgSsossB by MS (Table S3) and found 24 proteins in common with the TgAlba1 and TgAlba2 co-purified proteins (Fig. 3c). Among the protein partners shared by TgAlba and TgSsossB are the cap and the polyA binding proteins.

The transcript of Tgalba2 is increased in the $\Delta Tgalba1$ strains but is no longer translated

RT-PCR confirmed that $\Delta Tgalba1$ and $\Delta Tgalba2$ mutants in both genetic backgrounds lack the *Tgalba1* and *Tgalba2* transcript, respectively (Fig. 4a and c), whereas parental and complemented strains expressed both transcripts, as expected (Fig. 4a and c). We found that the *Tgalba2* transcript was overexpressed in the $\Delta Tgalba1$ mutant in both genetic backgrounds when compared to the parental strain as measured by microarray and quantitative real-time PCR (qRT-PCR) (Fig. S5). We double-checked expression of TgAlba1 and TgAlba2 proteins in the knock-out and parental parasites by Western blot (Fig. 4b and d). As expected, both TgAlba1 and TgAlba2 proteins were undetectable in $\Delta Tgalba1$ and $\Delta Tgalba2$ strains, respectively. Surprisingly, the TgAlba2 protein was undetectable in $\Delta Tgalba1$ strains (Fig. 4b and d, lower panel) despite the expression of the *Tgalba2* transcript in these strains (Fig. 4a and c). In contrast, the TgAlba1 protein was detected in the $\Delta Tgalba2$ strain (Fig. 4b and d, middle panel). It should be noted that Western blot detection of the *T. gondii* housekeeping actin protein ensured equal loading of protein for comparing parental parasites and knock-out mutants (Fig. 4b and d, upper panel). As shown in Fig. 4d, Western blot using an anti-HA antibody (right panel) demonstrates that the HA-tagged copy of the TgAlba1 protein was efficiently expressed in the Pru $\Delta Ku80$ $\Delta Tgalba1$ strain. This was sufficient to restore expression of the endogenous TgAlba2 protein (Fig. 4d, left-bottom panel).

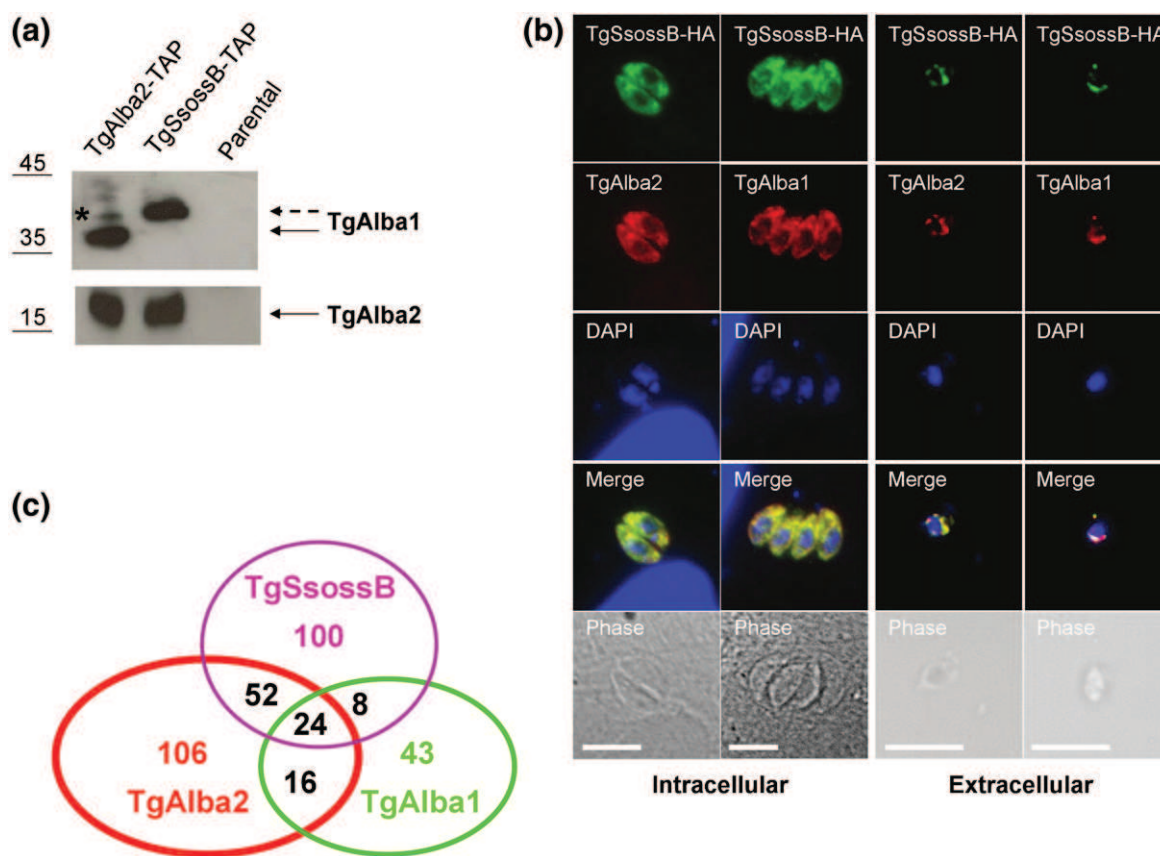


Fig. 3. Analysis of the proteins co-purified with the TgAlba complex. (a) TgAlba1 and TgAlba2 are co-purified with TgSsossB. To confirm the proteomic analysis of the co-purified proteins associated with TgAlba1 and TgAlba2, we purified a TAP-tagged version of TgSsossB. As controls, we used TgAlba2-TAP co-purified protein sample and a protein sample purified with the same strategy from the parental strain protein extracts (labeled “Parental”). Protein samples were then subjected to Western blots using the anti-TgAlba2 and anti-TgAlba1 antibodies. Both TgAlba proteins (arrowed) are pulled down with TgSsossB. The anti-TgAlba1 antibody detected two bands of different size in the purified TgAlba2-TAP and in the TgSsossB-TAP samples. One band is only present in the TgAlba2-TAP sample (solid arrow). The other TgAlba1 form is present in the TgSsossB-Tap sample (dashed arrow) and in the TgAlba2-TAP sample (star). Molecular markers are shown on the left side (in kilodaltons). (b) TgSsossB co-localizes with TgAlba1 and TgAlba2. The RH Δ Ku80 strain expressing an HA-tagged version of TgSsossB was subjected to IFA using an anti-HA (green), anti-TgAlba1 (red), or anti-TgAlba2 (red) antibodies together with nuclear DNA staining by DAPI (blue). In both intracellular (left panel) and extracellular (right panel) parasites, TgAlba1 or TgAlba2 co-localized with TgSsossB. The scale bar represents 5 μ m. (c) Venn diagram representing the proteins associated in each complex. Proteins identified as partners of TgAlba2 (red), TgAlba1 (green), and TgSsossB (pink) were compared and a core of 24 proteins is shared by each complex including TgAlba1 and TgSsossB.

TgAlba1 and TgAlba2 bind RNA *in vivo*

In order to confirm the direct involvement of the TgAlba complex in regulating the translation of *TgAlba* transcripts, we performed an immunoprecipitation of TgAlba proteins and harvested the total RNA associated with this complex in the RH parental strain. We then tested for the presence of *Tgalba* transcripts as well as two negative controls (*eno2* and *tubulin*) using qRT-PCR on DNase-treated samples. We showed that both proteins were able to pull down the two different *Tgalba* transcripts (Fig. 5a). In contrast, the transcripts of

tubulin and *eno2* were not enriched in the immunoprecipitated samples (Fig. 5a). As expected, none of these transcripts were enriched using the negative control (non-immune sera) (Fig. 5a). In order to identify transcripts that bind to the TgAlba protein complex, we hybridized the pull-down RNAs to the ToxoGeneChip microarray and compared it to the entire pool of RNA expressed at the tachyzoite stage. As a negative control, we also hybridized RNA pulled down using the non-immune sera. When compared to the transcriptome of the parental RH strain at the tachyzoite stage and after removing the genes present in both samples, we identified a list of

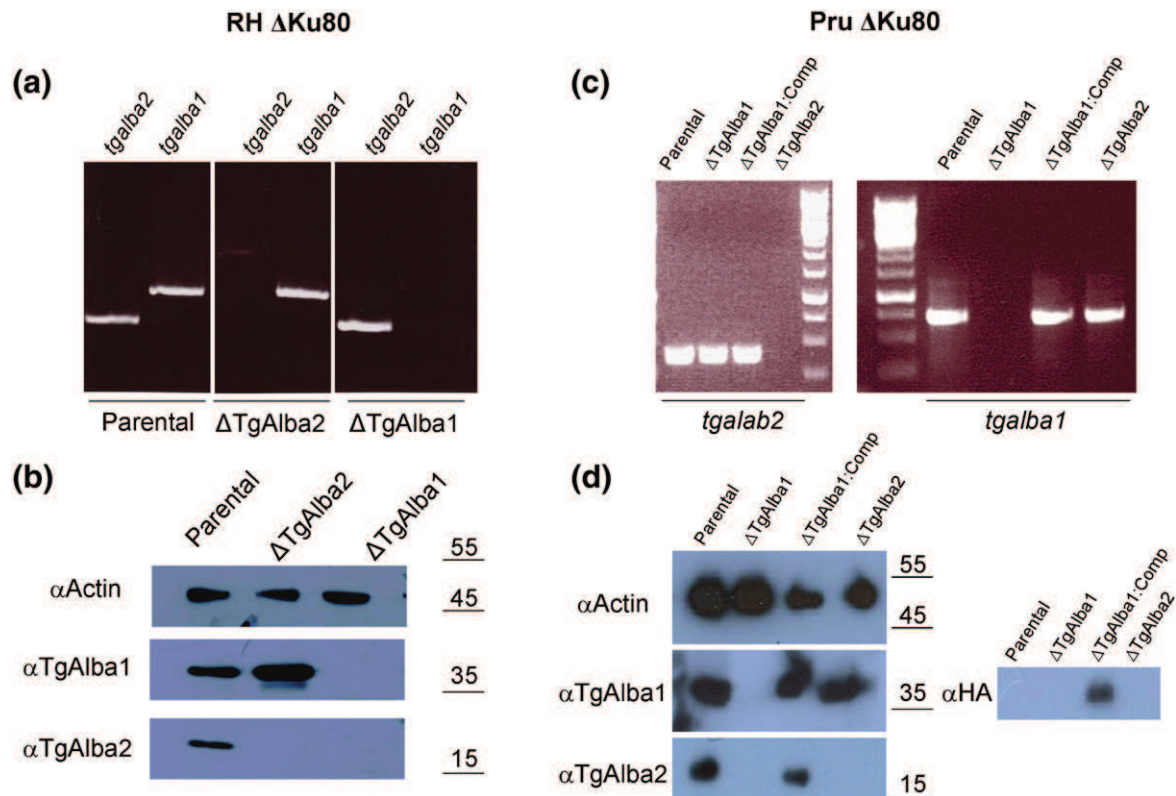


Fig. 4. Characterization of the Δ TgAlba2 and Δ TgAlba1 strains in the RH and Pru Δ Ku80 strains. (a) RT-PCR analysis of *TgAlba1* and *TgAlba2* transcript expression in the parental, Δ TgAlba2, and Δ TgAlba1 RH Δ Ku80 strains. RNA from intracellular parasites was extracted and subjected to RT-PCR using primers specific for the *TgAlba2* and *TgAlba1* transcripts in the parental, Δ TgAlba2, and Δ TgAlba1 strains. In the parental strain, both transcripts are detectable. In the Δ TgAlba2 strain, we were not able to detect the *TgAlba2* transcript but *TgAlba1* transcript was present, as expected. In the Δ TgAlba1 strain, we were not able to detect the *TgAlba1* transcript but the *TgAlba2* transcript was present, as expected. (b) The TgAlba2 protein is undetectable in the RH Δ TgAlba1 strain. Western blot analyses of the TgAlba2, TgAlba1, and actin proteins in the parental RH Δ Ku80, Δ TgAlba2, and Δ TgAlba1 strains. The TgAlba2 protein is undetectable in the Δ TgAlba1 strain. Equal numbers of parasites were loaded on a polyacrylamide gel and subjected to Western blot analyses using the anti-actin (top panel), anti-TgAlba1 (middle panel), and anti-TgAlba2 (bottom panel) antibodies. Actin is used as a loading control and shows little differences between the parental, Δ TgAlba2, and Δ TgAlba1 strains. Molecular markers are shown on the right side (in kilodaltons). (c) RT-PCR analysis of *Tgalba1* and *Tgalba2* transcript expression in the parental Pru Δ Ku80, Δ TgAlba2, and Δ TgAlba1 strains. RNA from intracellular parasites was extracted and subjected to RT-PCR using primers specific for the *Tgalba2* and *Tgalba1* transcripts in the parental Pru Δ Ku80, Δ TgAlba2, Δ TgAlba1, and Δ TgAlba1 Δ UPRT::TgAlba1-HA complemented (labeled Δ TgAlba1:Comp) strains. In the parental and the complemented strains, both transcripts are detectable. In the Δ TgAlba2 strain, we were not able to detect the *TgAlba2* transcript but the *TgAlba1* transcript was present, as above. In the Δ TgAlba1 strain, we were not able to detect the *TgAlba1* transcript but the *TgAlba2* transcript was present, as above. (d) The TgAlba2 protein is undetectable in the Pru Δ TgAlba1 strain. Western blot analyses of the TgAlba2, TgAlba1, and actin proteins in the Pru Δ Ku80 parental, Δ TgAlba1, Δ TgAlba1 Δ UPRT::TgAlba1-HA complemented (labeled Δ TgAlba1:Comp), and Δ TgAlba2 strains. The TgAlba2 protein is undetectable in the Δ TgAlba1 strain as for the Δ TgAlba2 strain generated in the RH Δ Ku80 strain, as described above. Actin is used as a loading control. The anti-actin (top panel), anti-TgAlba1 (middle panel), and anti-TgAlba2 (bottom panel) antibodies were used. The anti-HA antibodies (right panel) detected the TgAlba1-HA-tagged copy in the Δ TgAlba1 Δ UPRT::TgAlba1-HA complemented strain. Molecular markers are shown on the right side (in kilodaltons).

30 transcripts including *TgAlba1* (Table S5). We found that most of these transcripts encode small proteins with a mean size of 22 kDa. We verified by qRT-PCR that three transcripts were indeed pulled down by the TgAlba complex (Fig. 5a). TGME49_047580 encodes for a protein containing a glutaredoxin domain, generally implicated in redox

homeostasis while TGME49_019170 encodes for a protein of unknown function. TGME49_110400 encodes for a homolog of glycogenin, a protein involved in carbohydrate biosynthesis. TGME49_047580 and TGME49_019170 transcripts were significantly enriched in the TgAlba1 and TgAlba2 pulled-down RNAs but the enrichment

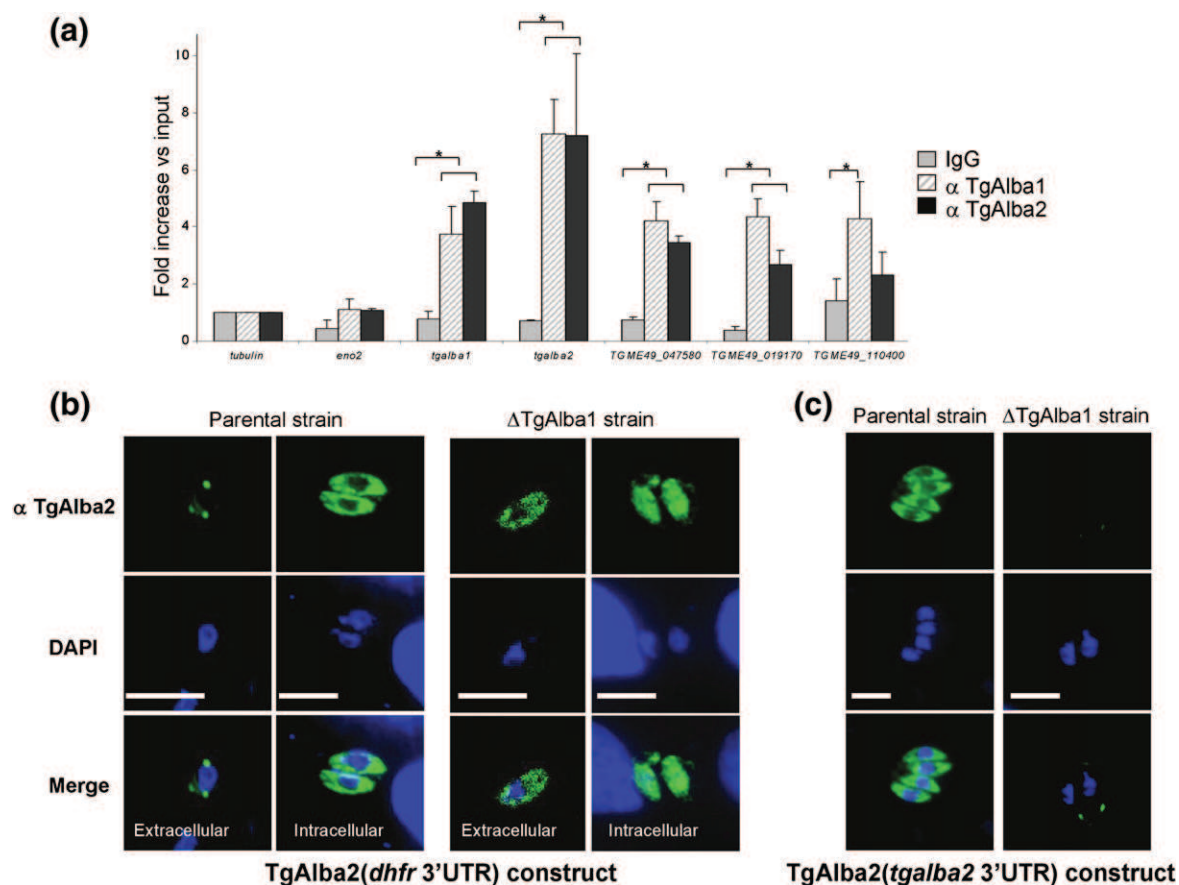


Fig. 5. Alba proteins are involved in gene-specific translation. (a) TgAlba1 and TgAlba2 proteins specifically bind transcripts *in vivo*. RNA pull down was performed using irrelevant IgG (gray bars), the anti-TgAlba1 antibody (black bars), and anti-TgAlba2 antibody (striped bars). qRT-PCR was performed using primers representing the following genes: *tubulin*, *eno2*, *TgAlba2*, *TgAlba1*, *TGME49_047580*, *TGME49_019170*, and *TGME49_110400*. The last four transcripts were identified as mRNAs that bind to the TgAlba complex after hybridization of the pull-down RNA onto microarrays. A Student *t* test was significant ($p < 0.05$) for the comparisons indicated with an asterisk (*). Three independent experiments were performed. Data are represented as mean \pm SD. (b) TgAlba2 expression depends on the presence of its own 3'UTR and the TgAlba1 protein. The *TgAlba2* locus was modified in the RH Δ Ku80 parental and Δ TgAlba1 strains by replacing the endogenous *TgAlba2* 3'UTR with the 3'UTR of the *TgDHFR* gene. Using this knock-in strategy, TgAlba2 is expressed under the control of its own promoter and of the *TgDHFR* 3'UTR. In the genetically engineered parental parasites, TgAlba2 localization was unchanged in extracellular (perinuclear granular staining) or intracellular (cytoplasmic and excluded from the nucleus) parasites. In the Δ TgAlba1 strain, TgAlba2 localized to the cytoplasm of extracellular parasites and to the cytoplasm and the nucleus of the intracellular parasites. TgAlba2 is identified using the mouse anti-TgAlba2 antibody (green). Nuclei are stained with DAPI (blue). The scale bar represents 5 μ m. (c) TgAlba2 expression depends on the presence of its own 3'UTR and the TgAlba1 protein. The *TgAlba2* locus was modified in the RH Δ Ku80 parental and Δ TgAlba1 strains reintroducing the 3'UTR of the *TgAlba2* gene. Using this knock-in strategy, TgAlba2 is expressed under the control of its own promoter and 3'UTR. In the parental parasites, TgAlba2 localization and expression are unchanged. In the Δ TgAlba1 strain, the TgAlba2 protein is undetectable. TgAlba2 is identified using the anti-TgAlba2 antibody (green). Nuclei are stained with DAPI (blue). The scale bar represents 5 μ m.

was only significant in the TgAlba1 pulled-down RNAs for the *TGME49_110400* transcript.

Translation of the *TgAlba2* transcript depends on the presence of the TgAlba1 protein

The absence of the TgAlba2 protein in the Δ TgAlba1 strain may be explained by (i) the dependence of the *TgAlba2* transcript on the TgAlba1

protein for its translation or (ii) the high instability of the protein. To clarify this point, we introduced the *tgAlba2* gene in the Δ TgAlba1 RH strain. In this experiment, the open-reading frame of the *tgAlba2* gene fused to YFP was expressed under the control of the *Tgtubulin* promoter and the 3' untranslated region (UTR) of the *Tgdhfr* gene. When the Δ TgAlba1 strain was transfected with the TgAlba2-YFP construct, we observed the ectopic expression of YFP-

tagged TgAlba2 (Fig. S6a), indicating that the TgAlba2 protein is stable in the absence of the TgAlba1 protein (Fig. S6a). This clearly demonstrates that the absence of TgAlba2 in the $\Delta TgAlba1$ strain is due to its inability to translate the *tgalba2* transcript and indicates that the TgAlba1 protein is necessary for the translation of the endogenous *Tgalba2* transcript. Next, the $\Delta TgAlba1$ strain was transfected with the TgAlba1-YFP construct, and we observed that expression of an ectopic copy of TgAlba1 is sufficient to restore the translation of the *Tgalba2* endogenous transcript (Fig. S6b). Finally, when complementing the $\Delta TgAlba2$ strain with the same vectors, YFP-tagged Alba proteins were efficiently expressed (Fig. S6c and d).

Translation of the *TgAlba2* transcript depends on the presence of its 3'UTR

In order to identify the crucial elements enabling the control of *TgAlba2* transcript translation, we performed a knock-in of the *TgAlba2* gene in the RH $\Delta Ku80 \Delta TgAlba1$ strain. In this strain, the expression of the *TgAlba2* gene is under the control of its own promoter and 5'UTR but has the 3'UTR of the *TgDHFR* gene. Accordingly, we observed that the TgAlba2 protein was expressed in the $\Delta TgAlba1$ RH strain background, indicating that the 3'UTR plays a critical role in regulating the translation of the *TgAlba2* transcript in this genetic background (Fig. 5b). However, we also noticed that TgAlba2 was mislocalized in the $\Delta TgAlba1$ strain (Fig. 5b). Indeed, TgAlba2 was localized throughout the entire parasite body, including the nucleus in intracellular parasites and in the cytoplasm of extracellular parasites (Fig. 5b). This indicates that TgAlba1 may have an important role in the proper localization of the TgAlba2 protein. To confirm the importance of the 3'UTR of the *TgAlba2* transcript in its expression, we performed a knock-in of the *TgAlba2* gene using the same vector (as in Fig. 5b), but we replaced the 3'UTR of the *TgDHFR* gene with that of *TgAlba2* itself. In the parental strain, transfection of the tagged version of TgAlba2 under the control of its own 3'UTR resulted in normal expression and correct localization of the protein (Fig. 5c). In contrast, when transfected in the $\Delta TgAlba1$ strain, the vector replacing the endogenous copy by a tagged version of TgAlba2 under the control of its own 3'UTR was integrated at the right loci but expression of the TgAlba2 protein could not be detected (Fig. 5c). These data also indicate that the 3'UTR of *TgAlba2* plays a crucial role for its correct translation.

Discussion

Little is known about the role of translational control in *T. gondii* gene regulation. We identified

two *T. gondii* proteins containing an Alba domain. Crystal structures of archeal Alba proteins revealed that their folds are similar to the prokaryotic translation initiation factor 3 and other RNA binding proteins.¹⁶ In human and yeast, the Alba proteins are nuclear and play a role in tRNA processing.^{30,31} In contrast, the two TgAlba proteins of *T. gondii* described herein have both an intriguing cytoplasmic localization in intracellular parasites and peculiar perinuclear foci in extracellular parasites. In extracellular parasites, we showed that the TgAlba proteins within these perinuclear foci co-localize with the markers of the endoplasmic reticulum and RNA granules. In both intracellular and extracellular parasites, TgAlba proteins may therefore be able to contact transcripts.

Noteworthy, *T. brucei* Alba proteins were also shown to co-localize with a subset of poly(A+) RNA in stress granules while these proteins localize to the cytoplasm in normal culture conditions.^{22,23} In *P. falciparum*, the Alba proteins were shown to localize to perinuclear foci in ring stages and also expand to the cytoplasm in mature forms.²⁵

In order to gain insights into the biological function of these proteins in *T. gondii*, we performed a knock-out in the $\Delta Ku80$ type I and type II background strains. The $\Delta TgAlba \Delta Ku80$ type I mutants displayed an increased sensitivity when extracellular parasites were exposed to alkaline pH. This phenotype is very similar to that of the *T. gondii* *eif2a* mutant previously described.¹³ However, in order to better assess the ability to respond to external stress conditions, we also performed a targeted gene disruption of *Tgalba* genes in the type II $\Delta Ku80$ strain. Because these $\Delta TgAlba1$ mutants are not expressing the TgAlba1 or the TgAlba2 protein, we conclude that both proteins are likely to be important for the response to stress signals. The TgAlba2 protein is undetectable in $\Delta TgAlba1$ strains despite the elevated levels of the *TgAlba2* transcript in both type I and type II background. We conclude that there is a strong link between the presence of the TgAlba1 protein and the ability to translate the *TgAlba2* transcript. This also suggests that the TgAlba complex may act as a positive regulator for the translation of the *TgAlba2* transcript.

It is interesting to note that in other organisms, the abrogation of one Alba protein impacts the expression of other Alba proteins. For example, in *T. brucei*, the knock-down of the TbAlba3 transcript resulted in the down-regulation of the TbAlba1 and TbAlba2 protein expression.²¹ The absence of the TgAlba2 protein in the $\Delta TgAlba1$ strain may be explained by either the dependence of the *Tgalba2* transcript on the TgAlba1 protein for its translation or the high instability of the protein. By replacing the 3'UTR of the *tgalba2* gene with unrelated DNA sequence, we showed that the TgAlba2 protein is expressed and stable in the absence of the TgAlba1 protein. In

these experiments, we show that the translation of the *TgAlba2* transcript relies on the presence of its cognate 3'UTR and of the TgAlba1 protein.

To further explore the potential role of the TgAlba complex in translational regulation, we purified the proteins interacting with TgAlba1 and TgAlba2. We identified 40 proteins that were common in the two sets of purified proteins. The poly-A binding protein (TGME49_024850) was present in the partners that were co-purified with both TgAlba proteins. In *T. brucei*, the poly-A binding protein was also shown to interact with Alba proteins.²¹ We also identified the cap binding protein (TGME49_049610), two subunits of eif3 (TGME49_073460 and TGME49_117720), and a protein containing a single-strand DNA-binding OB fold (TGME49_048810, TgSsoSB).

In archaea, SsoSB interacts with RNA polymerase and is able to counteract the action of Alba proteins on chromatin.³² A reverse-immunoprecipitation experiment showed that TgSsoSB is able to interact with both TgAlba proteins. We hypothesized that the known role of proteins specifically interacting with either TgAlba protein may reflect the role of the respective TgAlba complexes.

For instance, the TgAlba2 protein may interact with a large complex of proteins including 10 of the 12 eIF3 subunits of *T. gondii*, as well as 10/12 subunits of the chaperonin complex and 9/16 subunits of the regulatory subunits of the 26S proteasome. In yeast, a complex called the "translasome" has been characterized and is also composed of eIF3, elongation factors, tRNA synthetases, 40S and 60S ribosomal proteins, chaperones, and the proteasome.³³ These data indicate that *T. gondii* may also assemble a multi-complex protein to regulate translation initiation.

In contrast, protein partners that co-purified with TgAlba1 include RNA binding proteins and may therefore recruit RNA to be regulated by this complex. Indeed, we showed that the TgAlba complex was able to specifically pull down RNA, including the *TgAlba1* and *TgAlba2* transcripts.

Other proteins encompassing RNA binding motifs were also identified in the list of TgAlba1-interacting proteins and may participate in the recruitment of specific transcripts. In addition, TgAlba1 contains an RGG motif that might also be important for the recruitment of RNAs. We showed that the replacement of the 3'UTR of the *TgAlba2* transcript is sufficient to recapitulate its translation in the Δ TgAlba1 strain. This suggests that the TgAlba complex may bind directly or indirectly to the 3'UTR of specific transcripts even if we were not presently able to identify the specific motifs in the *TgAlba2* 3' UTR. In addition, there was no recognizable motif in the 3'UTR of the 30 other transcripts identified by microarray hybridization. Because numerous RNA binding proteins are co-purified with the TgAlba protein complex, it is likely that more than one RNA

binding protein is involved in recruiting the transcripts. Recently, a large repertoire of putative microRNA (miRNA) was identified in *T. gondii*, and TgAlba1 was identified as a part of the TgAGO complex, which binds to miRNA.¹⁴ TgAlba1 also interacts with a Tudor/nuclease domain-containing protein, which was also linked to TgAGO,¹⁴ suggesting that TgAlba1 may also play an alternative role in the translational repression mediated by miRNA.

Taken together, our study revealed the presence of a multi-protein complex containing TgAlba proteins, and we hypothesize that translation of specific gene transcripts in *T. gondii* may be the potential targets of this multi-protein complex. We also established that the TgAlba complex plays a role in regulating response to stress and differentiation. This may help uncover a link between translational regulation of gene expression and the mechanisms governing differentiation.

Materials and Methods

Parasite tissue culture and manipulation

Tachyzoites of *T. gondii* strain RH Δ Ku80³⁴ and Pru Δ Ku80²⁸ were propagated *in vitro* in HFFs using Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum (FCS), 2 mM glutamine, and 1% penicillin-streptomycin. Alkaline pH media were prepared with RPMI 1640 supplemented with 20 mM Hepes, 2 mM glutamine, 1% penicillin-streptomycin, and 1% FCS. Alkaline media pH was adjusted to 8.2. *T. gondii* tachyzoites were grown in ventilated tissue culture flasks at 37 °C and 5% CO₂. Transgenes were introduced by electroporation into tachyzoites of *T. gondii* strains, and stable transformants were selected by culture in the presence of 2 μ M pyrimethamine or xanthine (50 μ g/ml) and mycophenolic acid (25 μ g/ml). Clonal lines were obtained by limiting dilution. Prior to RNA and protein purification, intracellular parasites were purified by sequential syringe passage with 17-gauge and 26-gauge needles and filtration through a 3- μ m polycarbonate membrane filter.

Plaque assays and parasite growth assays

For plaque assays, extracellular parasites were incubated in the culture media either at 37 °C or at 4 °C and compared to those incubated with alkaline media (pH 8.2) at 37 °C for 2 h. Afterwards, 10³ parasites were inoculated onto confluent HFF monolayer in 24-well plates. Plaques were grown over 5 days in normal culture conditions. Fixation and staining of plaque assays were carried out using the RAL555 kit.

For proliferation assays, 8 \times 10⁵ parasites per well in a 24-well plate were incubated 3–4 h in normal media. After two phosphate-buffered saline (PBS) washes, coverslips were incubated for 24, 48, or 72 h in normal media or in alkaline pH media (pH 8.2). Fixation and staining of proliferation assays were carried out using the RAL555 kit. The number of parasites per vacuole was counted in 10 fields per slide with a minimum of 2 slides per condition.

Cyst isolation from mouse brain and enumeration

Four Balb/c mice were injected with 200 parasites in the intraperitoneal cavity. After 4 weeks, brains were collected and homogenized as previously described.²⁸ Cysts were counted after *D. biflorus* lectin labeling of the cyst wall for 30 min at room temperature in PBS. Ten slides for each brain were scored for the presence of lectin positive cysts.

DNA manipulation

TgAlba2 (TGME49_018820) and *TgAlba1* (TGME49_021380) were amplified from genomic DNA of the parental strain and cloned into the pLIC vectors,³⁴ kindly provided by Drs. Vern Carruthers (University of Michigan, USA) and Michael White (University of South Florida, USA). The pLIC TAP-tag vector was produced by amplifying the TAP-tag from the pcDNA3-NTAP vector³⁵ and cloned using the *EcoRI* and *NdeI* sites of the pLICHA vector. Replacement of the *TgDHFR* 3'UTR by the *Tgalba2* 3'UTR was performed using the *AscI* and *NdeI* sites of the pLIC-HA vector. The coding region of *TgAlba1* and *TgAlba2* was amplified by PCR and cloned into pTub-HSP60-YFP vector using *BglII* and *AvrII* sites. All PCR reactions were carried out using High-Fidelity PCR Enzyme Mix (Fermentas) according to the manufacturer's instructions. The sequences of all primers used in this study are listed in Supplementary Table S6. The 5' and 3' flanking regions of *TgAlba1* and *TgAlba2* were amplified from genomic DNA of Δ Ku80 RH *T. gondii* (type I strain). PCR products were then cloned using *Apal* and *HindIII* sites for the 5' flanking sequences and *SpeI* and *NotI* sites for the 3' flanking sequences into the pDHFR vector, which was linearized with *NotI* before transfection into Δ Ku80 RH or Δ Ku80 Pru strains.

For complementation, *Tgalba1* gene amplified from cDNA of the Pru Δ Ku80 strain was cloned in frame with an HA Tag cassette into *AscI* and *SmaI* sites of pUPRT targeting vector provided by Dr. Reese.³⁶ Subsequently, 2000 bp of the *Tgalba1* putative promoter was cloned into the *AscI* site.

Antibodies

The anti-*TgAlba2* and anti-*TgAlba1* mouse antibodies were produced after immunization of mice with a full-length recombinant version of the proteins and used at 1:500 dilutions in IFA and 1:2000 in Western blots. Anti-HA rabbit (Eurogentec) or mice (Invitrogen) were used at 1:500 in IFA and 1:2000 in Western blots. The anti-SERCA³⁷ was used at 1:500 dilutions in IFA. The anti-TgNF3¹⁵ was used at 1:3000 dilutions and the anti-LDH1 was used at 1:1000 dilutions in Western blots.

IFAs and confocal imaging

For IFAs, intracellular and extracellular tachyzoites were fixed with 4% paraformaldehyde in PBS for 15 min on ice, followed by two PBS washes. Extracellular parasites were dried on Teflon slides. Parasites were permeabilized with 0.1% Triton X-100 in PBS containing 0.1% glycine for 10 min at room temperature. Samples were blocked with

3% FCS in the same buffer and the primary antibodies were added on parasites in the same buffer for 1 h at room temperature. Secondary antibody coupled to Alexa-488 or to Alexa-633 (Molecular Probes) diluted at 1:1000 was added in addition to 4',6-diamidino-2-phenylindole (DAPI) for nucleus staining. Confocal imaging was performed with an LSM710 microscope (Zeiss) and a Plan Aplanachromat objective (Plan-Apochromat 63 \times /1.40 Oil DIC M27, Zeiss) as previously described.¹⁵

RNA FISH

Extracellular parasites were treated by a high [KCl] buffer for 1 h as previously described²⁷ and fixed with 4% paraformaldehyde for 30 min. Parasites were settled on RNA-treated slides and prepared as previously described.³⁸ Cy3-oligo d(T)₄₀ were hybridized on the slide at 50 °C overnight and washed as previously described.³⁸ After 10-min DAPI staining, slides were mounted on coverslips or processed for IFA as described above.

TAP-tag purification

Cytoplasmic and nuclear extracts of the *TgAlba1*-TAP Δ Ku80 RH, *TgAlba2*-TAP Δ Ku80 RH, and *TgSossB*-TAP strains were prepared as described above from a mixed population of intracellular and extracellular parasites. Cytoplasmic and nuclear extracts were combined for the subsequent steps. TAP-tag purification was performed as previously described³⁹ and the proteins were eluted after incubation overnight with 100 units of the AcTEV (Invitrogen) protease at 4 °C on a rotating wheel.

RNA pull down

Intracellular parasites (10⁹ parasites) were purified by sequential syringe passage with 17-gauge and 26-gauge needles and filtration through a 3- μ m polycarbonate membrane filter. Parasites were washed twice with ice-cold PBS and resuspended in 1.5 ml of buffer 1 (10 mM Hepes, pH 7.9, 1.5 mM MgCl₂, 10 mM KCl, 0.1 mM ethylenediaminetetraacetic acid, 0.65% NP40, 0.5 mM PMSF, and 40 U RNasin), incubated on ice for 10 min and centrifuged at 1500g for 10 min at 4 °C. The supernatant was then incubated with the anti-*TgAlba1*, anti-*TgAlba2*, anti-HA antibody, or irrelevant IgG overnight at 4 °C on a rotating wheel. A fraction of the supernatant was kept as the input. Pre-washed protein A beads were added to the samples and left for 1 h at 4 °C on a rotating wheel. Beads were then washed three times in buffer 1 and three times in buffer 2 (10 mM Hepes, pH 7.9, 1.5 mM MgCl₂, 10 mM KCl, 250 mM NaCl, 0.1 mM ethylenediaminetetraacetic acid, 1.1% Triton X-100, 0.5 mM PMSF, and 40 U RNasin). Beads were then resuspended in 300 μ l of Trizol (Invitrogen) and RNA was purified according to the manufacturer's instructions.

Quantitative real-time PCR

All primers were designed online using Primer 2 v.0.4.0†. The cDNA samples were synthesized from 100 ng of RNA samples using the Revert Aid First Strand cDNA Synthesis

Kit (Fermentas) after a treatment of 15 min at 37 °C with DNase (Promega). qRT-PCR was carried out on an Mx3000P System (Agilent Technologies, Palo Alto, CA). Individual reactions were prepared with 0.5 µM of each primer, ~5 ng of cDNA, and SYBR Green PCR Master Mix (Applied Biosystems, California) to a final volume of 20 µl. All experiments were performed twice with separate biological replicates. For each experiment, reactions were performed in triplicate and expression of individual genes was normalized to *Tgtubulin* Ct values.

Purification of RNA and microarray

Total RNA was purified from intracellular parasites after growth in HFF cells for 36 h. Total RNA was extracted using the NucleoSpin RNA II kit (Macherey-Nagel) according to the manufacturer's instructions. At least three independent experiments were performed. Experiments were performed starting from 0.5 or 0.1 µg (for RNA pull-down samples) of total RNA. A one-cycle labeling procedure for the synthesis of fluorescent targets was followed. This protocol allows for signal amplification due to the synthesis of a cDNA intermediate including a T7-RNA polymerase cassette, which is used to obtain *in vitro* transcription. Biotinylated ribonucleotide analogues are incorporated in target RNA during *in vitro* transcription. This biotinylated RNA is then fragmented before being hybridized on Affymetrix chip. After hybridization, the biotin is recognized by streptavidin–phycoerythrin and the signal is amplified by the presence in the labeling reaction of biotinylated anti-streptavidin antibody. All hybridizations were performed on the *Toxoplasma* GeneChip Affymetrix microarray^{40,41} as recommended by the manufacturers.

After hybridizations, the raw data were analyzed using the RMA (robust multi-array average)⁴² and LIMMA (linear models for microarray data)⁴³ packages running under the statistical language R. A normalization protocol, consisting of a within-array loess normalization to correct for dye and spatial effects, was applied on the background-corrected median or mean intensities.⁴⁴ After normalization, identification of statistically significant regulation was performed using moderated *t* statistic with empirical Bayes shrinkage of the standard errors.

Sample preparation for proteomic analysis and MS

Gel slices were treated with dithiothreitol and iodoacetamide and digested by trypsin as previously described.^{15,45} NanoLC-MS/MS analyses were performed using either a nanoACQUITY Ultra-Performance LC system (Waters, Milford, MA, USA) coupled to a Q-TOF maXis mass spectrometer (Bruker Daltonics, Bremen, Germany) or a nanoHPLC-Chip/MS (Agilent Technologies) coupled to an ion trap amaZon (Bruker Daltonics). Both systems were fully controlled by HyStar 3.2 (Bruker Daltonics). For the ion trap system, the chip contained a Zorbax 300SB-C18 column (150 mm×75 µm, 5 µm particle size) and a Zorbax 300SB-C18 enrichment column (40 nl). The solvent system consisted of 2% acetonitrile, 0.1% formic acid in water (A), and 0.1% formic acid in acetonitrile and 2% water (B). Tryptic digests from TgAlba2 and TgSsoSB co-purified proteins were loaded into the enrichment column at a flow rate set to 3.75 µl/min. Elution

was performed at a flow rate of 300 nl/min with 8–40% linear gradient (B) in 30 min. The MS instrument was operated with the following settings: source temperature was set to 135 °C while cone gas flow was at 3 L/min. The nanoelectrospray voltage was set to –1850 V. The MS spectra were acquired in the positive ion mode on the mass range 250 to 1500 *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 200,000 with a maximum accumulation time of 200 ms, and the number of averages was set to 2. For tandem MS experiments, the system was operated in the data-dependent mode with 6 MS/MS scans (in the range 100–2000 *m/z*) with a preference for doubly and triply charged ions (absolute threshold of 5000 and a relative of 5%). Ions were excluded after the acquisition of 1 MS/MS spectra and the exclusion was released after 0.12 min. The ion charge control was fixed at 300,000 and 2 scans were averaged to obtain an MS/MS spectrum.

For the Q-TOF system, tryptic digests from TgAlba1 co-purified proteins were loaded on a Symmetry C18 precolumn (20×0.18 mm, 5 µm) and the peptides were separated on an ACQUITY UPLC® BEH130 C18 column (75 µm×200 mm, 1.7 µm particle size, Waters). The solvent system consisted of 0.1% formic acid in water (A) and 0.1% formic acid in acetonitrile (B). Trapping was performed for 3 min at 15 µl/min with 99% of A and 1% of B. Elution was performed at a flow rate of 450 nl/min, using a 6–35% gradient of B for 21 min at 60 °C followed by a fast rise at 90% of B. The MS instrument was operating in positive mode, with source temperature at 200 °C while dry gas flow was at 4 L/min. The nanoelectrospray voltage was optimized to –4500 V. For tandem MS experiments, the system was operated with automatic switching between MS and MS/MS modes in the range 50–2200 *m/z* (MS acquisition time of 0.68 s, MS/MS acquisition time between 0.68 and 4.76 s). The four most abundant peptides, preferably doubly, triply, and quadruply charged ions, were selected on each MS spectrum for further fragmentation. Ions were excluded after the acquisition of 1 MS/MS spectra and the exclusion was released after 0.6 min.

Protein identification

MS/MS data sets were processed as previously described¹⁵ with minor adjustments. Searches were performed against an updated NCBI nr database (July 28, 2010) concatenated with reversed copies of all sequences (total 51,508 entries). Searches were performed for ion trap data with a mass tolerance of 0.25 Da in both MS and MS/MS mode and for Q-TOF data with a mass tolerance of 5 ppm on parent and 0.02 Da on fragment. Tryptic identified peptides and corresponding *T. gondii* proteins are provided as Supplementary Tables S2, S3, and S4.

Experimental infection in mice

All animal experiments were performed following the guidelines of the Pasteur Institute of Lille study board, which conforms to the Amsterdam Protocol on animal protection and welfare, a Directive 86/609/EEC for the Protection of Animals Used for Experimental and Other

Scientific Purposes, updated in the Council of Europe's Appendix A[†]. The animal work also complied with the French law (no. 87-848 dated 19-10-1987) and the European Communities Amendment of Cruelty to Animals Act 1976. All animals were fed with regular diet and all procedures were in accordance with national regulations on animal experimentation and welfare authorized by the French Ministry of Agriculture and Veterinary committee (permit number: 59-009145). The Pasteur Institute of Lille and the Centre National de la Recherche Scientifique Committee on the Ethics of Animal Experiments specifically approved this study. Purified tachyzoites from the parental, ΔTgalba2, ΔTgalba1, and ΔTgalba1 complemented strains were inoculated into a group of four female 6- to 8-week-old BALB/c mice at the indicated number per mouse and monitored until death or survival for 1 month, as previously described.^{15,41}

Acknowledgements

The authors would like to thank T. Mouveaux and J. Duflot for excellent technical assistance and other members of the laboratory for fruitful discussions. We also thank Dr. Steven G. Ball for carefully reading our manuscript. This work was supported by Centre National de la Recherche Scientifique, Institut National de la Santé et de la Recherche Médicale, and a grant from the French National Research Agency (number ANR-09-MIEN-002-01; BradyToxoFT).

Supplementary Data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.jmb.2013.01.039>

Received 12 September 2012;

Received in revised form 29 December 2012;

Accepted 26 January 2013

Keywords:

Toxoplasma;
gene expression;
translation;
ALBA

† <http://frodo.wi.mit.edu/primer3/>

‡ <http://conventions.coe.int/Treaty/EN/Treaties/PDF/123-Arev.pdf>

Abbreviations used:

IFA, immunofluorescence assay; HFF, human foreskin fibroblast; MS, mass spectrometry; qRT-PCR, quantitative real-time PCR; TAP, tandem affinity purification; UTR, untranslated region; miRNA, microRNA; FCS, fetal calf serum; PBS, phosphate-buffered saline; DAPI, 4',6-diamidino-2-phenylindole.

References

- Weiss, L. M. & Kim, K. (2000). The development and biology of bradyzoites of *Toxoplasma gondii*. *Front. Biosci.* **5**, D391–D405.
- Behnke, M. S., Wootton, J. C., Lehmann, M. M., Radke, J. B., Lucas, O., Nawas, J. *et al.* (2010). Coordinated progression through two subtranscriptomes underlies the tachyzoite cycle of *Toxoplasma gondii*. *PLoS One*, **5**, e12354.
- Behnke, M. S., Radke, J. B., Smith, A. T., Sullivan, W. J. & White, M. W. (2008). The transcription of bradyzoite genes in *Toxoplasma gondii* is controlled by autonomous promoter elements. *Mol. Microbiol.* **68**, 1502–1518.
- Sullivan, W. J. & Hakimi, M. A. (2006). Histone mediated gene activation in *Toxoplasma gondii*. *Mol. Biochem. Parasitol.* **148**, 109–116.
- Gissot, M., Kelly, K. A., Ajioka, J. W., Grealley, J. M. & Kim, K. (2007). Epigenomic modifications predict active promoters and gene structure in *Toxoplasma gondii*. *PLoS Pathog.* **3**, e77.
- Saksouk, N., Bhatti, M. M., Kieffer, S., Smith, A. T., Musset, K., Garin, J. *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.
- Balaji, S., Babu, M. M., Iyer, L. M. & 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.
- Gissot, M., Kim, K., Schaap, D. & Ajioka, J. W. (2009). New eukaryotic systematics: a phylogenetic perspective of developmental gene expression in the Apicomplexa. *Int. J. Parasitol.* **39**, 145–151.
- Alano, P. (2007). *Plasmodium falciparum* gametocytes: still many secrets of a hidden life. *Mol. Microbiol.* **66**, 291–302.
- Mair, G. R., Braks, J. A., Garver, L. S., Wiegant, J. C., Hall, N., Dirks, R. W. *et al.* (2006). Regulation of sexual development of *Plasmodium* by translational repression. *Science*, **313**, 667–669.
- Xia, D., Sanderson, S. J., Jones, A. R., Prieto, J. H., Yates, J. R., Bromley, E. *et al.* (2008). The proteome of *Toxoplasma gondii*: integration with the genome provides novel insights into gene expression and annotation. *Genome Biol.* **9**, R116.
- Van, T. T., Kim, S.-K., Camps, M., Boothroyd, J. C. & Knoll, L. J. (2007). The BSR4 protein is up-regulated in *Toxoplasma gondii* bradyzoites, however the dominant surface antigen recognised by the P36 monoclonal antibody is SRS9. *Int. J. Parasitol.* **37**, 877–885.
- Joyce, B. R., Queener, S. F., Wek, R. C. & Sullivan, W. J., Jr. (2010). Phosphorylation of eukaryotic initiation factor-2(α) promotes the extracellular survival of obligate intracellular parasite *Toxoplasma gondii*. *Proc. Natl Acad. Sci. USA*, **107**, 17200–17205.
- Braun, L., Cannella, D., Ortet, P., Barakat, M., Sautel, C. F., Kieffer, S. *et al.* (2010). A complex small RNA repertoire is generated by a plant/fungal-like machinery and effected by a metazoan-like Argonaute in the single-cell human parasite *Toxoplasma gondii*. *PLoS Pathog.* **6**, e1000920.

15. Olguin-Lamas, A., Madec, E., Hovasse, A., Werkmeister, E., Callebaut, I., Slomianny, C. *et al.* (2011). A novel *Toxoplasma gondii* nuclear factor TgNF3 is a dynamic chromatin-associated component, modulator of nucleolar architecture and parasite virulence. *PLoS Pathog.* **7**, e1001328.
16. Aravind, L., Iyer, L. M. & Anantharaman, V. (2003). The two faces of Alba: the evolutionary connection between proteins participating in chromatin structure and RNA metabolism. *Genome Biol.* **4**, R64.
17. Bell, S. D., Botting, C. H., Wardleworth, B. N., Jackson, S. P. & White, M. F. (2002). The interaction of Alba, a conserved archaeal chromatin protein, with Sir2 and its regulation by acetylation. *Science*, **296**, 148–151.
18. Jelinska, C., Conroy, M. J., Craven, C. J., Hounslow, A. M., Bullough, P. A., Waltho, J. P. *et al.* (2005). Obligate heterodimerization of the archaeal Alba2 protein with Alba1 provides a mechanism for control of DNA packaging. *Structure*, **13**, 963–971.
19. Guo, R., Xue, H. & Huang, L. (2003). Ssh10b, a conserved thermophilic archaeal protein, binds RNA *in vivo*. *Mol. Microbiol.* **50**, 1605–1615.
20. Mair, G. R., Lasonder, E., Garver, L. S., Franke-Fayard, B. M., Carret, C. K., Wiegant, J. C. *et al.* (2010). Universal features of post-transcriptional gene regulation are critical for *Plasmodium* zygote development. *PLoS Pathog.* **6**, e1000767.
21. Mani, J., Güttinger, A., Schimanski, B., Heller, M., Acosta-Serrano, A., Pescher, P. *et al.* (2011). Alba-domain proteins of *Trypanosoma brucei* are cytoplasmic RNA-binding proteins that interact with the translation machinery. *PLoS One*, **6**, e22463.
22. Subota, I., Rotureau, B., Blisnick, T., Ngwabyt, S., Durand-Dubief, M., Engstler, M. & Bastin, P. (2011). ALBA proteins are stage regulated during trypanosome development in the tsetse fly and participate in differentiation. *Mol. Biol. Cell*, **22**, 4205–4219.
23. Goyal, M., Alam, A., Iqbal, M. S., Dey, S., Bindu, S., Pal, C. *et al.* (2011). Identification and molecular characterization of an Alba-family protein from human malaria parasite *Plasmodium falciparum*. *Nucleic Acids Res.* **40**, 1174–1190.
24. Chêne, A., Vembar, S. S., Rivière, L., Lopez-Rubio, J. J., Claes, A., Siegel, T. N. *et al.* (2011). PfAlbas constitute a new eukaryotic DNA/RNA-binding protein family in malaria parasites. *Nucleic Acids Res.* **40**, 3066–3077.
25. Yang, S. & Parmley, S. F. (1997). *Toxoplasma gondii* expresses two distinct lactate dehydrogenase homologous genes during its life cycle in intermediate hosts. *Gene*, **184**, 1–12.
26. Ferguson, D. J. P., Parmley, S. F. & 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.
27. Lirussi, D. & Matrajt, M. (2011). RNA granules present only in extracellular *Toxoplasma gondii* increase parasite viability. *Int. J. Biol. Sci.* **7**, 960–967.
28. Fox, B. A., Falla, A., Rommereim, L. M., Tomita, T., Gigley, J. P., Mercier, C. *et al.* (2011). Type II *Toxoplasma gondii* KU80 knockout strains enable functional analysis of genes required for cyst development and latent infection. *Eukaryot. Cell*, **10**, 1193–1206.
29. Haseltine, C. A. & Kowalczykowski, S. C. (2002). A distinctive single-strand DNA-binding protein from the Archaeon *Sulfolobus solfataricus*. *Mol. Microbiol.* **43**, 1505–1515.
30. Hands-Taylor, K. L., Martino, L., Tata, R., Babon, J. J., Bui, T. T., Drake, A. F. *et al.* (2010). Heterodimerization of the human RNase P/MRP subunits Rpp20 and Rpp25 is a prerequisite for interaction with the P3 arm of RNase MRP RNA. *Nucleic Acids Res.* **38**, 4052–4066.
31. Perederina, A., Esakova, O., Koc, H., Schmitt, M. E. & Krasilnikov, A. S. (2007). Specific binding of a Pop6/Pop7 heterodimer to the P3 stem of the yeast RNase MRP and RNase P RNAs. *RNA*, **13**, 1648–1655.
32. Marsh, V. L., McGeoch, A. T. & Bell, S. D. (2006). Influence of chromatin and single strand binding proteins on the activity of an archaeal MCM. *J. Mol. Biol.* **357**, 1345–1350.
33. Sha, Z., Brill, L. M., Cabrera, R., Kleifeld, O., Scheliga, J. S., Glickman, M. H. *et al.* (2009). The eIF3 interactome reveals the translatome, a supercomplex linking protein synthesis and degradation machineries. *Mol. Cell*, **36**, 141–152.
34. Huynh, M.-H. & Carruthers, V. B. (2009). Tagging of endogenous genes in a *Toxoplasma gondii* strain lacking Ku80. *Eukaryot. Cell*, **8**, 530–539.
35. Gingras, A.-C., Caballero, M., Zarske, M., Sanchez, A., Hazbun, T. R., Fields, S. *et al.* (2005). A novel, evolutionarily conserved protein phosphatase complex involved in cisplatin sensitivity. *Mol. Cell. Proteomics*, **4**, 1725–1740.
36. Reese, M. L., Zeiner, G. M., Saeij, J. P., Boothroyd, J. C. & Boyle, J. P. (2011). Polymorphic family of injected pseudokinases is paramount in *Toxoplasma* virulence. *Proc. Natl Acad. Sci. USA*, **108**, 9625–9630.
37. Nagamune, K., Beatty, W. L. & Sibley, L. D. (2007). Artemisinin induces calcium-dependent protein secretion in the protozoan parasite *Toxoplasma gondii*. *Eukaryot. Cell*, **6**, 2147–2156.
38. Thompson, J. (2002). In situ detection of RNA in blood- and mosquito-stage malaria parasites. *Methods Mol. Med.* **72**, 225–233.
39. Puig, O., Caspary, F., Rigaut, G., Rutz, B., Bouveret, E., Bragado-Nilsson, E. *et al.* (2001). The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods*, **24**, 218–229.
40. Bahl, A., Davis, P. H., Behnke, M., Dziarszinski, F., Jagalur, M., Chen, F. *et al.* (2010). A novel multi-functional oligonucleotide microarray for *Toxoplasma gondii*. *BMC Genomics*, **11**, 603.
41. Walker, R., Gissot, M., Croken, M. M., Huot, L., Hot, D., Kim, K. & Tomavo, S. (2013). The *Toxoplasma* nuclear factor TgAP2XI-4 controls bradyzoite gene expression and cyst formation. *Mol. Microbiol.* **87**, 641–655.
42. Irizarry, R. A., Hobbs, B., Collin, F., Beazer-Barclay, Y. D., Antonellis, K. J., Scherf, U. & Speed, T. P. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*, **4**, 249–264.

43. Smyth, G. K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.*; 3, Art. 3.
44. Yang, Y. H., Dudoit, S., Luu, P., Lin, D. M., Peng, V., Ngai, J. & Speed, T. P. (2002). Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res.* **30**, e15.
45. Fauquenoy, S., Morelle, W., Hovasse, A., Bednarczyk, A., Slomianny, C., Schaeffer, C. *et al.* (2008). Proteomics and glycomics analyses of N-glycosylated structures involved in *Toxoplasma gondii*-host cell interactions. *Mol. Cell. Proteomics*, **7**, 891–910.

IV. Etude protéomique des partenaires de la sortiline (TgSORTLR)

Ce travail a fait l'objet d'un article publié dans *Cell Host & Microbe*⁵⁶ disponible dans l'annexe 3.

1. Contexte et objectif de cette étude

Lors des études protéomiques réalisées sur des extraits de glycoprotéines de *T. gondii* purifiées par chromatographie d'affinité avec la concanavine A, une protéine homologue à la sortiline de mammifère fut identifiée². L'analyse bioinformatique de la séquence d'acides aminés (962 acides aminés) de cette "*T. gondii* Sortilin-Like Receptor" (TgSORTLR) montre que celle-ci ne possède aucun site consensus de N-glycosylation. Cette protéine TgSORTLR a probablement été purifiée par le biais des glycoprotéines des rhoptries (ROP) et micronèmes (MIC). En effet ces glycoprotéines sont les protéines les plus abondantes et ont une grande affinité avec la concanavine A due à la présence de glycosylations^{2, 34, 57}. Cette hypothèse est renforcée par la forte homologie de séquence de TgSORTLR avec les orthologues de mammifères qui sont bien connus pour posséder une forte capacité de recrutement d'autres protéines^{42, 58}.

Cette "sortilin-like" située au niveau de la membrane de vésicules transgolgiennes, présente des caractéristiques structurales frappantes avec la sortiline humaine. TgSORTLR présente un domaine luminal (acides aminés compris entre les positions 37 et 789) en N-terminal, et un domaine cytoplasmique (acides aminés compris entre les positions 808 et 962) du côté C-terminal, et entre les deux domaines une zone transmembranaire (acide aminés compris entre les positions 790 et 807).

Afin d'avancer dans la compréhension du rôle de TgSORTLR chez *T. gondii*, nous avons mené une étude protéomique pour identifier les protéines impliquées dans les différents complexes protéiques des 2 domaines fonctionnels de TgSORTLR.

2. Stratégie d'analyse protéique des complexes protéiques de TgSORTLR

a. Description des échantillons

Dans un premier temps, les protéines des parasites (2×10^9 parasites) ont été réticulés *in vivo* à l'aide d'un agent pontant DSP (Dithiobis(succinimidyl protonate)) perméable à la membrane cytoplasmique du parasite *T. gondii*. Le DSP contient une amine, l'ester N-hydroxysuccinimide (NHS) à chaque extrémité d'un bras espaceur de 6 carbones. L'ester NHS réagit avec des amines primaires telles que l'amine latérale de la lysine (K) à pH 7-9 pour former des liaisons amides stables, et libère les 2 groupes N-hydroxysuccinimide. Cette réaction de DSP avec les lysines de différentes protéines voisines dans le cytoplasme entraîne leur pontage et donc leur co-purification ultérieure (Figure 6).

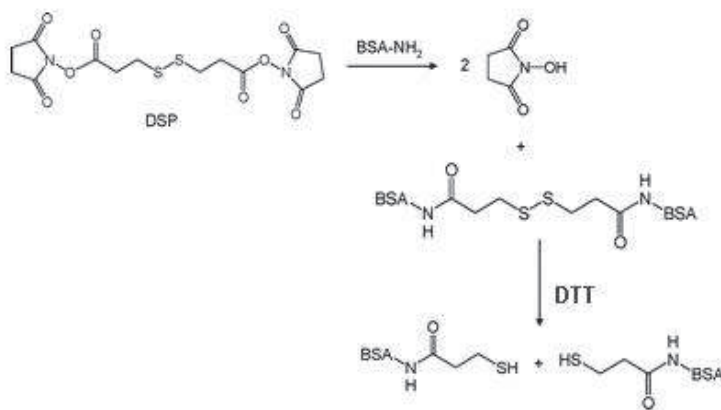


Figure 6: Exemple de pontage DSP sur la l'Albumine Bovine (BSA). Les amines primaires des lysines (K) de la BSA, réagissent avec la DSP. Les complexes protéiques de BSA pontées, sont ensuite séparés par clivage des ponts disulfures par réduction avec le DTT.

Les parasites ont été ensuite lysés pour extraire les protéines. Les complexes protéiques contenant TgSORTLR ont été purifiés par immunoprécipitation sur des billes anti-TgSORTLR. Les complexes protéiques ont ensuite été élués et réduits pour rompre les ponts DSP entre protéines avant d'être séparés sur gel SDS-PAGE.

Les protéines seront par la suite soumises à une étape de réduction suivie d'une alkylation. Les résidus de cystéines et les lysines modifiées par la DSP sont donc alkylés. Les peptides qui contiennent des lysines dans leur séquence vont donc avoir, un incrément de masse de +145 Da dû à la lysine modifiée par le DSP.

Dans un deuxième temps des mutants des différents domaines de TgSORTLR ont été analysés par nanoLC-MS/MS pour vérifier que les domaines de séquences protéiques attendus étaient correctement exprimés dans chaque clone. Puis, des billes greffées avec le coté C-terminal de la sortiline, TgSORTLR₍₈₀₈₋₉₆₂₎-GST et de billes de contrôle greffées avec GST ont été incubés en parallèle avec les extraits protéiques totaux de parasites (2×10^9 parasites pour chaque sorte de bille) pour piéger spécifiquement les protéines liées à TgSORTLR₍₈₀₈₋₉₆₂₎-GST ou GST qui sert de contrôle. Les billes TgSORTLR₍₈₀₈₋₉₆₂₎-GST portent les extrémités C-terminales (séquence protéique comprise entre les acides aminés 808 et 962 de TgSORTLR) de TgSORTLR correspondant à sa partie cytoplasmique fusionnées avec la GST (Glutathione S-transferase). La GST a été fusionnée à TgSORTLR₍₈₀₈₋₉₆₂₎ par le coté N-terminal pour que l'interaction protéine C-terminal de TgSORTLR soit conservée. Les protéines éluées de chacune des billes après 24 heures d'incubation ont été séparées ensuite sur gel SDS-PAGE. Les bandes obtenues sur chacun des gels dans les deux expériences ont été découpées de façon systématique pour l'analyse nanoLC-MS/MS.

b. Analyse nanoLC-MS/MS

La stratégie adoptée pour l'identification des partenaires des protéines TgAlba est décrite sur la Figure 7.

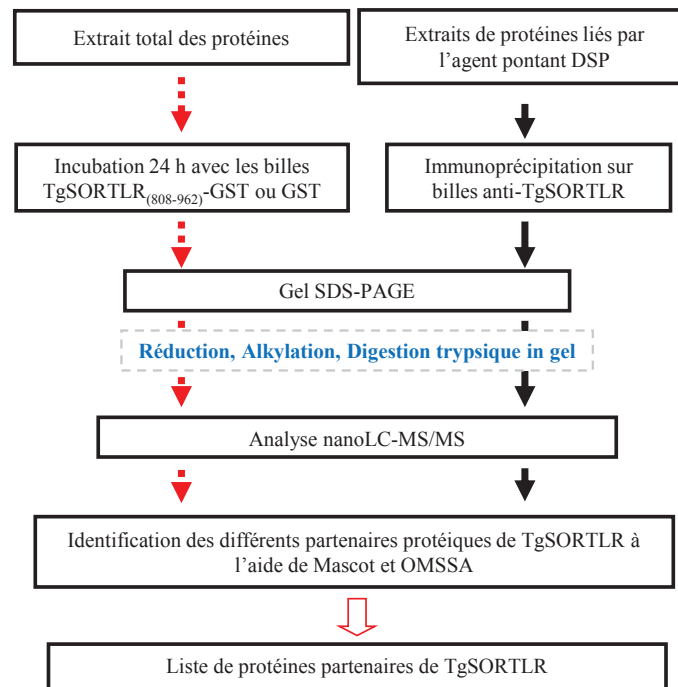


Figure 7. Stratégie globale d'analyse protéomique pour l'identification des partenaires protéiques des différents domaines de TgSORTLR.

Les conditions des analyses nanoLC-MS/MS sont les mêmes que celles utilisées pour l'identification des partenaires des protéines TgAlba (Annexe 2).

Les identifications des partenaires des protéines TgSORTLR ont été faites avec deux algorithmes (Mascot et OMSSA) dans une banque extraite de ToxoDB (www.toxodb.org).

3. Résultats publiés

Grâce à la stratégie mise en place pour l'analyse du complexe protéines-TgSORTLR formé par réticulation *in vivo* à l'aide de l'agent pontant DSP, nous avons identifié en plus de TgSORTLR des protéines micronèmes MIC1, MIC4 et des protéines des rhoptries ROP1, ROP2, ROP4 et ROP5 ainsi que d'autres protéines majoritaires du cytoplasme comme l'actine par exemple. TgSORLR a été identifiée avec plusieurs peptides dont 2 modifiés par l'agent pontant DSP comme le montre le spectre de fragmentation sur la Figure 8.

Un peptide de la protéine ROP5 été également identifié avec la modification induite par l'agent pontant DSP (Figure 9).

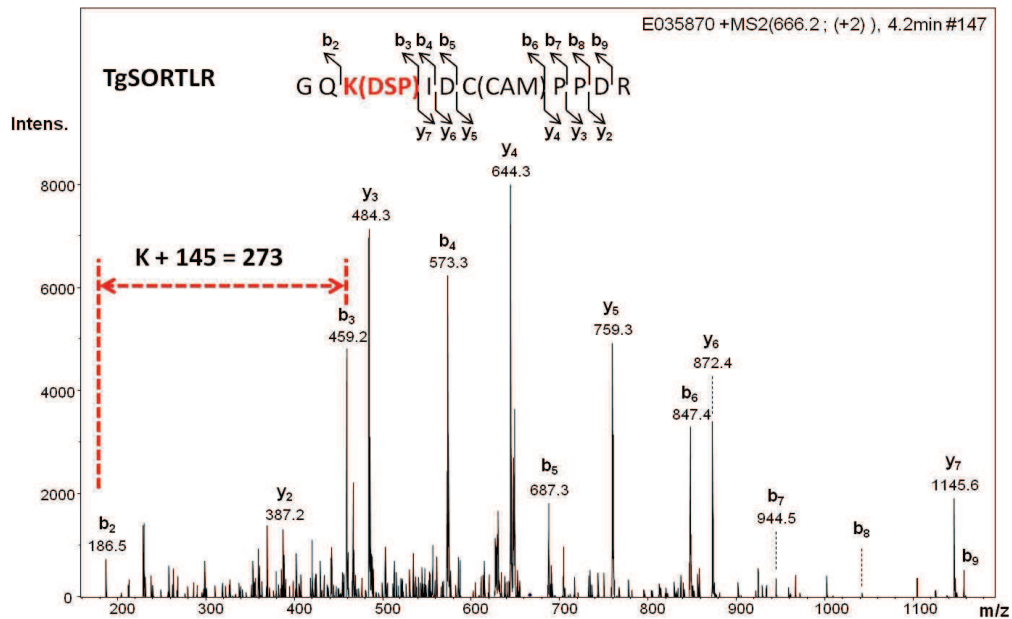


Figure 8: Spectre MS/MS annoté d'un peptide tryptique de TgSORTLR di-chargé de $m/z = 666,2$ portant une lysine (K) modifiée par l'agent pontant DSP. La liaison covalente avec l'agent pontant entraîne une différence de masse de +145 Da sur la lysine (K(DSP)). L'écart de masse entre les fragments b_2 et b_3 correspond à $K + 145$ Da.

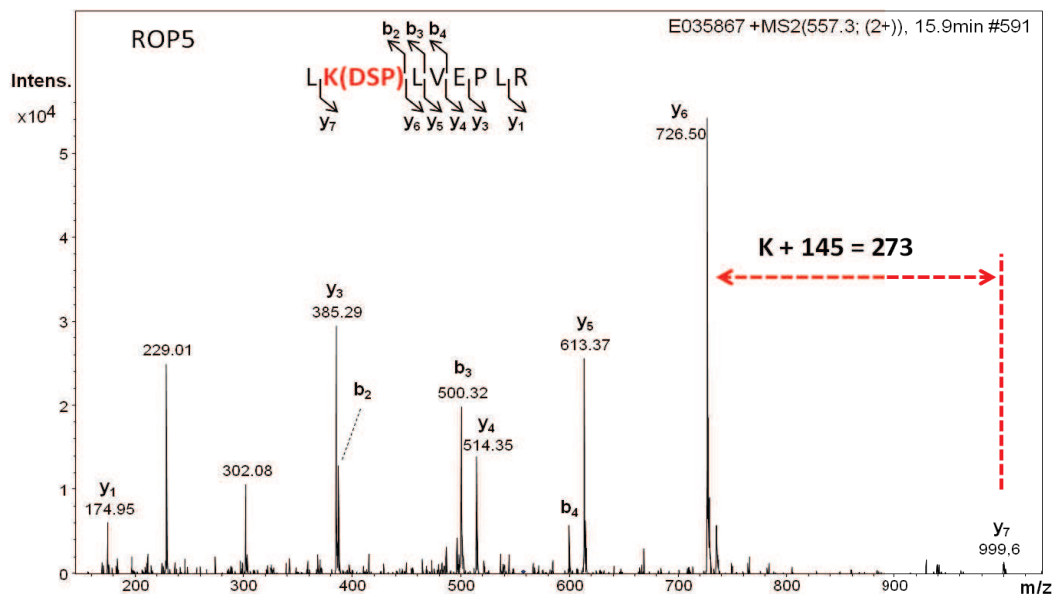


Figure 9: Spectre MS/MS annoté du peptide tryptique de ROP 5 di-chargé de $m/z = 557,3$ portant une lysine (K) modifiée par l'agent pontant DSP. La liaison covalente avec l'agent pontant entraîne une différence de masse de +145 Da sur la lysine (K(DSP)). L'écart de masse entre les fragments y_6 et y_7 correspond à $K + 145$ Da.

Ces résultats ont donc montré que les micronèmes (MIC1, MIC4) et les rhotries (ROP1, ROP2, ROP4, ROP5) sont liées à TgSORTLR, et donc interagissent avec cette protéine cible dans le milieu cytoplasmique *in vivo*.

L'analyse des mutants des différents domaines de TgSORTLR a permis de vérifier que les domaines de séquences protéiques attendus étaient correctement exprimés dans chaque parasite muté.

L'analyse du complexe protéique immunopurifié avec TgSORTLR₍₈₀₈₋₉₆₂₎, l'extrémité C-terminale cytoplasmique de TgSORTLR, a permis de mettre en évidence les interactions de protéines de protéines parasitaires homologues aux chaînes lourdes de Clathrin et les petites chaînes de protéines adaptatrices associées à la Clathrin (adaptins β , γ , $\mu 1$). Nous avons également identifié des homologues de protéines

(Sec23/Sec24) impliqués dans le transport antérograde des protéines du réticulum endoplasmique vers l'appareil de Golgi. Plusieurs partenaires protéiques cytoplasmiques du système endosomal homologues des rétromères associés au tri vacuolaire des protéines Vps9, Vps26, et Vps35 ont été mises en évidence par cette étude. Ce résultat est en accord avec les travaux de Kim et al., 2010 qui montrent que la sortiline se lie au complexe rétromère impliqué dans le transport rétrograde des endosomes médiateurs au réseau trans-Golgien chez des souris⁵⁸.

Conclusion

Grâce à la stratégie mise en place, nous avons montré que TgSORTLR se lie à des protéines impliquées dans la virulence de *T. gondii* telles que les protéines des micronèmes et des rhoptries. Nous avons aussi mis en évidence que la sortiline de *Toxoplasma gondii* est un récepteur golgien et post-golgien clé, dont le domaine peptidique C-terminal se lie à plusieurs partenaires protéiques cytoplasmiques du système endosomal (clathrin, adaptin, les protéines Vps).

Des expériences biochimiques complémentaires faites par nos collaborateurs sur différents domaines de TgSORTLR ont permis de montrer que les protéines micronèmes et les rhoptries sont spécifiquement liées au N-terminal de TgSORTLR. Par ailleurs ces expériences ont aussi montré que l'extrémité cytoplasmique C-terminale de TgSORTLR se fixe spécifiquement aux protéines endosomales, et joue un rôle capital dans le trafic antérograde et rétrograde du récepteur, dans sa localisation subcellulaire et ses fonctions biologiques.

L'ensemble des résultats de biologie cellulaire, de génétique et de protéomique a démontré, pour la première fois que ce récepteur membranaire clé est impliqué dans le tri des protéines, la biogenèse d'organites apicaux et l'infection de l'hôte.

Afin de valider les interactions spécifiques de ces protéines partenaires du domaine C-terminal, 6 protéines (Vps9, 26, 35, Clathrin μ 1, Sec23 et une protéine "hypothetical" Hp 20) ont été sélectionnés pour des expériences d'immuno-purification inverse. Il s'agira ensuite de caractériser les complexes protéiques par analyse protéomique et de déterminer le rôle spécifique de chacune de ces protéines, et plus particulièrement d'appréhender le rôle exact de certaines protéines dont la fonction biologique est encore inconnue et de construire un réseau d'interactions autour de ces 6 protéines sélectionnées et de TgSORTLR. Près de 600 protéines impliquées dans des voies métaboliques diverses et variées ont été identifiées. Une analyse bioinformatiques des résultats d'analyse protéomique est en cours afin de pouvoir construire ce réseau d'interactions protéine-protéine autour de TgSORTLR.

Enfin, pour confirmer les résultats d'analyse protéomique, nos collaborateurs biologistes, par des techniques de "knock in" ou remplacement du gène endogène de la protéine d'intérêt par une copie étiquetée avec l'épitope HA, ont pu montrer que des protéines homologues de μ -adaptine, Sec23, Vps29, Vps26 et Vps35 de *T. gondii* co-localisent et se fixent spécifiquement à TgSORTLR. Ces résultats confirment que l'extrémité C-terminal et cytoplasmique du récepteur TgSORTLR se comporte d'une manière similaire aux sortilines des cellules de mammifères.

Annexe 3: Publication

***Toxoplasma* Sortilin-like Receptor Regulates Protein Transport and Is Essential for Apical Secretory Organelle Biogenesis and Host Infection**

Sloves, P. J. ; Delhayé, S. ; Mouveaux, T. ; Werkmeister, E. ; Slomianny, C.; Hovasse, A.
Dilezitoko Alayi, T. ; Callebaut, I.; Gaji, R. Y. ; Schaeffer-Reiss, C. ; Van Dorsselear, A.
Carruthers, V. B. ; Tomavo, S.

Cell Host & Microbe, 2012, 11:5

Toxoplasma Sortilin-like Receptor Regulates Protein Transport and Is Essential for Apical Secretory Organelle Biogenesis and Host Infection

Pierre-Julien Sloves,¹ Stephane Delhaye,¹ Thomas Mouveaux,¹ Elisabeth Werkmeister,¹ Christian Slomianny,² Agnes Hovasse,³ Tchilabalo Dilezitoko Alayi,³ Isabelle Callebaut,⁴ Rajshekhar Y. Gaji,⁵ Christine Schaeffer-Reiss,³ Alain Van Dorsselaar,³ Vern B. Carruthers,⁵ and Stanislas Tomavo^{1,*}

¹Center for Infection and Immunity of Lille, CNRS UMR 8204, INSERM U 1019, Institut Pasteur de Lille, Université Lille Nord de France, 59000 Lille, France

²Laboratory of Cell Physiology, INSERM U 1003, Université Lille Nord de France, Villeneuve d'Ascq, France

³Laboratory of Mass Spectrometry, IPHC, CNRS UMR 7178, Université de Strasbourg, 67087 Strasbourg, France

⁴Institut de Minéralogie et de Physique des Milieux Condensés, CNRS UMR 7590, Université Pierre et Marie Curie, 75005 Paris, France

⁵Department of Microbiology and Immunology, University of Michigan Medical School, Ann Arbor, MI 48109, USA

*Correspondence: stan.tomavo@pasteur-lille.fr

DOI 10.1016/j.chom.2012.03.006

SUMMARY

Apicomplexan parasites have an assortment of unique apical secretory organelles (rhoptries and micronemes), which have crucial functions in host infection. Here, we show that a *Toxoplasma gondii* sortilin-like receptor (TgSORTLR) is required for the subcellular localization and formation of apical secretory organelles. TgSORTLR is a transmembrane protein that resides within Golgi-endosomal related compartments. The luminal domain specifically interacts with rhoptry and microneme proteins, while the cytoplasmic tail of TgSORTLR recruits cytosolic sorting machinery involved in anterograde and retrograde protein transport. Ectopic expression of the N-terminal TgSORTLR luminal domain results in dominant negative effects with the mislocalization of both endogenous TgSORTLR as well as rhoptry and microneme proteins. Conditional ablation of TgSORTLR disrupts rhoptry and microneme biogenesis, inhibits parasite motility, and blocks both invasion into and egress from host cells. Thus, the sortilin-like receptor is essential for protein trafficking and the biogenesis of key secretory organelles in *Toxoplasma*.

INTRODUCTION

The phylum Apicomplexa contains some of the most deadly pathogens of medical and veterinary importance, including *Plasmodium* (malaria), *Eimeria* (coccidiosis), and *Toxoplasma gondii*, which pervasively infects humans. As an obligate intracellular pathogen, *T. gondii* is capable of actively invading a wide range of nucleated host cells, wherein it replicates in a parasitophorous vacuole (PV). In addition to the canonical eukaryotic organelles, such as a central nucleus surrounded by an

extensive endoplasmic reticulum (ER) and a single Golgi situated in the proximal anterior region, *T. gondii* has three other morphologically distinct, specialized secretory organelles called micronemes, rhoptries, and dense granules that typify apicomplexan parasites (Pelletier et al., 2002; Boothroyd and Dubremetz, 2008). Whereas dense granule proteins are secreted constitutively, rhoptry (ROP) and microneme (MIC) contents are sequentially released upon binding to the host-cell surface and during the subsequent invasion phase in a highly regulated manner. Microneme proteins are not only crucial for binding to the host cell prior to entry, but are also essential for parasite gliding motility that drives active invasion (Huynh and Carruthers, 2006). Rhoptry secretion establishes the PV and delivers key virulence factors that modulate immune responses and other vital cellular functions (Boothroyd and Dubremetz, 2008).

T. gondii ROP and MIC proteins navigate through the ER, Golgi, and endosome-like organelles prior to being packaged into their respective apical secretory organelles (Ngô et al., 2003; Harper et al., 2006). A few components of the *T. gondii* vesicle budding, transport, and fusion machinery have been partially characterized (Chaturvedi et al., 1999; Hager et al., 1999; Stedman et al., 2003), and tyrosine-based endosomal sorting signals and associated adaptor complexes have been described (Hoppe et al., 2000). Also, a large dynamin-like GTPase, DrpB, expressed in the cytoplasm near the Golgi is required for the biogenesis of secretory organelles of *T. gondii* (Breinich et al., 2009). While these studies suggest that *T. gondii* possesses machinery involved in vesicular trafficking through the Golgi, endosomal system, and the secretory system, precisely how proteins are guided through these endomembranous structures remains poorly understood. Similar to virtually all other unicellular eukaryotes, *T. gondii* and other apicomplexans lack a mannose-6-phosphate receptor for sorting to the endosomal pathway. Although *T. gondii* possesses two putative plant-like vacuolar sorting receptors (TgVSR1 and TgVSR2), neither appear to be necessary for the protein trafficking to apical organelles based on expression and knockout experiments (R.Y.G. and VBC, unpublished data). Sortilin, also known as VPS10 in yeast, is a type I single pass transmembrane cargo

receptor that functions in mannose-6-phosphate independent sorting to the endolysosomal system. Expressed in a number of vertebrate tissues, notably brain, spinal cord, testis, and skeletal muscle, sortilin also functions as a surface coreceptor for induction of neural apoptosis in the brain (Kim and Hempstead, 2009) and is linked to type 2 diabetes (Clee et al., 2006) and Alzheimer's disease (Rogaeva et al., 2007). However, mice or yeast deficient in sortilin/VPS10 are viable and show relatively mild phenotypes (Marcusson et al., 1994; Jansen et al., 2007). Here, we demonstrate that a *Toxoplasma gondii* sortilin-like receptor, TgSORTLR is a major cargo receptor that is essential for protein trafficking, the biogenesis of apical secretory organelles, and the development of host infection and pathogenicity. Our findings illustrate a key role for an endolysosomal cargo receptor in protein shepherding through the exocytic system and also point toward the indication that, unlike yeast and higher eukaryotes, lower branching eukaryotes, including pathogenic species, lack redundant mechanisms for protein trafficking.

RESULTS

T. gondii Sortilin-like Receptor Resides in the Golgi and Endosomal-Related Compartments

Using concanavalin-A lectin affinity purification and mass spectrometry, we previously identified a *T. gondii* ortholog of mammalian sortilin (TgSORTLR), which copurified with several ROP and MIC proteins involved in host cell invasion (Fauquenoy et al., 2008). Cloning and sequencing the full-length cDNA revealed a 962 amino acid protein with predicted molecular mass of 105 kDa. Interestingly, TgSORTLR itself is not *N*-glycosylated (Fauquenoy et al., 2008, 2011), suggesting that it bound to the lectin column via interactions with ROP and MIC glycoproteins. This notion is also supported by the high degree of similarity of TgSORTLR to mammalian sortilins (Figures 1A and S1A), which are well known for their protein-binding capabilities. The luminal domain of TgSORTLR exhibits conserved ten-bladed propellers (blades A to J) and two small 10-cysteine domains (10CCa and 10CCb), which interact with the β -propeller and influence ligand binding (Figures 1A and S1A; Quistgaard et al., 2009). Modeling of the 3D structure of TgSORTLR luminal domain also showed striking specific structural features that are shared by sortilin homologs of other apicomplexans (Figures S1B and S1C). These unique features include sequence insertions organized as three large loops in the luminal cargo-binding domain (inserts 1 to 3, Figures 1A and S1A). Additionally, the C-terminal tail of TgSORTLR contains an acidic cluster and a dileucine motif (Figures 1A and S1D), which in mammalian sortilin are crucial for intracellular sorting (Hermey, 2009).

Antibodies to recombinant TgSORTLR recognized a ~105 kDa protein in *T. gondii* lysate (Figure 1B). Confocal imaging of intracellular parasites stained with anti-TgSORTLR revealed labeling in the Golgi region (Figures 1C and 1D). Indeed, analysis of confocal images showed that $73 \pm 6\%$ of TgSORTLR colocalizes with GRASP-RFP, a marker of *T. gondii* cis-medial Golgi (Figures 1D and 1E; Pelletier et al., 2002). Three-dimensional reconstructions from confocal images confirmed the presence of TgSORTLR in the Golgi (Movie S1). In addition, we showed that TgSORTLR also colocalized significantly with markers of the TGN (GalNAcT), early endosomes (Rab5a), late

endosomes (VP1 and Rab7), and a lysosomal-related compartment (CPL), with colocalization estimated at 70%, 65%, 62%, and 40%, respectively (Figures 1D and 1E). In contrast, TgSORTLR showed little colocalization with the ER marker HDEL-RFP (10%) or the dynamin DrpB (3%) (Figures 1D and 1E). Cryoimmuno-electron microscopy (cryo-IEM) of TgSORTLR in intracellular parasites ectopically expressing TgSORTLR-cMyc (Figure 1F, left panel) or endogenous wild-type protein (Figure 1F, middle panel) confirmed Golgi cisternae. Some of the labeled components seen by cryo-IEM (e.g., Figure 1F, middle panel) resembled the tubular post-Golgi structures noted by Harper and colleagues (Harper et al., 2006). TgSORTLR was also seen by cryo-IEM in membrane vesicles in the central anterior region (Figure 1F, right panel), suggesting that a significant proportion of TgSORTLR may localize beyond the Golgi, likely in the *trans*-Golgi Network (TGN) and/or endosome-like compartments. We therefore conclude that the Golgi and endosomal localization of TgSORTLR defines it as a plausible candidate type I transmembrane receptor for protein sorting and trafficking through the endocytic/exocytic system of *T. gondii*.

TgSORTLR Is a Cargo Receptor that Binds ROP and MIC Proteins

The localization of TgSORTLR in Golgi and endosome-like compartments prompted us to identify its protein ligands. Live parasites were crosslinked with the reduction sensitive, cell-permeable crosslinker dithiobis succinimidyl propionate (DSP) (Fauquenoy et al., 2011). A detergent extract was immunoprecipitated with anti-TgSORTLR beads, and the eluted protein complexes (Figure S2A) were identified by mass spectrometry as microneme MIC1, MIC4, and rhoptry ROP1, ROP2, ROP4 and ROP5 proteins (Table S1). Our analysis revealed seven TgSORTLR peptides (Table S1), including two that were modified by DSP (Figures S2B and S2C). One DSP-modified peptide from ROP5 was also identified (Figure S2D). Furthermore, incubation of parasite extracts with recombinant luminal domain rTgSORTLR₃₇₋₇₈₉ fused to GST confirmed binding to MIC4, ROP2, and ROP4 proteins and further revealed interactions with microneme proteins MIC5 and MA2P (Figure S2E). In contrast, rTgSORTLR₃₇₋₇₈₉ did not bind GRA3. GST alone also failed to bind any of the MIC or ROP proteins examined. Furthermore, we showed that TgSORTLR colocalizes partially with MIC5 (Figure 2B, merge), ROP1 (Figure 2C, merge), and ROP5 (Figure S2F) using confocal microscopy. Rhoptries and micronemes are made de novo in daughter parasites undergoing replication, and their resident propeptides are subjected to proteolytic processing within endosomal compartments (Carey et al., 2004; Harper et al., 2006; Parussini et al., 2010). Using antibodies specific to propeptides of ROP4 and MIC5 proteins, we also found that TgSORTLR partially colocalizes with the pro-ROP4 and proMIC5, respectively, within the proximal anterior region (Figures S2G and S2H). However, the level of colocalization between TgSORTLR and proROP4 or proMIC5 varies considerably from one vacuole to another (Figures S2G and S2H), suggesting potential transient interactions that may occur in a cell-cycle dependent manner. Collectively, these data suggest TgSORTLR is a cargo receptor that interacts physically with ROP and MIC proteins in the Golgi-endosomal compartments of *T. gondii*.

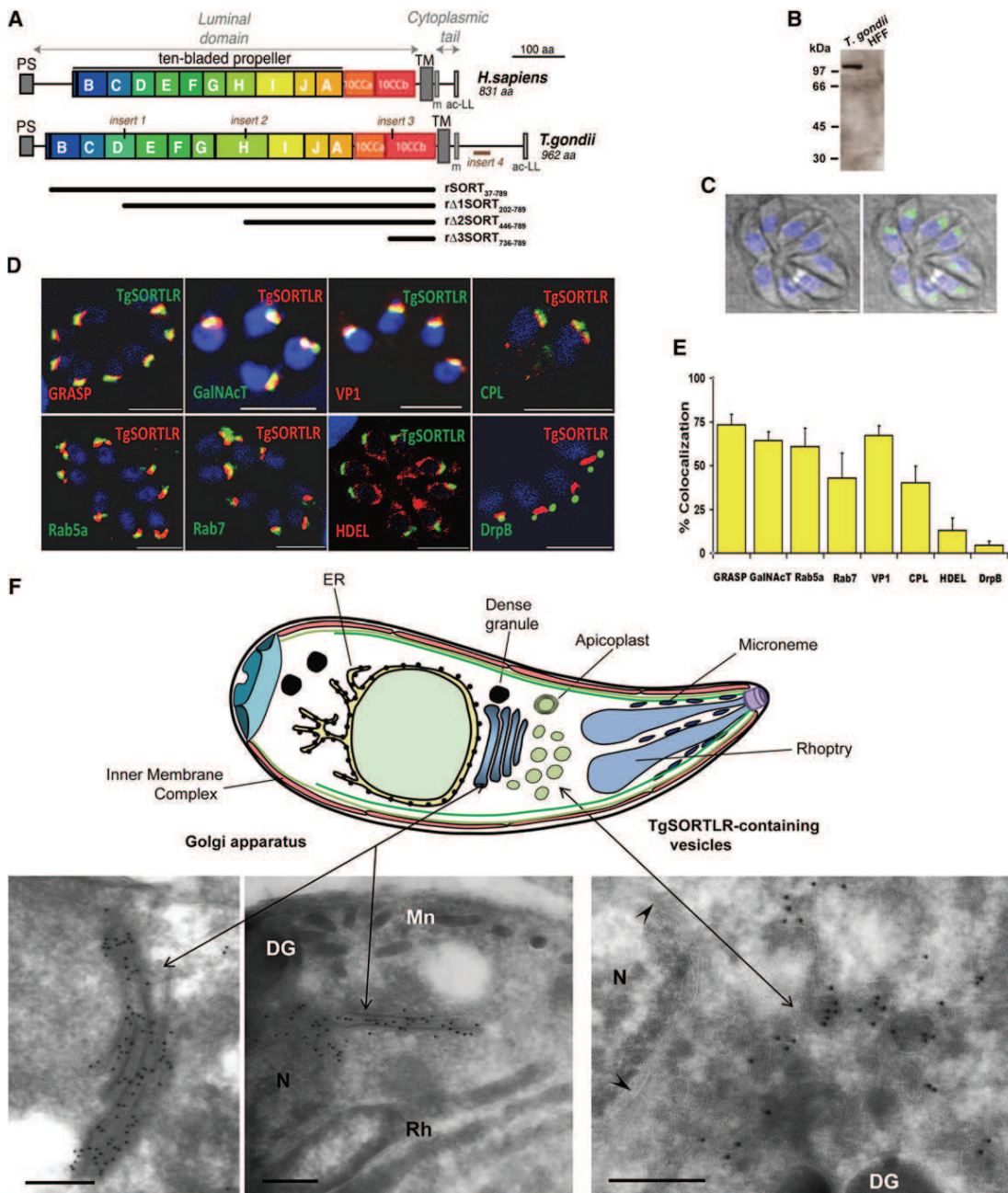


Figure 1. *T. gondii* Sortilin-like Receptor Resides in the Golgi and Endosome-Related Structures

(A) Schematic representations of key domains of human sortilin, TgSORTLR, the luminal domain of TgSORTLR, and its truncated versions fused to GST. PS, peptide signal; TM, transmembrane; m, membrane-adjacent motif; ac-LL, acidic cluster-dileucine motif. See also Figure S1.

(B) Immunoblot probed with mouse polyclonal anti-TgSORTLR antibodies reveals a single protein of ~105 kDa in *T. gondii* whereas no protein was detected in human foreskin fibroblast (HFF) cell extract.

(C) Localization of TgSORTLR by confocal images using the mouse polyclonal anti-TgSORTLR antibodies. Bars, 5 μ m.

(D) TgSORTLR colocalized with Golgi-endosomal and trafficking markers (see also Movie S1). Intracellular *T. gondii* tachyzoites stably expressing GRASP-RFP (*cis*-Golgi and medial Golgi), HDEL-RFP (endoplasmic reticulum), and DrpB-YFP (Golgi proximal cytoplasmic inclusion) or transiently expressing GalNAcT-GFP (*trans*-Golgi network), Rab5a (early endosome), and Rab7 (late endosome), or rabbit polyclonal anti-VP1 (late endosome and lysosome-related compartment) antibodies were used. Bars, 5 μ m.

(E) Quantification of TgSORTLR colocalization Golgi-endosomal markers. Bars represent mean values of $n = 3$ independent experiments \pm standard deviation.

(F) Schematic representation of *T. gondii* showing its different compartments and organelles. Cryo-TEM of TgSORTLR localized in the Golgi apparatus of transgenic TgSORTLR-cMyc parasite (left panel), wild-type parasite (middle panel), and Golgi-proximal vesicles of transgenic TgSORTLR-cMyc strain (right panel). Black arrowheads indicate the nuclear envelope. Mn, microneme; Rh, rhoptry; DG, dense granule; N, nuclei. Bars, 200 nm.

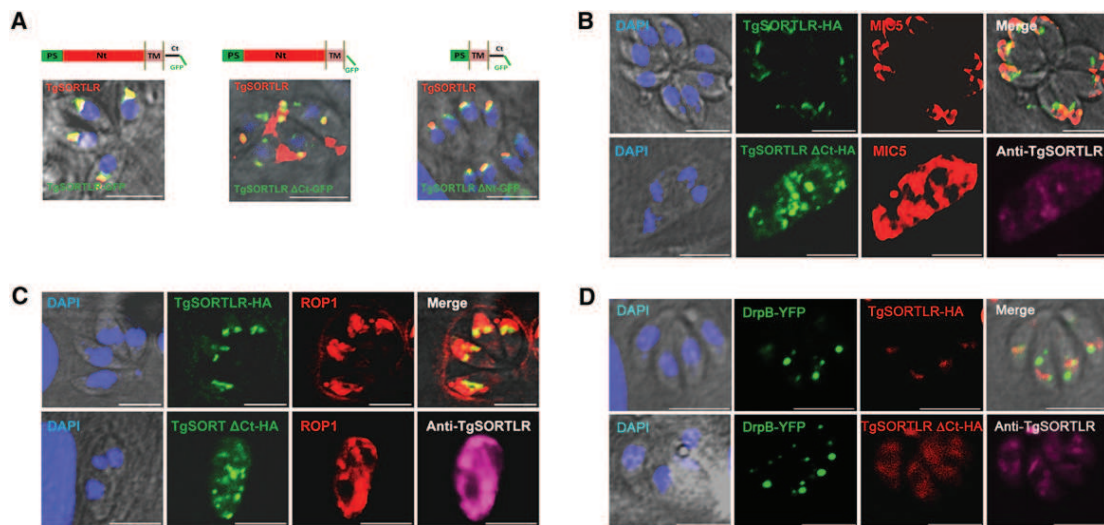


Figure 2. The Cytoplasmic Tail of TgSORTLR Is Essential for Correct Subcellular Localization of Endogenous TgSORTLR, Rhopty, and Microneme Proteins

(A) Expression of TgSORTLR protein lacking its C-terminal tail (TgSORTLR Δ Ct-GFP, green) resulted in the mislocalization of endogenous TgSORTLR stained with the anti-TgSORTLR antibodies (middle panel, red). TgSORTLR Δ Nt-GFP (right panel) and the full-length TgSORTLR-GFP (left panel) mediate the correct localization in the Golgi-endosomal structures. The endogenous TgSORTLR was labeled with the mouse anti-TgSORTLR and goat secondary antibodies.

(B) Expression of TgSORTLR Δ Ct-HA (rat polyclonal anti-HA9 and 488 nm-Alexa goat secondary antibodies, green) resulted in the mislocalization of MIC5 (rabbit anti-MIC5 and 594 nm-Alexa goat secondary antibodies, red). Endogenous TgSORTLR was stained with the mouse polyclonal anti-TgSORTLR and 633 nm-Alexa goat secondary antibodies.

(C) Expression of TgSORTLR Δ Ct-HA resulted in the mislocalization of ROP1. TgSORTLR Δ Ct-HA, endogenous TgSORTLR, and ROP1 were labeled as above.

(D) Expression of TgSORTLR Δ Ct-HA has no effect in the Golgi proximal localization of DrpB-YFP. These data indicate that rhopty and microneme proteins are present in the Golgi-endosomal associated structures in which TgSORTLR resides (see also Figure S2). Bars, 5 μ m for all panels.

Ectopic Expression of a C-Terminally Truncated Mutant of TgSORTLR Abolishes the Transport to Rhoptries and Micronemes

To substantiate that *T. gondii* uses a sortilin-like receptor to traffic invasion-related proteins to the apical organelles, wild-type parasites were transiently transfected with full-length or truncated mutants of TgSORTLR fused to either GFP or an HA epitope tag. Whereas full-length TgSORTLR-GFP showed the expected proximal anterior localization (Figure 2A, left panel), C-terminally truncated TgSORTLR fused to GFP (TgSORTLR Δ Ct-GFP) displayed multiple puncta both anterior and posterior to the parasite nucleus (Figure 2A, middle panel). These data suggest an essential role for the C-terminal tail in the proper localization of TgSORTLR. Furthermore, truncation of the luminal domain demonstrated that the C-terminal tail alone is sufficient for Golgi-endosomal localization (Figure 2A, right panel). Interestingly, transient expression of C-terminally truncated TgSORTLR-HA (Figures 2B and 2C, TgSORTLR Δ Ct-HA, green signal and lower panels) appeared to exert a dominant negative effect, causing the mislocalization of endogenous TgSORTLR (Figures 2B and 2C, anti-TgSORTLR, lower and right panels) along with ROP1 and MIC5, which accumulated throughout the cytoplasm of transfected parasites (Figures 2B and 2C, red signal and lower panels). In contrast, no effect on the Golgi proximal localization of the DrpB was noted (Figure 2D). Transient expression of TgSORTLR Δ Ct-HA in *T. gondii* tachyzoites also led to a severe defect in intracellular parasite division, which is consistent with our repeated failure to generate *Toxoplasma* lines stably expressing this mutant. These findings

suggest that the dominant negative effect of expressing TgSORTLR Δ Ct compromises parasite replication.

Conditional Ablation of TgSORTLR Disrupts Protein Trafficking and Biogenesis of Rhoptries and Micronemes

Since dominant negative effects can be problematic to interpret, we conditionally ablated TgSORTLR expression in anhydrotetracycline (ATc)-inducible TgSORTLR knockout (iKO) parasites (Figures 3A and 3B). ATc treatment resulted in the loss of TgSORTLR protein in iKO parasites but had no effect on expression in wild-type parasites (Figure 3C). Confocal imaging confirmed the absence of TgSORTLR after ATc treatment (Figure 3D). ROP2–4 and MIC1 coimmunoprecipitated with TgSORTLR in wild-type and cMyc-tagged transgenic parasites, even in the absence of DSP (Figure 3E), but failed to do so in iKO mutants grown with ATc (Figure 3F). These findings confirm the specificity of the interactions, which is also supported by the failure of TgSORTLR to coimmunoprecipitate with DrpB-YFP (Figure S2I). To decipher the molecular basis of ROP and MIC protein binding to TgSORTLR, we tested the complete luminal domain of TgSORTLR along with a series of truncation mutants (see schematic representations in Figure 1A) for binding to either native or recombinant ROP and MIC proteins. Figure 3G shows that the recombinant TgSORTLR proteins were purified to homogeneity with similar yields. The recombinant complete luminal domain of TgSORTLR (rSORT₃₇₋₇₈₉-GST) bound to MIC5 and ROP1 (Figure 3I) in total protein extracts from the iKO depleted of endogenous TgSORTLR by ATc treatment (Figure 3H,

compare lanes 5 and 6). Similarly, the first two truncation mutants, rSORT₂₀₂₋₇₈₉-GST and rSORT₄₄₄₋₇₈₉-GST, also bound to MIC5 and ROP1, with the latter mutant showing enhanced binding to ROP1 (Figure 3I). Neither the third recombinant mutant rTgSORT₇₃₆₋₇₈₉-GST nor GST protein alone bound to MIC5 and ROP1, confirming that the bladed propeller domains of TgSORTLR are necessary for cargo binding. To test for direct binding, recombinant MIC5-HA and ROP1-HA proteins expressed in wheat germ extracts (Figure 3H, lanes 2 and 4) were incubated with the series of recombinant TgSORTLR proteins. No binding was seen between MIC5-HA or ROP1-HA and any of the recombinant TgSORTLR proteins (Figure 3J). *E. coli*-derived recombinant MIC5 and MIC10 also failed to bind recombinant TgSORTLR. These findings suggest that the binding of MIC and ROP proteins to TgSORTLR is indirect or requires features of the native MIC and ROP proteins that are not present in recombinant forms of these proteins.

Next, we treated iKO parasites with ATc to ablate expression of TgSORTLR and examine the effect on ROP and MIC protein localization. TgSORTLR deficiency strikingly altered the apical distribution of all ROP and MIC markers tested. Figure 4 illustrates the mislocalization of ROP1, ROP2-4, ROP5, ROP7, and RON1 proteins (Figure 4A) and MIC2, MIC3, MIC4, and M2AP proteins (Figure 4B). We also noticed that in some vacuoles of the TgSORTLR-deficient mutant, ROP1 and MIC4 appear as residual cytoplasmic puncta whereas other vacuoles showed the complete absence of ROP1 and MIC4 (Figures 4A and 4B). In addition, ROP7 (Figure 4A), MIC3, and ROP2-4 (data not shown) were detected in the host cytoplasm and localized around the nucleus, consistent with leaking of the PV contents into the host cell cytosol. In contrast, no morphological changes were observed for dense granules (GRA3), inner membrane complex (GAP45), the remnant plastid (Api, Atrx2; DeRocher et al. [2008]), the endoplasmic reticulum (SERCA; Nagamune et al. [2007]), or the mitochondrion (Mito, HSP60; Toursel et al. [2000]) (Figure 4C). Finally, complementation of iKO mutants by transient transfection of TgSORTLR-GFP restored the classical microneme and rhoptry apical patterns, while the TgSORTLR luminal domain (TgSORTLR Δ Ct-GFP) failed to rescue the defect localizations (Figure S3), confirming a functional requirement of the C-terminal cytosolic domain.

TgSORTLR Is Essential for the Formation of Rhoptries and Micronemes in *T. gondii*

Following 72 hr of ATc treatment, we observed that TgSORTLR-deficient parasites were devoid of the typical electron-dense club-shaped rhoptry and flattened microneme organelles (Figures 5B and 5C) compared to wild-type parasites (Figure 5A). Five serial ultrathin sections spaced 100 nm apart through five adjacent TgSORTLR-deficient parasites confirmed the absence of typical and morphologically defined rhoptries in 23 out of the 25 serial electron micrographs (Figure S4). Furthermore, no typical micronemes were seen in these 25 serial ultrathin sections of TgSORTLR-deficient parasites, whereas numerous dense granules (DG) were still present (Figures 5C and S6). We also observed by cryo-IEM that ROP2-4 and MIC5 proteins were diverted to the PV (Figures 5E and 5G), while these proteins correctly resided in their respective apical organ-

elles in the absence of ATc (Figures 5D and 5F). Finally, MIC5 was detected in the dense granules after ATc treatment (Figure 5H, inset), suggesting that it and possibly the other mislocalized secretory proteins are rerouted to this default, constitutive secretory pathway in the absence of regulated secretory organelles.

TgSORTLR Cytoplasmic Tail Binds to Cargo Proteins that Mediate Both Anterograde and Retrograde Golgi-Endosome Trafficking

To further substantiate the function of TgSORTLR as cargo receptor for sorting microneme and rhoptry proteins, we fused the cytoplasmic tail of TgSORTLR to GST and incubated it with parasite extracts to test for interactions with components of the cytosolic sorting machinery. As shown in Table S2, the TgSORTLR cytoplasmic tail bound to parasite homologs of clathrin heavy chain, clathrin-associated adaptor complex small chain, β , γ , and μ 1 adaptins, three components of the AP1 and AP2 adaptor complexes known for clathrin-mediated endocytosis (Sosa et al., 2012). Interestingly, we also identified homologs of the coat complex transport proteins Sec23/Sec24 that ensure the directionality of anterograde membrane flow from the ER to the Golgi apparatus (Lord et al., 2011). Finally, the C terminus of TgSORTLR also pulled down three homologs of retromer-associated vacuolar sorting proteins Vsp9, Vsp26, and Vsp35. This is consistent with TgSORTLR binding to the retromer complex that mediates retrograde transport from endosomes to the *trans*-Golgi network (Kim et al., 2010). To further investigate if these cytosolic cargo-sorting proteins are similarly present in the Golgi-endosomal compartment, we determined the subcellular localization of endogenous Vps26 protein in *T. gondii* by chromosomally appending an encoded HA epitope tag. TgVps26-HA displayed better colocalization with TgSORTLR than any of the previously tested markers (Figure 6A compared to Figure 1D). Altogether, these findings suggest that TgSORTLR functions as a cargo transport protein by coordinating with cytosolic proteins involved in Golgi-endosomal trafficking including TgVsp26. Furthermore, we showed that the Golgi-endosomal localization of TgVps26-HA transiently expressed in the iKO mutant (Figure 6B, left panels) was abolished upon treatment with ATc (Figure 6B, right panels). Similarly, the Golgi-endosomal localization of TgVps26-HA stably expressed in the wild-type parasites (Figure 6C, left panels) was disrupted upon transiently expressing the dominant negative TgSORTLR Δ Ct protein (Figure 6C, right panels). Instead, only cytoplasmic signal was observed. Taken together, these data indicate that endosomal compartments might not form correctly in the absence of TgSORTLR.

Host Infection by *T. gondii* Is Crippled by the Absence of TgSORTLR

As a first assessment of the essentiality of TgSORTLR, we tested the ability of parasites to form plaques on host cell monolayers. Whereas the wild-type parasites grew normally and developed equal sized plaques in the absence or presence of ATc (Figure 7A, yellow arrows), growth of iKO lines gave rise to normal plaque sizes only in the absence of ATc (Figure 7B, left panel). The presence of ATc led to a complete block in plaque formation in TgSORTLR-deficient parasites (Figure 7B, right panel). To

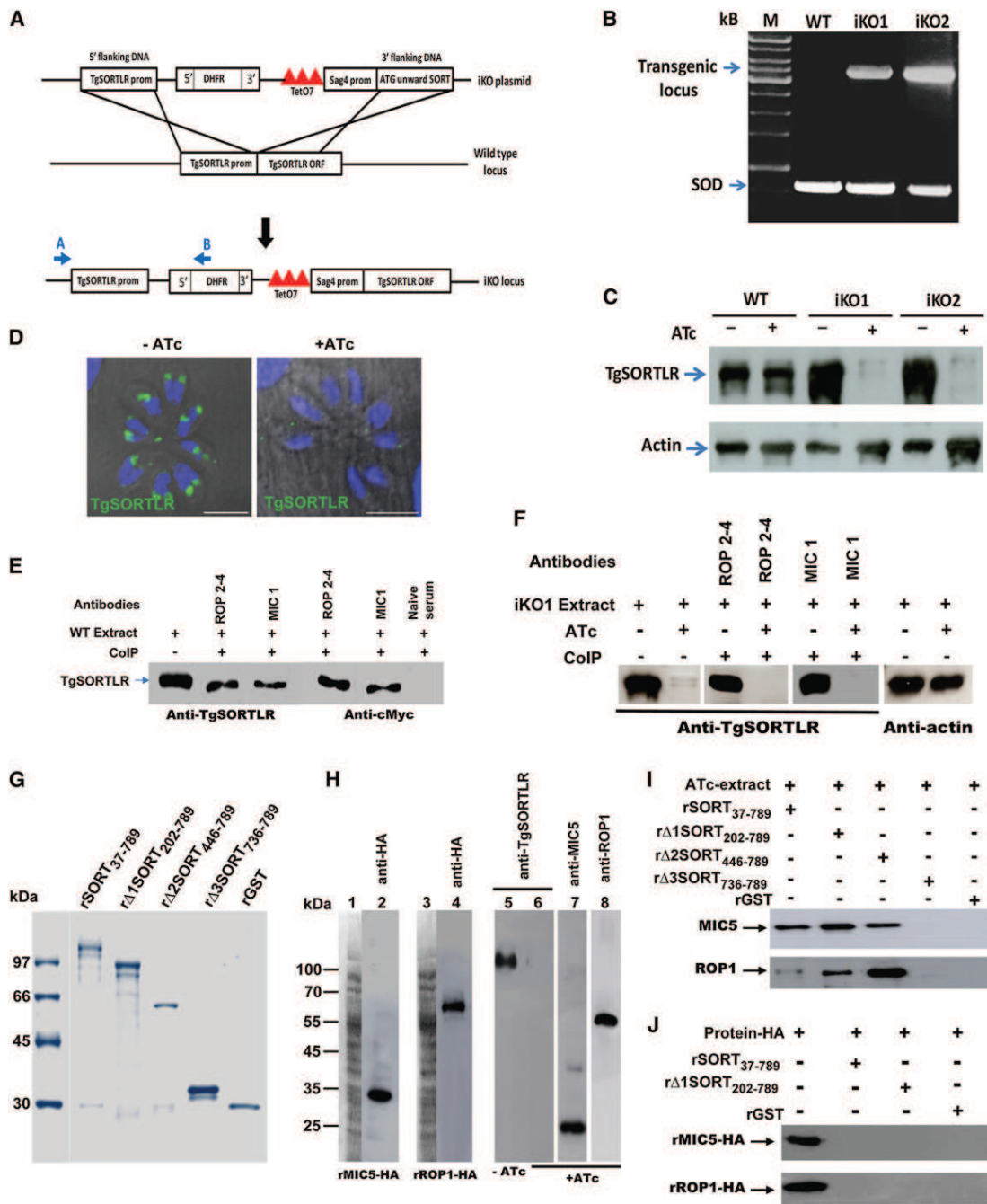


Figure 3. Conditional Ablation of TgSORTLR Gene and Phenotypic Studies

(A) Schematic of the approach used for the conditional ablation of TgSORTLR gene.
 (B) PCR analysis using the primers indicated in (A) confirms the conditional ablation *TgSORTLR*. Superoxide dismutase (SOD), positive control.
 (C) Immunoblots of wild-type and TgSORTLR-deficient lines grown in the presence or absence of ATc for 48 hr and probed with anti-TgSORTLR antibodies. Actin was detected as a loading control.
 (D) Confirmation of TgSORTLR conditional depletion by confocal imaging using anti-TgSORTLR antibodies after 48 hr ATc treatment.
 (E) Immunoblots of wild-type TgSORTLR or ectopic TgSORTLR-cMyc coimmunoprecipitated with monoclonal anti-ROP2-4 or anti-MIC1 antibodies. See also Table S1.
 (F) TgSORTLR does not coimmunoprecipitate with ROP2-4 or MIC1 after conditional depletion with ATc for 48 hr.
 (G) Coomassie blue staining of the purified recombinant luminal domain of TgSORTLR, its different truncations or GST alone.
 (H) Expression of recombinant rMIC5-HA (lane 2) and rROP1-HA (lane 4) proteins by wheat germ in vitro transcription/translation. Immunoblots were probed with rabbit polyclonal anti-HA9 antibodies. Lanes 1 and 3, total protein extracts stained with Ponceau red. Immunoblots of TgSORTLR-deficient mutants grown without (lane 5) or with ATc for 48 hr (lane 6) incubated with anti-TgSORTLR, rabbit polyclonal anti-MIC5 (lane 7), or monoclonal anti-ROP1 (lane 8) antibodies.

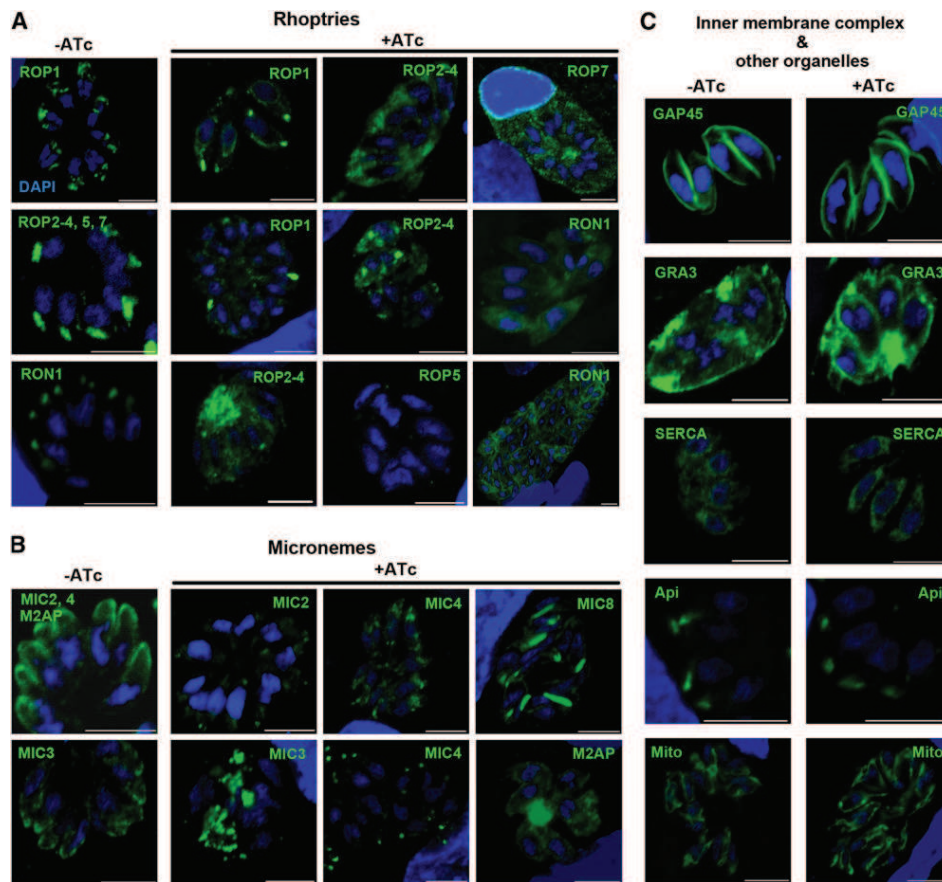


Figure 4. The Conditional Depletion of TgSORTLR Mislocalizes Microneme and Rhoptry Proteins

(A) In the presence of ATc for 48 hr, several rhoptry ROP1, ROP2-4, ROP5, ROP7, and RON1 proteins were mislocalized in TgSORTLR-deficient mutant, whereas the apical localization was observed in the absence of ATc, as expected. Only one confocal image of wild-type parasite corresponding to the first anti-ROP antibody is shown when several proteins have an identical pattern.

(B) Several microneme proteins including MIC2, MIC3, MIC4, and MA2P were mislocalized in TgSORTLR-deficient mutant after 48 hr ATc treatment. Only one confocal image of wild-type parasite was shown as above.

(C) Subcellular localizations of the inner membrane complex (GAP45), dense granules (GRA3), endoplasmic reticulum (SERCA), apicoplast (Atrx2; Api), or mitochondrial (HSP60; Mito) markers were unchanged in the TgSORTLR deficient mutant. These data indicate that TgSORTLR regulates ROP and MIC protein transport during rhoptry and microneme biogenesis (see also Figure S3).

assess egress, live video microscopy was used to visualize parasite exit in response to induction with the calcium ionophore A23187. The iKO parasites efficiently egressed from host cells when grown without ATc (Movie S2, left), yet it failed to egress upon ionophore stimulation after growth in ATc (Movie S2, right). To monitor if the parasites depleted for TgSORTLR can still glide, iKO mutants grown in ATc were mechanically released from infected host cells, and trail deposition assays performed. Figure 7C shows that the iKO line grown in the absence of ATc showed a significant number of circular trails (yellow arrows, left panel). In contrast, conditional ablation of TgSORTLR paralyzed the parasite (Figure 7C, right panel). As expected based on the absence of gliding, mechanically released iKO mutants

grown in ATc were also severely impaired in their ability to invade host cells (Figure 7D).

To examine the role of the TgSORTLR in *Toxoplasma* infection, mice were infected with iKO parasites that were cultured without ATc, and TgSORTLR expression levels were suppressed in vivo by providing ATc in the drinking water. Strikingly, all mice inoculated with iKO parasites in the absence of ATc treatment succumbed to the infection by day 9, whereas ATc-treated infected mice survived regardless of whether treatment was initiated on day 0 postinfection or day 5 postinfection, when the mice exhibit symptomatic acute toxoplasmosis (Figure 7E). This demonstrates that mice can be cured of acute toxoplasmosis by suppression of TgSORTLR expression. Similarly,

(I) Pull-down assays using the protein extracts from TgSORTLR-deficient mutant and rSORT₃₇₋₇₈₉-GST, rΔ1SORT₂₀₂₋₇₈₉-GST, rΔ2SORT₄₄₆₋₇₈₉-GST, or rΔ3SORT₇₃₆₋₇₈₉-GST (see schematic representations in Figure 1A). Blots were incubated with rabbit polyclonal anti-MIC5 or anti-ROP1 antibodies.

(J) Pull-down assays using protein extracts containing rMIC5-HA or rROP1-HA protein incubated with rSORT₃₇₋₇₈₉-GST or rΔ1SORT₂₀₂₋₇₈₉-GST beads (upper panel). Blots were stained with rabbit anti-HA antibodies.

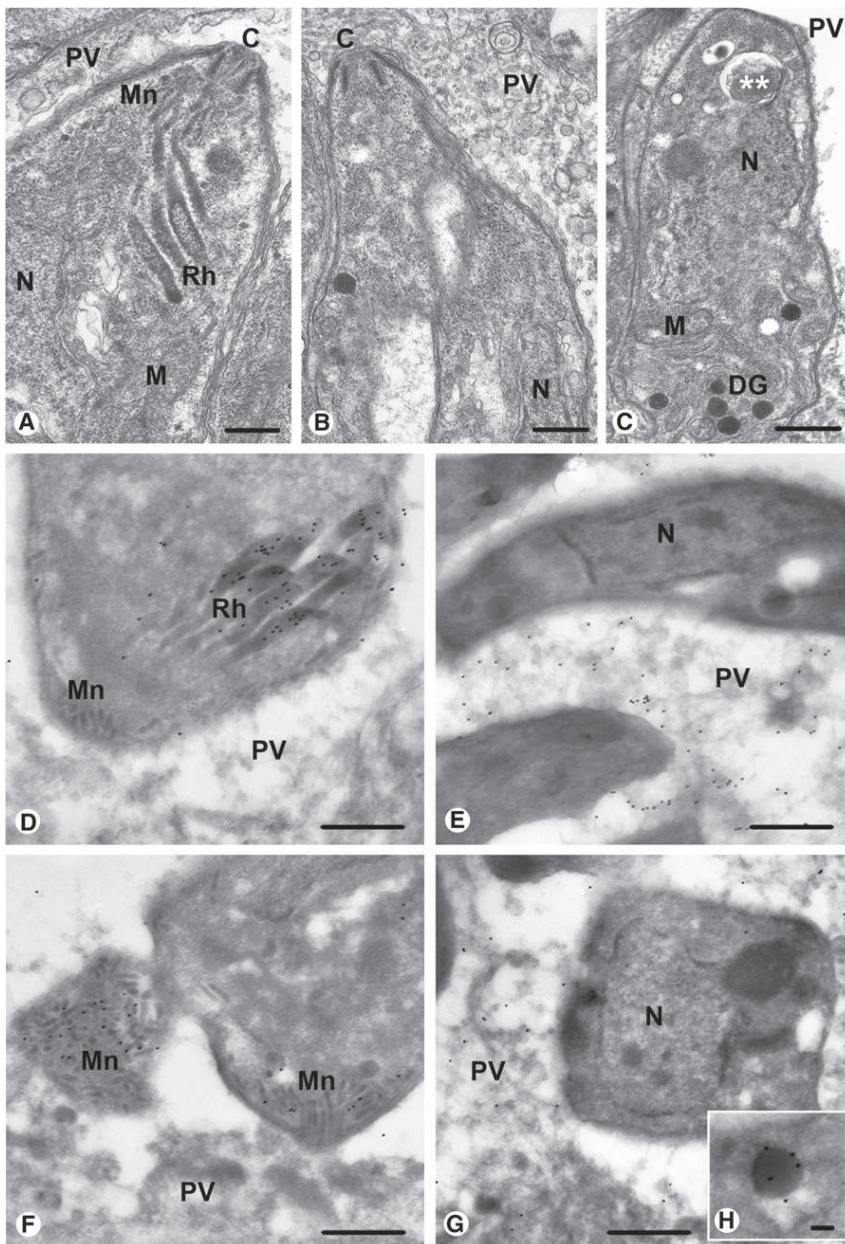


Figure 5. TgSORTLR Is Crucial for the Biogenesis of Micronemes and Rhoptries

(A) Transmission electron micrograph showing rhoptries (Rh) and micronemes (Mn) in the wild-type parasite. (B) Ultrathin section of a TgSORTLR-deficient parasite lacking typical micronemes and rhoptries after ATc treatment for 72 hr. (C) Longitudinal ultrathin section of TgSORTLR-deficient parasite lacking typical micronemes and rhoptries but accumulating unknown vesicles (asterisks) after ATc treatment for 72 hr. See also Figure S4. (D) Cryo-IEM of rhoptries in the wild-type parasite using anti-ROP1 antibodies. (E) In the TgSORTLR-deficient parasites, gold-labeled ROP1 was only detected in the vacuole after ATc treatment for 48 hr. (F) Cross-sections through the anterior end of wild-type parasite showing numerous gold-labeled micronemes using anti-MIC5 antibodies. (G) TgSORTLR-deficient parasites showed gold-labeled MIC5 mainly in the PV after ATc treatment for 48 hr as a result of mistargeting from the microneme pathway. The remaining gold-labeled MIC5 was detected in the dense granules (DG, see the inset (H)). N, nucleus; PV, parasitophorous vacuole; M, mitochondrion. Bars, 500 nm for all panels.

not confer sterile immunity, perhaps because of the rapid clearance of iKO parasites.

DISCUSSION

We showed that TgSORTLR localizes to the Golgi cisternae and proximal vesicles and acts as a cargo receptor that is necessary for the biogenesis of secretory organelles. Further, TgSORTLR is essential for host cell egress, gliding motility, cell invasion, and in vivo infection by *T. gondii*, likely because of the crucial roles for apical secretory organelles in all events. Despite its high conservation throughout the tree of life, sortilin has not been reported to be essential in any

mice infected with the TgSORTLR iKO at higher doses survived the acute infection when treated with ATc beginning at day 0 postinfection (Figure 7F). Moreover, all mice survived the iKO infection when ATc treatment was stopped at day 15, suggesting complete clearance of infection (Figure 7F, black arrow). We conclude that the conditional ablation of TgSORTLR transformed a *T. gondii* virulent strain into nonlethal parasites. Mice that survived the primary infection with iKO parasites were challenged with a lethal dose of virulent wild-type RH parasites. Unexpectedly, all of the mice previously infected with iKO parasites succumbed to the challenge infection, albeit with a modest delay that was dependent upon the dose of the primary infection (Figure 7G). This suggests that infection with iKO parasites does

other biological system (Marcusson et al., 1994; Jansen et al., 2007). In contrast, our findings reveal an indispensable role for *T. gondii* sortilin, which is the indication that, unlike yeast and higher eukaryotes, lower branching eukaryotes such as apicomplexan parasites do not have redundant mechanisms of protein trafficking. This provides a unique glimpse into a “simplified” sorting system in lower eukaryotes. We suggest that TgSORTLR guides ROP and MIC proteins through the Golgi for packaging into vesicles at the *trans*-Golgi network. Supporting this notion, we observed Golgi proximal, proROP4, and TgSORTLR-positive structures that could be transport vesicles destined for the endolysosomal system. Regardless, TgSORTLR likely plays a role in the formation and packaging of these vesicles based on the

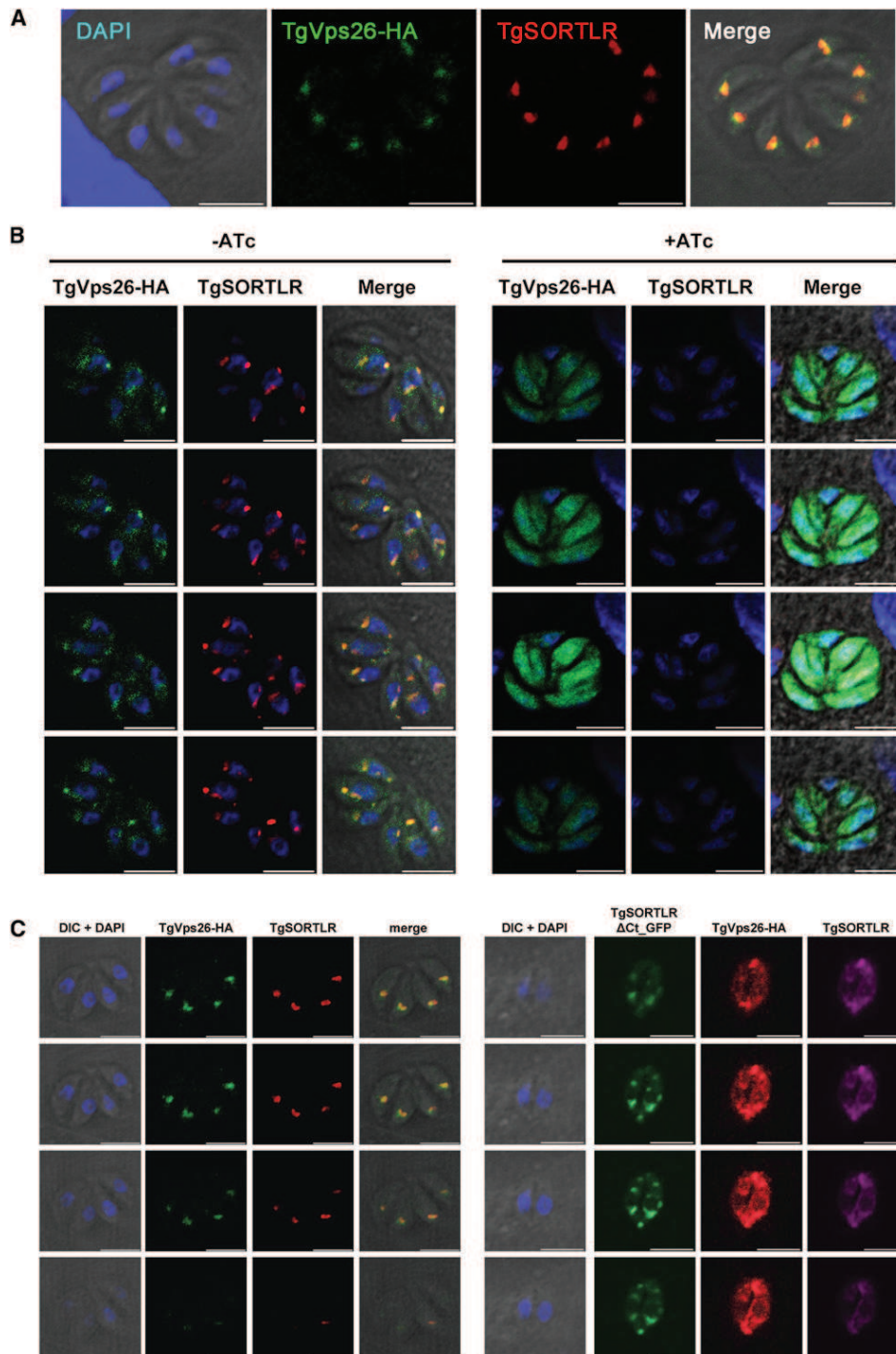


Figure 6. TgSORTLR Colocalizes with TgVsp26 and Dictates Its Membrane Association

(A) Extensive colocalization between TgSORTLR (mouse polyclonal antibody, red) and endogenously tagged TgVps26-HA (rabbit polyclonal antibody, green). (B) Z stack acquisitions by confocal microscopy of transiently transfected iKO tachyzoites expressing TgVSP26-HA under the control of the GRA1 and grown in the absence (left panels) or presence (right panels) of ATc for 48 hr. Antibodies used are the same as in (A). (C) Z stack confocal acquisitions of tachyzoites stably expressing endogenously tagged TgVps26-HA in RHΔku80 strain before (left panels) or after (right panels) transient transfection with TgSORTLR lacking its C-terminal tail. These data indicate that the C-terminal tail of TgSORTLR recruits cytoplasmic sorting proteins (see also Table S2). The confocal images were captured in increments of 0.37 or 0.43 μm for (B) and (C), respectively. Bars, 5 μm for all panels.

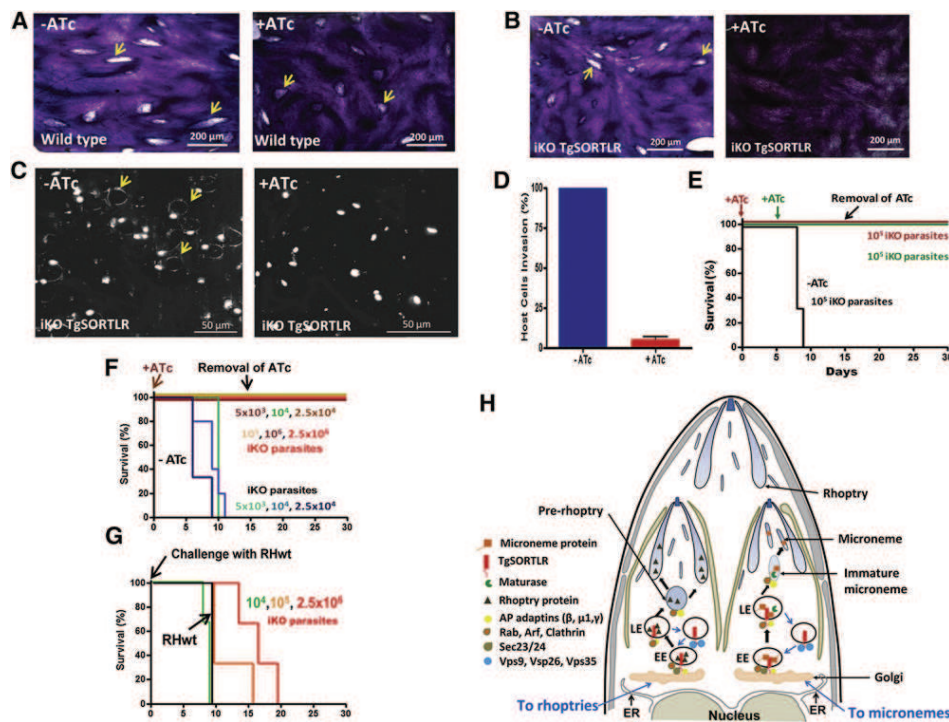


Figure 7. TgSORTLR Is Essential for Gliding Motility, Host Infection, and Parasite Virulence

- (A) The wild-type parasites grown in the presence or absence of ATc after 48 hr.
- (B) The conditional TgSORTLR depletion dramatically prevents the lytic and invasion cycles, thus affecting plaque formation in the presence of ATc for 48 hr. Bars, 200 μ m.
- (C) Motility was impaired in the conditional TgSORTLR-deficient line after it was grown in ATc for 48 hr (see also *Movie S2*). Bars, 50 μ m.
- (D) Host cell invasion was severely impaired in TgSORTLR deficient parasites after ATc treatment for 48 hr. Bars represent means \pm standard deviations ($n = 3$, $p < 0.001$ by Student's *t* test).
- (E) TgSORTLR depletion abrogates virulence in mice. A group of mice infected with TgSORTLR-deficient mutants for 5 days (green) prior to the addition of ATc to drinking water.
- (F) A group of six mice inoculated with doses ranging from 5×10^3 up to 2.5×10^6 iKO parasites survived with ATc treatment.
- (G) TgSORTLR-depleted parasites are avirulent in mice but do not confer protective immunity to reinfection with 1×10^3 of wild-type RH tachyzoites.
- (H) Model of TgSORTLR function in protein sorting and the biogenesis of apical secretory organelles.

properties of sortilin through its cytoplasmic tail known to specifically interact with the monomeric adaptor protein, Golgi-associated gamma-adaptin homologous and ARF-interacting proteins (Hermeijer, 2009). We provide evidence that the cytoplasmic tail of TgSORTLR specifically interacts with AP adaptins (β , γ , and $\mu 1$ adaptins), clathrin, and one of its associated factors and three vesicular sorting proteins (Vsp9, Vsp26, and Vps35). Moreover, expression of C-terminal tail mutant in *T. gondii* disrupted the subcellular distribution of endogenous TgSORTLR and that of ROP and MIC proteins, suggesting that the TgSORTLR cytoplasmic tail contains potent signals for Golgi-endosome sorting.

Whereas regulated secretory granules of higher eukaryotes are derived directly from the *trans*-Golgi network (Sossin et al., 1990), *T. gondii* and possibly other apicomplexan parasites appear to use their endolysosomal system as an intermediate conduit for trafficking of proteins destined for regulated secretion. While Joiner and colleagues proposed this for the rhoptries (Ngô et al., 2004), more recent studies suggest that microneme proteins also traffic through the endolysosomal system. Microneme proproteins reside in late endosomes (Harper et al., 2006), and at least two such proteins are proteolytically

processed by an endolysosomal cathepsin L protease (Parussini et al., 2010). Thus, one advantage of merging the endocytic and exocytic pathways is that it allows convenient access to proteases necessary for the activation of secretory proteins. A second advantage revealed by the current study is that the parasite utilizes an endosomal cargo receptor, TgSORTLR, for sorting proteins to the regulated secretory organelles. Our work further suggests that, likely due to the lack of a mannose-6-phosphate sorting system, *T. gondii* appears to critically rely on TgSORTLR for shuttling secretory proteins through the hybrid endocytic/exocytic system. Thus, the identification of TgSORTLR promises to provide insights into the evolutionary development and relationships of basic cellular functions in protists. The biogenesis of these parasite-specific organelles, their evolutionary origins and relationships to organelles of higher eukaryotes, as well as the precise sorting events that endow them with distinctive sets of proteins are poorly understood. Despite being touted as a multiprotein receptor, precisely how sortilin recognizes different cargo proteins from rhoptries and micronemes remains to be precisely elucidated. Rhoptries are formed prior to micronemes during daughter cell formation (Nishi et al., 2008),

suggesting that TgSORTLR sorts ROP proteins prior to attending to MIC proteins. We showed that TgSORTLR binding to ROP and MIC proteins requires the presence of the N-terminal bladed propeller domains. This observation is consistent with TgSORTLR being a receptor that binds to multiple ligands, a property shared with mammalian sortilin. For example, the neurotensin C terminus binds to mammalian sortilin in a tunnel formed by the β -propeller domains (Quistgaard et al., 2009), whereas proneurotrophin binds to a linear surface-exposed sequence (Serup Andersen et al., 2010). Rhoptry and microneme proteins may bind to TgSORTLR in a similar manner. Finally, we found that recombinant ROP and MIC proteins do not interact with the recombinant TgSORTLR, suggesting that posttranslational modifications are required for proper binding to this cargo receptor. These putative posttranslational modifications may allow discrimination in binding of ROP and MIC proteins to TgSORTLR. Alternatively, ROP and MIC proteins might bind to TgSORTLR in multiprotein complexes with only a subset of proteins forming direct contacts with the cargo receptor. We propose a model that during daughter parasite formation, TgSORTLR binds to rhoptry (Figure 7H, left panel) or microneme (Figure 7H, right panel) proteins in the Golgi lumen and recruits cytosolic cargo sorting proteins such as AP adaptins, clathrin, and vacuolar sorting proteins. This leads to the formation of a sorting complex, which is presumably associated with vesicles that exit from the external Golgi cisternae. TgSORTLR then guides its cargo through the early endosomes (EE), possibly the late endosomes (LE), prerhoptries, or immature micronemes before releasing its payload. The “empty” cargo receptor then recruits components of the retromer complex for retrograde translocation and recycling to the Golgi to reload with new cargo. Our ability to conditionally extinguish sortilin expression will allow functional dissection that should reveal valuable insights into basic mechanisms of sortilin function and protein-protein interactions. The TgSORTLR-deficient parasites described herein underscore its key role and suggest that TgSORTLR plays a crucial role as a type I transmembrane cargo receptor regulating a common sorting mechanism that is important for the biogenesis of apical secretory organelles in *T. gondii*.

EXPERIMENTAL PROCEDURES

Growth of Host Cells and Parasite Strains

T. gondii RH strain, RHTaTi (a Δ Trans-activator Δ trap-identified inducible anhydrotetracycline [ATc] strain; Meissner et al. [2002]), RH Δ ku80 (a strain with high homologous integration of transfected DNA; Huynh and Carruthers [2009]), RH Δ ku80TaTi (a strain with ATc inducible system and high homologous recombination; Sheiner et al. [2011]), RH strain expressing GRASP-RFP, or HDEL-RFP were also used.

Generation of Transgenic *T. gondii* Strains

The RH tachyzoites (5×10^6) were transfected with 50 μ g of the full-length TgSORTLR or truncated TgSORTLR Δ Ct or TgSORTLR Δ Nt genes amplified by PCR and fused to GFP or HA epitope tag. All primers used for PCR during this study are listed in Supplemental Experimental Procedures. Ectopic expression of TgSORTLR-cMyc driven by TetO7SAG1 promoter and stable lines were obtained in RHTaTi strain using 5 μ g/ml of phleomycin. The TgSORTLR iKO lines were generated using RH Δ ku80TaTi and pG13-D-T7S4 plasmid containing 2 kb 5' and 3' noncoding genomic DNA, respectively (Sheiner et al., 2011). For HA-tagged TgVps26 protein by knockin strategy, a DNA fragment of 2 kb upstream of the stop codon from the TGME49_063500 genomic sequence was cloned in the pLIC-HA-DHFR plasmid (Huynh and

Carruthers, 2009). The plasmid (25 μ g) was linearized with EcoRV and transfected in 5×10^6 Ku80 tachyzoites followed by pyrimethamine selection. The full-length cDNA encoding TgVps26 protein was PCR amplified. The amplified cDNA was cloned in frame with GFP in pMAH14 vector containing GRA1 promoter as previously described (Fauquenoy et al., 2011).

Immunofluorescence Assay, Electron and Live Video Microscopy

Paraformaldehyde-fixed intracellular parasites were used for IFA and confocal microscopy as previously described (Olguin-Lamas et al., 2011). Ultrastructural morphology was performed using conventional and immuno-gold electron microscopy (Olguin-Lamas et al., 2011), except that 8% paraformaldehyde containing 0.01% glutaraldehyde was used for cryo-IEM. Egress was induced with 4 μ M A23187 (Sigma) at 37°C under CO₂ chamber, and videos were recorded with a Zeiss AxioObserver Z1 microscope. The list of monoclonal and polyclonal antibodies used during this study is shown in Supplemental Experimental Procedures. Confocal images were analyzed by ImageJ (NIH) software, and colocalization between green and red confocal images was determined by Pearson's coefficient using 4–7 vacuoles containing at least 4–8 parasites per vacuole.

Coimmunoprecipitation, Proteomic, and Immunoblot Analyses

TgSORTLR and binding partners were coimmunoprecipitated after crosslinking 2×10^9 live tachyzoites with dithiobis-[succinimidyl-propionate] (DSP) followed by extraction with Triton X-100 (Fauquenoy et al., 2011). Binding protein complexes were reduced by dithiothreitol before SDS-PAGE followed by silver staining. The gel lane was cut into slices, in-gel tryptic digested, and then processed for mass spectrometry as described in Supplemental Experimental Procedures.

Synthesis of Recombinant Proteins in *E. coli* and Wheat Germ In Vitro Translation

For expression in *E. coli*, TgSORTLR DNA fragments corresponding to amino acids 37–714 or 37–789 termed rSORT₃₇₋₇₁₄ or rSORT₃₇₋₇₈₉ were amplified by PCR. The amplified DNA was cloned in frame with GST in pGEX-6P (Pharmacia). Three N terminus TgSORTLR mutants, rSORT₂₀₂₋₇₈₉-GST, rSORT₄₄₆₋₇₈₉-GST, or rSORT₇₃₆₋₇₈₉-GST, corresponding to the respective deletion of bladed propeller domains B–D, B–H or B–10CCb, were also produced. The recombinant rSORT₈₀₈₋₉₆₂-GST fusion protein corresponding to TgSORTLR cytoplasmic tail was also generated as above. Production of fusion proteins in BL21 (DE3) bacteria and purification were performed by standard techniques. The full-length recombinant rROP1-HA and rMIC5-HA proteins were synthesized using the wheat germ in vitro transcription/translation kit (Promega). Five Balb/C mice were intraperitoneally (IP) immunized with 50 μ g of purified recombinant N-terminal TgSORTLR₃₇₋₇₈₉-GST protein per mouse in complete Freund's adjuvant (Sigma). Two boosts were performed by IP with 50 μ g of recombinant protein per mouse in incomplete Freund's adjuvant (Sigma).

GST Pull-Down Experiments

Total parasite extract from 10^8 tachyzoites or 20 μ l of in vitro translation reaction were resuspended in GST binding buffer (Tris.HCl 10 mM, NaCl 150 mM, MnCl₂ 1 mM, CaCl₂ 1 mM, Triton X-100 1%, and protease inhibitors) and incubated with 10–40 μ l of beads containing 3–6 μ g of rSORT₃₇₋₇₈₉-GST, rSORT₂₀₂₋₇₈₉-GST, rSORT₄₄₆₋₇₈₉-GST, and rSORT₇₃₆₋₇₈₉-GST or GST beads alone for overnight at 4°C. Precipitants were washed with the same buffer, eluted by SDS buffer and analyzed by western blots. For the pull-down of TgSORTLR C-terminal tail partners, cytoplasmic extracts from 4×10^9 tachyzoites were incubated with 100 μ l of rSORT₈₀₈₋₉₆₂-GST or to GST beads alone containing about 200 μ g of each protein. The bound partners were silver stained and processed for proteomics analyses.

Phenotypic Assays

Plaque assays were performed using 8-well Labtek slides containing human fibroblast cells infected with 100 parasites per well in media with or without 1 μ g/ml ATc, fixed after 7 days post infection and stained with crystal violet. Experiments were repeated thrice in triplicate wells. Intracellular TgSORTLR iKO parasites grown with or without ATc were released mechanically and used in three independent experiments each for gliding motility and invasion

assays as previously described (Supplemental Experimental Procedures; Fauquenoy et al., 2008, 2011).

Mouse Infectivity Studies

Six female Balb/C mice (6–8 weeks old) were injected intraperitoneally with 5×10^3 to 2×10^6 iKO parasites that were cultured without ATc. To suppress TgSORTLR expression in vivo, the drinking water was supplemented with 0.2 mg/ml of ATc and 5% of sucrose. A second group of six mice were also infected with 5×10^3 to 2.5×10^4 of iKO tachyzoites without ATc. As control, 5×10^3 to 2.5×10^4 of the parental line (RHΔku80TaTi strain) was used to infect groups of six mice with or without ATc. Survival was monitored over a period of 4 weeks. All experiments were repeated at least three times.

Statistics

Student's t test was used for statistical analysis. The Mann-Whitney test was used for analysis of mice survival curves.

SUPPLEMENTAL INFORMATION

Supplemental Information includes four figures, two tables, Supplemental Experimental Procedures, and two movies and can be found with this article online at doi:10.1016/j.chom.2012.03.006.

ACKNOWLEDGMENTS

We thank J.-F. Dubremetz for the anti-ROP, anti-MIC antibodies; B. Striepen and L. Sheiner for RHΔku80TaTi and the pG13-D-T7S4 plasmid; M. Meissner for DrpB-YFP strain; G. Ward for anti-proROP4 antibodies; D. Soldati for anti-GAP45 antibodies; D. Sibley for SERCA antibodies; and P. Bradley for the Atrx2 antibody. We are grateful to Gordon Langsley and Steven Ball for critical reading of this manuscript. This work was supported by a Ph.D. fellowship from the French Ministry of Research and Technology to P.J.S., by financial supports from the CNRS, INSERM, Pasteur Institute of Lille to S.T., and from the NIH (AI063263) to V.B.C.

Received: October 27, 2011

Revised: February 2, 2012

Accepted: March 28, 2012

Published: May 16, 2012

REFERENCES

- Boothroyd, J.C., and Dubremetz, J.F. (2008). Kiss and spit: the dual roles of *Toxoplasma* rhoptries. *Nat. Rev. Microbiol.* 6, 79–88.
- Breinich, M.S., Ferguson, D.J., Foth, B.J., van Dooren, G.G., Lebrun, M., Quon, D.V., Striepen, B., Bradley, P.J., Frischknecht, F., Carruthers, V.B., and Meissner, M. (2009). A dynamin is required for the biogenesis of secretory organelles in *Toxoplasma gondii*. *Curr. Biol.* 19, 277–286.
- Carey, K.L., Jongco, A.M., Kim, K., and Ward, G.E. (2004). The *Toxoplasma gondii* rhoptry protein ROP4 is secreted into the parasitophorous vacuole and becomes phosphorylated in infected cells. *Eukaryot. Cell* 3, 1320–1330.
- Chaturvedi, S., Qi, H., Coleman, D., Rodriguez, A., Hanson, P.I., Striepen, B., Roos, D.S., and Joiner, K.A. (1999). Constitutive calcium-independent release of *Toxoplasma gondii* dense granules occurs through the NSF/SNAP/SNARE/Rab machinery. *J. Biol. Chem.* 274, 2424–2431.
- Clee, S.M., Yandell, B.S., Schueler, K.M., Rabaglia, M.E., Richards, O.C., Raines, S.M., Kabara, E.A., Klass, D.M., Mui, E.T., Stapleton, D.S., et al. (2006). Positional cloning of Sorcs1, a type 2 diabetes quantitative trait locus. *Nat. Genet.* 38, 688–693.
- DeRocher, A.E., Coppens, I., Karnataki, A., Gilbert, L.A., Rome, M.E., Feagin, J.E., Bradley, P.J., and Parsons, M. (2008). A thioredoxin family protein of the apicoplast periphery identifies abundant candidate transport vesicles in *Toxoplasma gondii*. *Eukaryot. Cell* 7, 1518–1529.
- Fauquenoy, S., Morelle, W., Hovasse, A., Bednarczyk, A., Slomianny, C., Schaeffer, C., Van Dorsselaer, 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.
- Fauquenoy, S., Hovasse, A., Sloves, P.J., Morelle, W., Dilezitoko Alayi, T., Slomianny, C., Werkmeister, E., Schaeffer, C., Van Dorsselaer, A., and Tomavo, S. (2011). Unusual N-glycan structures required for trafficking *Toxoplasma gondii* GAP50 to the inner membrane complex regulate host cell entry through parasite motility. *Mol. Cell. Proteomics* 10, M111.008953.
- Hager, K.M., Striepen, B., Tilney, L.G., and Roos, D.S. (1999). The nuclear envelope serves as an intermediary between the ER and Golgi complex in the intracellular parasite *Toxoplasma gondii*. *J. Cell Sci.* 112, 2631–2638.
- Harper, J.M., Huynh, M.H., Coppens, I., Parussini, F., Moreno, S., and Carruthers, V.B. (2006). A cleavable propeptide influences *Toxoplasma* infection by facilitating the trafficking and secretion of the TgMIC2-M2AP invasion complex. *Mol. Biol. Cell* 17, 4551–4563.
- Hermey, G. (2009). The Vps10p-domain receptor family. *Cell. Mol. Life Sci.* 66, 2677–2689.
- Hoppe, H.C., Ngô, H.M., Yang, M., and Joiner, K.A. (2000). Targeting to rhoptry organelles of *Toxoplasma gondii* involves evolutionarily conserved mechanisms. *Nat. Cell Biol.* 2, 449–456.
- Huynh, M.H., and Carruthers, V.B. (2006). *Toxoplasma* MIC2 is a major determinant of invasion and virulence. *PLoS Pathog.* 2, e84.
- Huynh, M.H., and Carruthers, V.B. (2009). Tagging of endogenous genes in a *Toxoplasma gondii* strain lacking Ku80. *Eukaryot. Cell* 8, 530–539.
- Jansen, P., Giehl, K., Nyengaard, J.R., Teng, K., Lioubinski, O., Sjoegaard, S.S., Breiderhoff, T., Gotthardt, M., Lin, F., Eilers, A., et al. (2007). Roles for the pro-neurotrophin receptor sortilin in neuronal development, aging and brain injury. *Nat. Neurosci.* 10, 1449–1457.
- Kim, T., and Hempstead, B.L. (2009). NRH2 is a trafficking switch to regulate sortilin localization and permit proneurotrophin-induced cell death. *EMBO J.* 28, 1612–1623.
- Kim, E., Lee, Y., Lee, H.J., Kim, J.S., Song, B.S., Huh, J.W., Lee, S.R., Kim, S.U., Kim, S.H., Hong, Y., et al. (2010). Implication of mouse Vps26b-Vps29-Vps35 retromer complex in sortilin trafficking. *Biochem. Biophys. Res. Commun.* 403, 167–171.
- Lord, C., Bhandari, D., Menon, S., Ghassemian, M., Nycz, D., Hay, J., Ghosh, P., and Ferro-Novick, S. (2011). Sequential interactions with Sec23 control the direction of vesicle traffic. *Nature* 473, 181–186.
- Marcusson, E.G., Horazdovsky, B.F., Cereghino, J.L., Gharakhanian, E., and Emr, S.D. (1994). The sorting receptor for yeast vacuolar carboxypeptidase Y is encoded by the VPS10 gene. *Cell* 77, 579–586.
- Meissner, M., Schlüter, D., and Soldati, D. (2002). Role of *Toxoplasma gondii* myosin A in powering parasite gliding and host cell invasion. *Science* 298, 837–840.
- Nagamune, K., Beatty, W.L., and Sibley, L.D. (2007). Artemisinin induces calcium-dependent protein secretion in the protozoan parasite *Toxoplasma gondii*. *Eukaryot. Cell* 6, 2147–2156.
- Ngô, H.M., Yang, M., Paprotka, K., Pypaert, M., Hoppe, H., and Joiner, K.A. (2003). AP-1 in *Toxoplasma gondii* mediates biogenesis of the rhoptry secretory organelle from a post-Golgi compartment. *J. Biol. Chem.* 278, 5343–5352.
- Ngô, H.M., Yang, M., and Joiner, K.A. (2004). Are rhoptries in Apicomplexan parasites secretory granules or secretory lysosomal granules? *Mol. Microbiol.* 52, 1531–1541.
- Nishi, M., Hu, K., Murray, J.M., and Roos, D.S. (2008). Organellar dynamics during the cell cycle of *Toxoplasma gondii*. *J. Cell Sci.* 121, 1559–1568.
- Olguin-Lamas, A., Madec, E., Hovasse, A., Werkmeister, E., Callebaut, I., Slomianny, C., Delhaye, S., Mouveaux, T., Schaeffer-Reiss, C., Van Dorsselaer, A., and Tomavo, S. (2011). A novel *Toxoplasma gondii* nuclear factor TgNF3 is a dynamic chromatin-associated component, modulator of nucleolar architecture and parasite virulence. *PLoS Pathog.* 7, e1001328.
- Parussini, F., Coppens, I., Shah, P.P., Diamond, S.L., and Carruthers, V.B. (2010). Cathepsin L occupies a vacuolar compartment and is a protein maturase within the endo/exocytic system of *Toxoplasma gondii*. *Mol. Microbiol.* 76, 1340–1357.

- Pelletier, L., Stern, C.A., Pypaert, M., Sheff, D., Ngô, H.M., Roper, N., He, C.Y., Hu, K., Toomre, D., Coppens, I., et al. (2002). Golgi biogenesis in *Toxoplasma gondii*. *Nature* **418**, 548–552.
- Quistgaard, E.M., Madsen, P., Grøftehauge, M.K., Nissen, P., Petersen, C.M., and Thirup, S.S. (2009). Ligands bind to Sortilin in the tunnel of a ten-bladed beta-propeller domain. *Nat. Struct. Mol. Biol.* **16**, 96–98.
- Rogaeva, E., Meng, Y., Lee, J.H., Gu, Y., Kawarai, T., Zou, F., Katayama, T., Baldwin, C.T., Cheng, R., Hasegawa, H., et al. (2007). The neuronal sortilin-related receptor SORL1 is genetically associated with Alzheimer disease. *Nat. Genet.* **39**, 168–177.
- Serup Andersen, O., Boisguerin, P., Glerup, S., Skeldal, S., Volkmer, R., Willnow, T.E., Nykjaer, A., and Andersen, O.M. (2010). Identification of a linear epitope in sortilin that partakes in pro-neurotrophin binding. *J. Biol. Chem.* **285**, 12210–12222.
- Sheiner, L., Demerly, J.L., Poulsen, N., Beatty, W.L., Lucas, O., Behnke, M.S., White, M.W., and Striepen, B. (2011). A systematic screen to discover and analyze apicoplast proteins identifies a conserved and essential protein import factor. *PLoS Pathog.* **7**, e1002392.
- Sosa, R.T., Weber, M.M., Wen, Y., and O'Halloran, T.J. (2012). A single β adaptin contributes to AP1 and AP2 complexes and clathrin function in *Dictyostelium*. *Traffic* **13**, 305–316. Published online December 4, 2011. 10.1111/j.1600-0854.2011.01310.x.
- Sossin, W.S., Fisher, J.M., and Scheller, R.H. (1990). Sorting within the regulated secretory pathway occurs in the *trans*-Golgi network. *J. Cell Biol.* **110**, 1–12.
- Stedman, T.T., Sussmann, A.R., and Joiner, K.A. (2003). *Toxoplasma gondii* Rab6 mediates a retrograde pathway for sorting of constitutively secreted proteins to the Golgi complex. *J. Biol. Chem.* **278**, 5433–5443.
- Toursel, C., Dziarszinski, F., Bernigaud, A., Mortuaire, M., and Tomavo, S. (2000). Molecular cloning, organellar targeting and developmental expression of mitochondrial chaperone HSP60 in *Toxoplasma gondii*. *Mol. Biochem. Parasitol.* **111**, 319–332.

Conclusion et perspectives

Nous avons abordé dans ce projet d'étude du parasite *T. gondii*, 3 aspects fondamentaux de la biologie du parasite. Grâce à des stratégies basées sur des approches développées pour l'étude des complexes protéiques présents en faibles quantités et l'utilisation combinée de systèmes nanoLC-MS/MS, nous avons pu identifier les différentes protéines de divers complexes protéiques qui ont permis d'approfondir les fonctions biologiques de plusieurs protéines d'intérêt. Ainsi nous avons apporté des éléments de réponse moléculaires nécessaires à la compréhension de la biologie du parasite *T. gondii*.

Nous avons identifié les protéines partenaires de la glycoprotéine TgGAP50 et démontré l'importance de la glycosylation de cette protéine dans le maintien et la stabilité du "glidéosome", ce complexe protéique nécessaire à la motilité du parasite. La caractérisation des glycosylations de TgGAP50 et l'identification des différents partenaires de TgGAP50 combinées aux différentes expériences de biologie moléculaire ont permis de mettre en évidence le rôle clé de cette glycoprotéine et de la N-glycosylation dans la motilité parasitaire et dans l'invasion du parasite à l'intérieur des cellules hôtes.

Nous avons également identifié les différents partenaires des protéines TgAlba et avons montré que ces TgAlba interagissent avec un large complexe de protéines. Les résultats d'analyses protéomiques, en combinaison avec les expériences de biologie moléculaire ont permis de démontrer la double fonction de régulation transcriptionnelle et traductionnelle conservée chez ces protéines TgAlba.

Nous avons aussi identifié les différents partenaires recrutés par la sortiline de *T. gondii*. L'étude des protéines partenaires des différents domaines de TgSORTLR (domaine luminal N-terminal de la protéine et cytoplasmique C-terminal) ont permis de montrer que les protéines des micronèmes et des rhoptries sont spécifiquement liées au N-terminal de TgSORTLR. Tandis que l'extrémité cytoplasmique C-terminale de TgSORTLR, qui se fixe spécifiquement aux protéines endosomales, joue un rôle capital dans le trafic antérograde et rétrograde du récepteur, dans sa localisation subcellulaire et ses fonctions biologiques. Nous avons montré pour la première fois que ce récepteur membranaire clé est impliqué dans le tri des protéines, la biogenèse d'organites apicaux et l'infection de l'hôte.

Les difficultés majeures rencontrées au cours de ce travail sont liées à la faible quantité de matériel disponible pour certains échantillons, qui limitait donc le nombre de replicats techniques d'analyse et exacerbait les problèmes de contamination par les kératines, classiquement rencontrés en analyse protéomique. En effet, lorsque les quantités de matériel sont faibles, des précautions encore plus drastiques doivent être prises, à toutes les étapes de l'étude d'un complexe protéique (des manipulations génétiques des parasites jusqu'à l'injection dans le système nanoLC-MS/MS) pour éviter que l'analyse protéomique de certains échantillons ne conduise majoritairement à l'identification de kératines et «masque» les protéines minoritaires.

Les approches méthodologiques mises en place au cours de mon travail de thèse seront utilisées dans la suite de ce projet d'étude du parasite *T. gondii*, et plus particulièrement pour l'étude des partenaires des 19 autres facteurs de transcription en cours d'étude.

L'étude protéomique des protéines partenaires clés du récepteur TgSORTLR va dès à présent se poursuivre par une étude quantitative de détermination des stœchiométries des complexes protéiques (Vps9, 26, 35, Clathrin μ 1, Sec23 et Hp 20) par une approche LC-SRM (LC-MS, MS en mode Selected-Reaction Monitoring). Le but de cette expérience est de construire un réseau d'interactions autour de TgSORTLR et des différentes protéines partenaires.

Références bibliographiques du chapitre 1

1. Hovasse, A. Protéomique: maîtrise de l'instrumentation, application et étude de glycosylations. *Thèse de l'Université de Strasbourg*, Strasbourg, 2010.
2. Fauquenoy, S.; Morelle, W.; Hovasse, A.; Bednarczyk, A.; Slomianny, C.; Schaeffer, C.; Van Dorsselaer, A.; Tomavo, S. Proteomics and glycomics analyses of N-glycosylated structures involved in *Toxoplasma gondii*-host cell interactions. *Mol Cell Proteomics* **2008**, *7*, (5), 891-910.
3. Olguin-Lamas, A.; Madec, E.; Hovasse, A.; Werkmeister, E.; Callebaut, I.; Slomianny, C.; Delhay, S.; Mouveaux, T.; Schaeffer-Reiss, C.; Van Dorsselaer, A.; Tomavo, S. A novel *Toxoplasma gondii* nuclear factor TgNF3 is a dynamic chromatin-associated component, modulator of nucleolar architecture and parasite virulence. *PLoS Pathog* **2011**, *7*, (3), e1001328.
4. Dubey, J. P. The history of *Toxoplasma gondii*--the first 100 years. *The Journal of eukaryotic microbiology* **2008**, *55*, (6), 467-75.
5. Innes, E. A. A brief history and overview of *Toxoplasma gondii*. *Zoonoses and public health* **2010**, *57*, (1), 1-7.
6. Tenter, A. M.; Heckeroth, A. R.; Weiss, L. M. *Toxoplasma gondii*: from animals to humans. *International journal for parasitology* **2000**, *30*, (12-13), 1217-58.
7. Jones, J. L.; Kruszon-Moran, D.; Wilson, M.; McQuillan, G.; Navin, T.; McAuley, J. B. *Toxoplasma gondii* infection in the United States: seroprevalence and risk factors. *American journal of epidemiology* **2001**, *154*, (4), 357-65.
8. AFSSA. Toxoplasmose : état des connaissances et évaluation du risque lié à l'alimentation. *Rapport de l'Agence Française de Sécurité Sanitaire des Aliments* **2005**.
9. Naqi, R.; Azeemuddin, M.; Ahsan, H. Cerebral toxoplasmosis in a patient with acquired immunodeficiency syndrome. *The Journal of the Pakistan Medical Association* **2010**, *60*, (4), 316-8.
10. Dubremetz, J. F.; Lebrun, M. Virulence factors of *Toxoplasma gondii*. *Microbes and infection / Institut Pasteur* **2012**.
11. Walsh, C. P.; Hammond, S. E.; Zajac, A. M.; Lindsay, D. S. Survival of *Toxoplasma gondii* tachyzoites in goat milk: potential source of human toxoplasmosis. *J Eukaryot Microbiol* **1999**, *46*, (5), 73S-74S.
12. Skinner, L. J.; Timperley, A. C.; Wightman, D.; Chatterton, J. M.; Ho-Yen, D. O. Simultaneous diagnosis of toxoplasmosis in goats and goatowner's family. *Scandinavian journal of infectious diseases* **1990**, *22*, (3), 359-61.
13. Dunn, D.; Wallon, M.; Peyron, F.; Petersen, E.; Peckham, C.; Gilbert, R. Mother-to-child transmission of toxoplasmosis: risk estimates for clinical counselling. *Lancet* **1999**, *353*, (9167), 1829-33.
14. Mandelbrot, L. Prevention of mother-to-child transmission of toxoplasmosis: Perspectives. *Gynecologie, obstetrique & fertilité* **2012**, *40*, (10), 591-8.
15. Rogers, N. M.; Peh, C. A.; Faull, R.; Pannell, M.; Cooper, J.; Russ, G. R. Transmission of toxoplasmosis in two renal allograft recipients receiving an organ from the same donor. *Transplant infectious disease : an official journal of the Transplantation Society* **2008**, *10*, (1), 71-4.
16. Giordano, L. F.; Lasmar, E. P.; Tavora, E. R.; Lasmar, M. F. Toxoplasmosis transmitted via kidney allograft: case report and review. *Transplantation proceedings* **2002**, *34*, (2), 498-9.
17. Nelson, J. C.; Kauffmann, D. J.; Ciavarella, D.; Senisi, W. J. Acquired toxoplasmic retinochoroiditis after platelet transfusions. *Annals of ophthalmology* **1989**, *21*, (7), 253-4.
18. Saadatnia, G.; Golkar, M. A review on human toxoplasmosis. *Scandinavian journal of infectious diseases* **2012**.

19. Kim, J. H.; Kang, K. I.; Kang, W. C.; Sohn, H. J.; Jean, Y. H.; Park, B. K.; Kim, Y.; Kim, D. Y. Porcine abortion outbreak associated with *Toxoplasma gondii* in Jeju Island, Korea. *Journal of veterinary science* **2009**, 10, (2), 147-51.
20. Dubey, J. P.; Lindsay, D. S.; Speer, C. A. Structures of *Toxoplasma gondii* tachyzoites, bradyzoites, and sporozoites and biology and development of tissue cysts. *Clin Microbiol Rev* **1998**, 11, (2), 267-99.
21. Robert-Gangneux, F.; Darde, M. L. Epidemiology of and diagnostic strategies for toxoplasmosis. *Clinical microbiology reviews* **2012**, 25, (2), 264-96.
22. Ajioka, J. W.; Fitzpatrick, J. M.; Reitter, C. P. *Toxoplasma gondii* genomics: shedding light on pathogenesis and chemotherapy. *Expert reviews in molecular medicine* **2004**, 3, (01).
23. Carruthers, V. B.; Giddings, O. K.; Sibley, L. D. Secretion of micronemal proteins is associated with *Toxoplasma* invasion of host cells. *Cellular microbiology* **1999**, 1, (3), 225-35.
24. Carruthers, V. B. Host cell invasion by the opportunistic pathogen *Toxoplasma gondii*. *Acta tropica* **2002**, 81, (2), 111-22.
25. Bradley, P. J.; Sibley, L. D. Rhoptries: an arsenal of secreted virulence factors. *Current opinion in microbiology* **2007**, 10, (6), 582-7.
26. Keeley, A.; Soldati, D. The glideosome: a molecular machine powering motility and host-cell invasion by Apicomplexa. *Trends in cell biology* **2004**, 14, (10), 528-32.
27. Tomavo, S. The differential expression of multiple isoenzyme forms during stage conversion of *Toxoplasma gondii*: an adaptive developmental strategy. *International journal for parasitology* **2001**, 31, (10), 1023-31.
28. Dzierszinski, F.; Mortuaire, M.; Dendouga, N.; Popescu, O.; Tomavo, S. 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* **2001**, 309, (5), 1017-27.
29. Radke, J. R.; Behnke, M. S.; Mackey, A. J.; Radke, J. B.; Roos, D. S.; White, M. W. The transcriptome of *Toxoplasma gondii*. *BMC biology* **2005**, 3, 26.
30. Amin, A.; Mazloomzadeh, S.; Haniloo, A.; Mohammadian, F.; Fazaeli, A. Evaluation of Anti-*Toxoplasma* IgG, IgM, and IgA in Mothers with Spontaneous Abortion in Zanjan, Northwest Iran. *The Korean journal of parasitology* **2012**, 50, (4), 371-4.
31. Selseleh, M.; Modarressi, M. H.; Mohebal, M.; Shojaee, S.; Eshragian, M. R.; Azizi, E.; Keshavarz, H. Real-time RT-PCR on SAG1 and BAG1 gene expression during stage conversion in immunosuppressed mice infected with *Toxoplasma gondii* Tehran strain. *The Korean journal of parasitology* **2012**, 50, (3), 199-205.
32. Buxton, D.; Innes, E. A. A commercial vaccine for ovine toxoplasmosis. *Parasitology* **1995**, 110 Suppl, S11-6.
33. de Carvalho, L.; Souto-Padron, T.; de Souza, W. Localization of lectin-binding sites and sugar-binding proteins in tachyzoites of *Toxoplasma gondii*. *The Journal of parasitology* **1991**, 77, (1), 156-61.
34. Odenthal-Schnittler, M.; Tomavo, S.; Becker, D.; Dubremetz, J. F.; Schwarz, R. T. Evidence for N-linked glycosylation in *Toxoplasma gondii*. *Biochem J* **1993**, 291 (Pt 3), 713-21.
35. Doliwa, C. Caractérisation par protéomique des mécanismes de résistance à la sulfadiazine chez *Toxoplasma gondii*. *Thèse de l'Université de Reims Champagne-Ardenne*, Reims, 2012.
36. Gissot, M.; Kim, K.; Schaap, D.; Ajioka, J. W. New eukaryotic systematics: a phylogenetic perspective of developmental gene expression in the Apicomplexa. *International journal for parasitology* **2009**, 39, (2), 145-51.
37. Banks, C. A.; Kong, S. E.; Washburn, M. P. Affinity purification of protein complexes for analysis by multidimensional protein identification technology. *Protein expression and purification* **2012**, 86, (2), 105-119.
38. Isono, E.; Schwechheimer, C., Co-immunoprecipitation and protein blots. *Methods in molecular biology* **2010**, 655, 377-87.

39. Aravind, L.; Iyer, L. M.; Anantharaman, V. The two faces of Alba: the evolutionary connection between proteins participating in chromatin structure and RNA metabolism. *Genome biology* **2003**, 4, (10), R64.
40. Wardleworth, B. N.; Russell, R. J.; Bell, S. D.; Taylor, G. L.; White, M. F. Structure of Alba: an archaeal chromatin protein modulated by acetylation. *The EMBO journal* **2002**, 21, (17), 4654-62.
41. Dube, J. B.; Johansen, C. T.; Hegele, R. A. Sortilin: an unusual suspect in cholesterol metabolism: from GWAS identification to in vivo biochemical analyses, sortilin has been identified as a novel mediator of human lipoprotein metabolism. *BioEssays : news and reviews in molecular, cellular and developmental biology* **2011**, 33, (6), 430-7.
42. Zhao, Y.; Cui, J. G.; Lukiw, W. J. Reduction of sortilin-1 in Alzheimer hippocampus and in cytokine-stressed human brain cells. *Neuroreport* **2007**, 18, (11), 1187-91.
43. Fauquenoy, S.; Hovasse, A.; Sloves, P. J.; Morelle, W.; Dilezitoko Alayi, T.; Slomianny, C.; Werkmeister, E.; Schaeffer, C.; Van Dorsselaer, A.; Tomavo, S. Unusual N-glycan structures required for trafficking *Toxoplasma gondii* GAP50 to the inner membrane complex regulate host cell entry through parasite motility. *Mol Cell Proteomics* **2011**, 10, (9), M111 008953.
44. Elias, J. E.; Haas, W.; Faherty, B. K.; Gygi, S. P. Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. *Nature methods* **2005**, 2, (9), 667-75.
45. Balgley, B. M.; Laudeman, T.; Yang, L.; Song, T.; Lee, C. S. Comparative evaluation of tandem MS search algorithms using a target-decoy search strategy. *Molecular & cellular proteomics* **2007**, 6, (9), 1599-608.
46. Fauquenoy, S. Implication de la N-glycosylation dans les mecanismes de motilite et d'invasion cellules hotes chez *Toxoplasma gondii*. *Thèse de l'Universite de Lille, Lille*, **2010**.
47. Braun, L.; Cannella, D.; Ortet, P.; Barakat, M.; Sautel, C. F.; Kieffer, S.; Garin, J.; Bastien, O.; Voinnet, O.; Hakimi, M. A. A complex small RNA repertoire is generated by a plant/fungal-like machinery and effected by a metazoan-like Argonaute in the single-cell human parasite *Toxoplasma gondii*. *PLoS pathogens* **2010**, 6, (5), e1000920.
48. Bell, S. D.; Botting, C. H.; Wardleworth, B. N.; Jackson, S. P.; White, M. F. The interaction of Alba, a conserved archaeal chromatin protein, with Sir2 and its regulation by acetylation. *Science* **2002**, 296, (5565), 148-51.
49. Jelinska, C.; Conroy, M. J.; Craven, C. J.; Hounslow, A. M.; Bullough, P. A.; Waltho, J. P.; Taylor, G. L.; White, M. F. Obligate heterodimerization of the archaeal Alba2 protein with Alba1 provides a mechanism for control of DNA packaging. *Structure* **2005**, 13, (7), 963-71.
50. Mani, J.; Guttinger, A.; Schimanski, B.; Heller, M.; Acosta-Serrano, A.; Pescher, P.; Spath, G.; Roditi, I. Alba-domain proteins of *Trypanosoma brucei* are cytoplasmic RNA-binding proteins that interact with the translation machinery. *PLoS One* **2011**, 6, (7), e22463.
51. Subota, I.; Rotureau, B.; Blisnick, T.; Ngwabyt, S.; Durand-Dubief, M.; Engstler, M.; Bastin, P. ALBA proteins are stage regulated during trypanosome development in the tsetse fly and participate in differentiation. *Molecular biology of the cell* **2011**, 22, (22), 4205-19.
52. Haseltine, C. A.; Kowalczykowski, S. C., A distinctive single-strand DNA-binding protein from the Archaeon *Sulfolobus solfataricus*. *Molecular microbiology* **2002**, 43, (6), 1505-15.
53. Jeffers, V.; Sullivan, W. J., Jr. Lysine acetylation is widespread on proteins of diverse function and localization in the protozoan parasite *Toxoplasma gondii*. *Eukaryotic cell* **2012**, 11, (6), 735-42.
54. Sha, Z.; Brill, L. M.; Cabrera, R.; Kleifeld, O.; Scheliga, J. S.; Glickman, M. H.; Chang, E. C.; Wolf, D. A. The eIF3 interactome reveals the translasome, a supercomplex linking protein synthesis and degradation machineries. *Molecular cell* **2009**, 36, (1), 141-52.
55. Zhou, C.; Arslan, F.; Wee, S.; Krishnan, S.; Ivanov, A. R.; Oliva, A.; Leatherwood, J.; Wolf, D. A. PCI proteins eIF3e and eIF3m define distinct translation initiation factor 3 complexes. *BMC biology* **2005**, 3, 14.

56. Sloves, P. J.; Delhay, S.; Mouveaux, T.; Werkmeister, E.; Slomianny, C.; Hovasse, A.; Dilezitoko Alayi, T.; Callebaut, I.; Gaji, R. Y.; Schaeffer-Reiss, C.; Van Dorsselaar, A.; Carruthers, V. B.; Tomavo, S. *Toxoplasma Sortilin-like Receptor Regulates Protein Transport and Is Essential for Apical Secretory Organelle Biogenesis and Host Infection. Cell Host Microbe* **2012**, 11, (5), 515-27.
57. Luk, F. C.; Johnson, T. M.; Beckers, C. J. N-linked glycosylation of proteins in the protozoan parasite *Toxoplasma gondii*. *Molecular and biochemical parasitology* **2008**, 157, (2), 169-78.
58. Kim, E.; Lee, Y.; Lee, H. J.; Kim, J. S.; Song, B. S.; Huh, J. W.; Lee, S. R.; Kim, S. U.; Kim, S. H.; Hong, Y.; Shim, I.; Chang, K. T. Implication of mouse Vps26b-Vps29-Vps35 retromer complex in sortilin trafficking. *Biochemical and biophysical research communications* **2010**, 403, (2), 167-71.

Chapitre 2 :

Etudes protéomique et génomique des changements induits par la maladie de l'esca sur le *Vitis vinifera L.cv. Chardonnay*

Cette étude a été réalisée en collaboration avec le Dr. F. Fontaine de l'Université de Reims Champagne-Ardenne et s'inscrit dans le cadre d'un programme d'étude des maladies de dépérissement de la vigne (Esca et DBA (Dead Black Arm)).

Ce travail de thèse a consisté à mettre au point des stratégies d'analyse protéomique permettant l'identification des protéines de vigne et de champignons différentiellement exprimées lors de la maladie de l'esca.

Ce travail devrait aider à mieux comprendre la maladie de l'esca afin de développer des moyens de lutte plus efficace contre cette maladie.

I. Contexte de cette étude

1. La maladie de l'esca

A l'origine appelé iska en Grèce ¹, "esca proper" est le nom proposé pour désigner l'une des maladies mondiales du bois très répandue dans les milieux viticoles ². La maladie de l'esca est connue depuis des décennies, pour être très nuisible pour le patrimoine viticole. L'esca s'attaque à la fois aux bois de jeunes et de vieux plants de vignobles ³, et conduit à plus ou moins long terme à la mort des ceps, à la baisse de la production du raisin et de la qualité du vin ⁴. En France, selon les estimations de l'Observatoire national mis en place pour suivre l'évolution des maladies du bois, la perte économique liée à l'esca et les autres pathogènes de la vigne est d'environ 11% en moyenne par année. Mais cette moyenne masque des disparités importantes avec des dégâts pouvant dépasser largement ces 11 % de pertes (www.viti-net.com).

L'esca est une maladie complexe causée par divers champignons et qui peut être aggravée par des facteurs climatiques ². Deux symptômes de dépérissement de vignes sont connus : la forme aiguë de la maladie, qui correspond à l'esca apoplectique, et la forme lente de l'esca (Fle) qui se manifeste par la tigrure et jaunissement des feuilles.

On compte aujourd'hui plus d'une dizaine d'agents pathogènes responsables de l'esca et qui sont exclusivement des champignons. Les 3 agents pathogènes de l'esca les plus cités comme responsables de l'esca sont: *Fomitiporia mediterranea* qui entraîne la pourriture du bois, *Phaeoacremonium aleophilum* et *Phaeomoniella chlamydospora* qui conduisent à la tigrure et jaunissement des feuilles de vignes (Fle) ^{5,6}.

Les symptômes de la forme lente se produisent aussi bien dans les tissus ligneux du tronc et des branches ainsi que sur les feuilles et les baies. Les dégâts de la forme lente de l'esca chez la vigne varient souvent d'une année à l'autre, mais se caractérisent en général par la tigrure des feuilles, la nécrose du bois des tiges et/ou du tronc, et l'apparition de différentes sortes de tâches sur les baies (Figure 1). La vigne atteinte par la forme lente de l'esca survit mais finit par mourir au bout de quelques années.



Figure 1. Symptômes de la forme lente de l'esca: (A) symptôme foliaire sur un cépage blanc, la tigrure de feuilles, (B) coupe transversale d'un tronc présentant au centre et en périphérie des nécroses typiques de l'esca, (C) apparition de tâches violettes sur les baies.

L'apoplexie se manifeste par l'apparition soudaine d'un flétrissement partiel ou complet de la plante infectée et est considérée comme la forme aiguë de la Fle et/ou esca (Figure 2). L'apoplexie apparaît après une longue période d'infection de la vigne par les agents pathogènes et se révèle dès lors que les conditions climatiques (par exemple lors de l'augmentation de température après une forte pluie) deviennent rudes pour la vigne qui va finir par mourir.



Figure 2. Symptôme de la forme aiguë de l'esca: (A) dépérissement foudroyant partiel de la vigne, (B) dépérissement total du pied de vigne.

Différents fongicides ⁷, des nutriments et bio-activateurs ⁸ ou encore des pratiques phytosanitaires ⁹ ont été testés dans le but d'éliminer ou de contrôler les agents pathogènes responsables de la maladie de l'esca. Mais très rapidement, ces traitements se sont révélés inefficaces pour diverses raisons (les coûts, le caractère

complexe de la maladie ou encore le durcissement des normes réglementant l'usage des fongicides). Par exemple, l'arsénite de sodium, un puissant fongicide prometteur contre l'esca a été interdit en France depuis 2001 à cause de sa toxicité et sa nuisance sur la santé des agriculteurs. La complexité de cette maladie du bois, la méconnaissance des champignons responsables de cette maladie et de leurs mécanismes d'action pourraient expliquer l'échec constaté jusqu'à ce jour dans le développement des stratégies de lutte contre la maladie de l'esca. Il est donc indispensable d'apporter des connaissances sur la physiologie de la vigne en réponse à une infection par ces agents pathogènes afin de développer des stratégies de lutte plus efficaces contre la maladie de l'esca.

Des étapes importantes dans la compréhension des mécanismes de réponse de la vigne vis-à-vis de l'esca ont été entamées grâce au développement des techniques d'analyse. Par exemple, la fluorescence et la Résonance Magnétique Nucléaire ont largement contribué dans l'étude des toxines et la physiologie de la vigne infectée. Des aspects physiologiques tels que l'état de l'altération de la photosynthèse¹⁰, l'augmentation de phytotoxines d'origine fongiques dans les feuilles de vignes lors d'une infection de la vigne par l'esca^{11, 12} ont pu être montrés.

Cependant, si quelques études de métabolisme ont été faites, peu de données moléculaires au niveau protéique de la plante et des champignons sont disponibles dans la littérature. En effet, des études de protéomique comparative entre des vignes saines et infectées n'ont pas encore été réalisées à notre connaissance. Ces études comparatives devraient permettre d'apporter des éléments de réponse moléculaire pouvant expliquer les processus biologiques activés ou modifiés lors de l'infection de la vigne par l'esca.

2. Problématique et objectifs de l'étude

Les champignons phytopathogènes sont les organismes parasites les plus dommageables pour les plantes, car ils sont responsables de maladies graves et de pertes importantes dans les milieux agricoles. L'étude biologique de quelques champignons et des interactions avec leur hôte s'est considérablement développée ces dernières années grâce à :

- (1) des progrès réalisés dans le développement des techniques dites –omiques (génomique, protéomique, glycomique etc.) et de la bio-informatique ;
- (2) le nombre croissant de projets de séquençage des génomes.

Parmi toutes ces techniques –omiques, l'analyse protéomique par spectrométrie de masse est l'outil couramment utilisé pour l'étude des interactions plantes-champignons parasites¹³.

Nous avons réalisé dans ce chapitre de thèse, une étude protéomique comparative entre des plantes de vignes saines et celles infectées par l'esca, afin d'identifier les protéines sous- ou sur-exprimées dans les différents organes de la plante (tiges herbacées et troncs).

La stratégie classique d'analyse protéomique par spectrométrie de masse, consiste à comparer des données expérimentales de MS/MS des peptides, générés par protéolyse des extraits protéiques avec les données théoriques de MS/MS de peptides générés *in silico* à partir des banques protéiques. Mais dans certains cas, les séquences protéiques de certains organismes ne sont pas disponibles pour différentes raisons (génomes non séquencés ou partiellement séquencés).

Dans le contexte de cette étude, nous sommes confrontés à une problématique qui met en jeu des organismes différents qui sont :

- (1) l'hôte, qui est la vigne majoritairement représentée par *Vitis vinifera*, dont les génomes de deux espèces de Pinot noir les plus cultivées et économiquement importantes ont été séquencés et annotés depuis 2007 ;
- (2) les parasites, qui sont des champignons responsables de la maladie de l'esca dont les séquençages de génome sont inachevés ou totalement inexistantes.

En effet, malgré le développement des techniques de séquençage de génomes à haut débit, très peu de données protéiques sont disponibles sur les champignons impliqués dans la maladie de l'esca.

A titre d'exemple, nous avons recensé le nombre de protéines de 21 espèces de champignons, décrits comme agents pathogènes responsables des maladies du bois de vigne, qui sont disponibles dans les banques protéiques de NCBI (Tableau 1), dont les 3 principaux agents pathogène de l'esca (*Fomitiporia mediterranea*, *Phaeoacremonium aleophilum* et *Phaeoconiella chlamydospora*). Parmi les 3 champignons, seul le génome de *Fomitiporia mediterranea* est en cours de séquençage (avec déjà 11347 séquences protéiques disponibles) par DOE Joint Genome Institute (USA) (www.genomesonline.org). Les protéines des autres espèces de

champignons proviennent d'études isolées qui ont conduit uniquement au séquençage spécifique des quelques gènes.

Tableau 1. Liste des 21 espèces fongiques impliquées dans les maladies de l'esca chez la vigne et d'autres pathologies de bois des plantes (www.mycobank.com), à chaque espèce est associée le numéro de taxonomie et le nombre de séquences protéiques extraites de NCBI (28 décembre 2012).

Champignons	N° de Taxonomie	Séquences protéiques	Hôtes	Maladies
<i>Phaemoniella chlamydospora</i>	158046	18	Vigne	Esca
<i>Phaeacemonium aleophilum</i>	65417	205	Vigne, cerisier	Esca, nécrose de bois
<i>Fomitiporia mediterranea</i>	208960	11347	Vigne	Esca
<i>Botryosphaeria dothidea</i>	55169	213	Divers plantes	Chancre
<i>Diplodia seriata</i>	420778	101	Vigne, pommier, poirier	Esca, dépérissement
<i>Diplodia corticola</i>	236234	31	Divers plantes, spécialement sur les Chênes	Dépérissement et chancre
<i>Diplodia mutila</i>	85927	46	Diverses plantes	Dépérissement et chancre
<i>Phaeobotryosphaeria porosa</i>	240344	2	Vigne	Esca
<i>Lasioidiplodia theabromae</i>	45133	145	Coton, vigne, bananier	Pourriture de récoltes
<i>Lasioidiplodia crassispora</i>	336254	10	Santal blanc	Chancre
<i>Lasioidiplodia plurivora</i>	448127	2	Prunier du Japon	Chancre
<i>Dothiorella viticola</i>	323561	31	Vigne	Maladie des feuilles, chancres
<i>Dothiorella iberica</i>	286282	47	Vigne, avocatier, diverses plantes subtropicales	Chancre
<i>Dothiorella sarmentorum</i>	240345	25	Vigne, agrumes	Chancre, dépérissement
<i>Neofusicoccum parvum</i>	310453	311	Vigne, arbres fruitiers	Esca, chancre
<i>Neofusicoccum ribis</i>	45134	56	Vigne	Pourriture de fruits et des branches, chancre
<i>Neofusicoccum australe</i>	240362	78	Vigne, acacia, olivier	Dépérissement et chancre
<i>Neofusicoccum luteum</i>	120395	31	Vigne, arbres fruitiers	Chancre
<i>Neofusicoccum mediterraneum</i>	467987	234	Vigne, arbres fruitiers	Dépérissement, chancre
<i>Neofusicoccum viticlavatum</i>	240342	2	Vigne	Esca, chancre
<i>Neofusicoccum vitifusiforme</i>	240343	17	Vigne, myrtier	Chancre, nécrose de bois

Le très faible nombre de séquences protéiques disponibles a donc nécessité l'utilisation d'une approche d'identification de protéines différente de l'approche de recherche conventionnelle d'identification de protéines basée sur une banque protéique. Cette stratégie est basée sur le séquençage *de novo* des spectres MS/MS des peptides, sans assistance de banques protéiques, et suivi d'une approche MS-BLAST¹⁴. Le séquençage *de novo* permet d'obtenir des tags de séquences d'acides aminés, et la recherche par MS-BLAST est une comparaison de ces tags (peptides) issus du séquençage *de novo* avec des protéines homologues d'autres espèces de plants de vigne et de champignons.

Dans une première partie, notre étude a porté sur **les tiges herbacées** de la vigne. Ces organes connexes situés entre les feuilles où aucun champignon n'a été isolé, jouent un grand rôle dans le transport de nutriments des racines vers les feuilles et inversement.

Dans une seconde partie, nous avons étudié **les troncs de vignes**, qui sont les organes de la plante où siègent les champignons responsables de l'esca.

Nous avons utilisé les deux stratégies d'identification de protéines adaptées à chaque contexte d'étude, afin de permettre l'identification des protéines de vigne et/ou des protéines de champignons homologues à ceux étudiés. Cette étude sur la variété de vigne Chardonnay a ensuite été étendue à d'autres variétés telles que le Mourvèdre et le Gewurztraminer.

Cette étude devrait aider à mieux comprendre la réponse de la vigne lors de l'expression des symptômes visibles de la maladie de l'esca. Une meilleure connaissance des mécanismes d'infection de ces pathogènes, et des réponses de la plante, est nécessaire, et permettra de lutter plus efficacement contre cette maladie de la vigne pour laquelle il n'existe à ce jour aucun remède efficace.

II. Etude des tiges herbacées de vignes infectées par l'esca

Les résultats obtenus au cours de ce travail sur les tiges herbacées ont fait l'objet d'un article publié dans *Journal of Proteome Research*¹⁵ disponible dans l'annexe 4.

1. Description des échantillons

Les échantillons ont été extraits à partir de 3 catégories de tiges herbacées de différents plants de vigne de Chardonnay (Figure 3): (1) les vignes atteintes de l'apoplexie (aA et A, Figure 3), la forme aiguë de l'esca présentant (A) ou non (aA) des symptômes visibles ; (2) les vignes atteintes de la forme moins sévère (aE et E, Figure 3) de l'esca présentant (E) ou non (aE) de symptômes visibles ; et (3) les vignes saines, c'est-à-dire n'ayant jamais manifesté une forme de l'esca servant de contrôle (C).

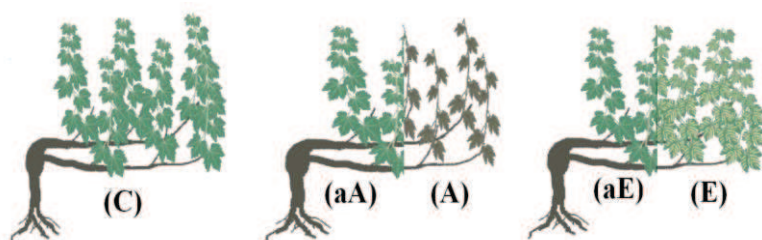


Figure 3. Représentation schématique de l'échantillonnage relatif aux différents symptômes de la maladie de l'esca sur les tiges herbacées de vignes.

2. Stratégie d'analyse protéomique des tiges herbacées de vignes

a. Stratégie globale de l'analyse protéomique

Les extraits protéiques issus des tiges herbacées de vignes sont déposés et séparés sur un gel 2D. Pour chaque extrait protéique, 3 répliquats techniques ont été réalisés et pour chaque symptôme 3 répliquats biologiques, soit un total de 45 gels 2D (en quantité équivalente). L'analyse d'image a permis de détecter 123 spots dont 28 différentiels. Les 28 spots ont été découpés et analysés par nanoLC-MS/MS (Figure 4).

Les analyses nanoLC-MS/MS ont été réalisées en utilisant deux systèmes : un système nanoACQUITY Ultra-performance-LC (Waters) couplé à un spectromètre de masse Q-TOF SYNAPT G1 (Waters) et un système nanoLC-Chip/MS (Agilent) couplé à une trappe ionique amaZon (Bruker) pour bénéficier des différents avantages qu'offrent les deux systèmes.

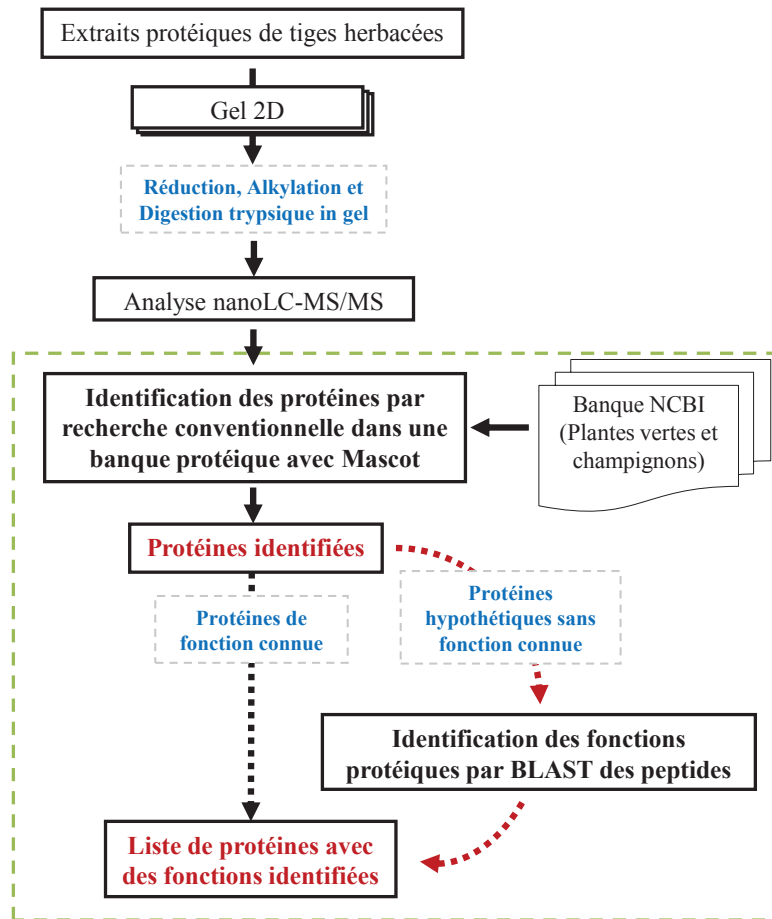


Figure 4. Stratégie globale d'analyse protéomique adoptée pour l'étude des tiges herbacées de vignes infectées par les deux formes de l'esca.

Les protocoles détaillés sont décrits dans la partie expérimentale de la publication des résultats. Nous détaillerons uniquement la partie concernant la stratégie d'identification des protéines (encadrée en vert sur la figure 4) dans les prochains paragraphes.

b. Identification des protéines des tiges herbacées de vignes par recherche conventionnelle dans une banque protéique

L'identification des protéines a été faite sur la base de la stratégie conventionnelle utilisée lors de l'identification des protéines par spectrométrie de masse. La génération des listes de masses, le paramétrage de l'algorithme de recherche Mascot et la validation des résultats de la recherche par Scaffold sont décrits dans la partie expérimentale de l'article publié dans l'annexe 4. Cette approche a conduit à l'identification de nombreuses protéines hypothétiques qui n'ont pas d'annotations fonctionnelles ce qui nous a amené à faire une recherche d'homologie de séquence avec les séquences protéiques d'autres espèces de plante mieux annotées.

c. Identification des fonctions des protéines hypothétiques par homologie de séquence BLAST

Afin d'attribuer des fonctions biologiques à certaines protéines de vignes identifiées comme "hypothetical protein" (Tableau 2.), c'est-à-dire des protéines dont aucune fonction biologique n'est encore décrite, une étape supplémentaire d'**Identification des fonctions protéiques par BLAST des peptides** a été effectuée (Figure 4). L'identification des fonctions de ces protéines hypothétiques est basée sur le fait que des protéines qui ont les mêmes rôles dans les organismes conservent des régions partielles ou totales d'acides aminés d'une espèce à une autre au cours de l'évolution¹⁶. Les conservations de séquences de ces protéines en font de ces protéines des homologues (orthologues pour des protéines de différentes espèces ou paralogues pour des protéines au sein d'une même espèce¹⁷).

Notre stratégie de recherche a consisté à rechercher chez les autres plantes vertes, les protéines homologues ayant des peptides identiques ou fortement semblables à ceux identifiés pour les protéines hypothétiques.

Validation des résultats BLAST

Pour la validation des résultats, seule la fonction de la protéine homologue identifiée avec le plus de peptides avec de bons scores d'homologie de séquence peptidique minimum de 90% sont validés, comme l'illustre la Figure 5.

Hypo	ME TFLFTSESVNEGHPDKLCDQISDAVLDACLQQDPDSK VACETCTKTNM
S-ad	ME TFLFTSESVNEGHPDKLCDQVSDAVLDACLAQDPDSK VACETCTKTNM
Hypo	VMV FGEITTK ANVDYEKIVRDTCR EIGFVSDDVGLDADNCK VLVNIEQQS
S-ad	VMV FGEITTK ANVDYEKIVRDTCR SI G FVSDDVGLDADNCK VLVNIEQQS
Hypo	PDIAQGVHGH LTKRPEEIGAGDQGHMFGYATDETP ELMPLSHVLATK LGA
S-ad	PDIAQGVHGH F TKRPEEIGAGDQGHMFGYATDETP E LMPLSHVLATK LGA
Hypo	RLTEVRKNGT CPWLRPDGKTQVTVEYHNDGGARVPIRVHTVLISTQHDET
S-ad	RLTDVRKNGT CPWLRPDGKTQVTVEY YND NGAMVPIRVHTVLISTQQYET
Hypo	VTNDEIAADLKEHV IKPVIPEKYLDEKTI FHLNPSGR FVIGGPHGDAGLT
S-ad	VTNDEIAADLKEHV IKPVIPEKYLDEKTI FHLNPSGR FVIGGPHGDAGLT
Hypo	GRKIIIDTYGGWGAHGGGAFSGK DPTKVDRSGAYIVRQAAKSIVANGLAR
S-ad	GRKIIIDTYGGWGAHGGGAFSGK DPTKVDRSGAYIVRQAAKSIVANGLAR
Hypo	RCIVQVSYAIGVPEPLSVFVDSYGTGK IPDREILKIVKENFDFRPGMIS I
S-ad	RCIVQVSYAIGVPEPLSVFVDSYGTGK IPDKEILQIVKENFDFRPGM IT I
Hypo	NL DLK RGGNGRFLK TAAYGHFGRDDP DFT WEVVK PLKWEKTQA
S-ad	NL DLK RGGNSRFLK TAAYGHFGRDDP DFT WE AVKPLKWDK PQS

Figure 5. Exemple de validation d'homologie de séquence BLAST de (Hypo), l'hypothetical protein VITISV_002362 [*Vitis vinifera*] (gi|147791974) et son homologue (S-ad), S-adenosylmethionine synthetase [*Gossypium hirsutum*] (gi|307948774) avec indication des 10 peptides (en rouge) identifiés par nanoLC-MS/MS. En couleur rouge, les peptides identifiés pour la l'hypothetical protein, en surligné jaune, les tags de séquences peptidiques communs identiques entre les deux protéines, en surligné gris, les acides aminés différents entre les deux protéines.

Le répertoire final des protéines identifiées comprend : (1) les protéines identifiées par la recherche conventionnelle dans la banque protéique ; (2) les fonctions recouvrées pour les protéines hypothétiques par l'approche BLAST avec les autres plantes vertes.

Toutes les protéines identifiées ont été déposées par la suite sur le serveur PRIDE (www.ebi.ac.uk/pride)¹⁸, sous les numéros d'accèsion # 19645 pour les données provenant du Q-TOF SYNAPT G1 (Waters) et # 19646 pour les données provenant de la trappe ionique amaZon (Bruker).

3. Résultats publiés

Grâce à notre stratégie d'analyse protéomique mise en place, 33 protéines ont été identifiées dans 28 spots différentiels analysés par nanoLC-MS/MS. Dans la plupart des cas (22 cas sur 33), la recherche classique dans la banque protéique de plantes vertes et champignons a conduit à l'identification des protéines de fonction biologique non décrite. Ce qui a nécessité une étape supplémentaire de recherche par approche BLAST pour identifier les fonctions de ces protéines par homologie de séquence avec les autres protéines de plantes vertes (Tableau 2).

Tableau 2. ^(a) Liste des 33 protéines identifiées par Mascot. * *Vitis vinifera*, ** *Glycine max*, *** *Gossypium hirsutum*,*** *Cicer arietinum*, ^(b) numéro d'accèsion correspondant à la protéine identifiée par Mascot dans NCBI, ^(c) identification de fonction protéique par homologie de séquence BLAST.

Protéines identifiées par Mascot ^(a)	N° d'accèsion NCBI ^(b)	Résultats BLAST ^(c) Fonctions identifiées
hypothetical protein [*]	gi 147811887	polyphenol oxidase
S-adenosylmethionine synthase 5 [*]	gi 223635289	
catechol oxidase precursor [*]	gi 1076272	
Proteasome subunit alpha type-5 [***]	gi 12229923	
hypothetical protein [*]	gi 225455238	small heat shock protein
hypothetical protein [*]	gi 225426793	thaumatin-like
hypothetical portein [*]	gi 147799889	PDI-like protein
unnamed protein [*]	gi 296084626	heat shock protein
hypothetical protein [*]	gi 225446775	oxygen-evolving enhancer protein 2
hypothetical protein [*]	gi 225449595	actin depolymerizing factor 1
hypothetical protein isoform 2 [*]	gi 225451120	superoxide dismutase
glutathione S-transferase 5[*]	gi 158323772	
hypothetical protein [*]	gi 225439868	diaminopimelate epimerase
hypothetical protein [*]	gi 147843260	60S acidic ribosomal protein P0
hypothetical protein [*]	gi 225469270	myo inositol monophosphatase
similar to HSC70-1 [*]	gi 225449497	
ATP synthase subunit alpha [***]	gi 110832063	
GroEL-like chaperone, ATPase [*]	gi 225433375	
Hypothetical protein [*]	gi 147798422	heat-shock protein
unnamed protein [*]	gi 297741493	cytosolic ascorbate peroxidase
hypothetical protein [*]	gi 225460600	Peroxidase
S-adenosylmethionine synthase 4[*]	gi 223635287	
hypothetical protein [*]	gi 225442981	phospholipase D alpha
hypothetical protein [*]	gi 225429592	small heat shock protein
hypothetical protein [*]	gi 225429614	small heat shock protein
predicted: O-acetylserine(thiol)-lyase [*]	gi 225434133	
hypothetical protein [*]	gi 147853311	aldo/keto reductase
hypothetical protein [*]	gi 147791974	s-adenosylmethionine synthetase
Hypothetical protein [*]	gi 225428898	phosphoglycerate dehydrogenase
similar to pyruvate decarboxylase [*]	gi 225443847	
hypothetical protein [*]	gi 225430776	succinate dehydrogenase, putative
cytosolic malate dehydrogenase [***]	gi 10334493	
hypothetical protein [*]	gi 147790725	aldo/keto reductase

Les annotations fonctionnelles ont permis de classer ces 33 protéines en 3 groupes (Figure 6) : (i) protéines impliquées dans le métabolisme et l'énergie (ME) ; (ii) protéines impliquées dans la tolérance aux stress (TS) ; (iii) protéines imputables à la réponse de défense de la vigne (RD).

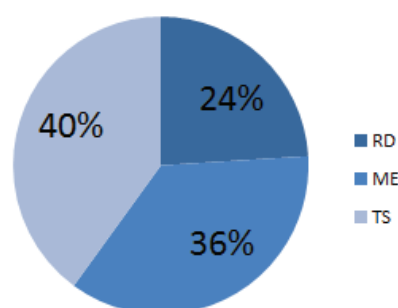


Figure 6. Pourcentages de protéines identifiées et réparties entre les trois groupes. (ME) Métabolisme et Energie, (RD) Réponse de Défense, (TS) Tolérance au Stress.

Des 33 protéines identifiées par analyse protéomique, 9 ont été sélectionnées pour la quantification de l'expression des gènes par réaction en chaîne par polymérase "Polymerase Chain Reaction" (PCR). Les résultats de la PCR (cf. Annexe 4) ont révélés dans certains cas des discordances entre les quantités de gènes et les quantités correspondantes des protéines analysées en protéomique.

Conclusion

Pour identifier les protéines impliquées dans la réponse de la vigne lors de l'infection de l'esca, nous avons focalisé nos investigations sur la variation du protéome des tiges herbacées de vignes infectées présentant ou non des symptômes visibles, par rapport aux tiges herbacées de vignes saines.

L'analyse protéomique a permis de donner des candidats protéiques à nos collaborateurs biologistes qui ont en suite quantifié leurs gènes par PCR. Les résultats ont montré que les mêmes réponses protéiques sont activées dans les tiges herbacées asymptomatiques (aA et aE) au même titre que dans les tiges symptomatiques (E et A) de vignes infectées. Cependant, même si les mêmes agents pathogènes sont responsables des deux formes de la maladie de l'esca, une variation significative de l'expression des gènes de protéines apparait dès que les symptômes deviennent visibles¹⁵.

Les feuilles et les tiges herbacées de vignes touchées par la maladie de l'esca ont montré des changements physiologiques et métaboliques liés aux symptômes extérieurs, même si des agents pathogènes associés à la maladie de l'esca ou d'autres maladies du bois n'ont pas été isolés à partir de ces tiges herbacées étudiées^{15, 19}. Aucune différence notable en termes de nombre de protéines identifiées par chaque système LC-MS/MS, n'a pu être observée le peu de protéines disponibles par spots (4-9 protéines par spot dont au moins 3 sont homologues). Nous nous limiterons donc dans la suite de ce travail à l'usage du système nanoLC-Chip/MS.

Annexe 4: Publication

Physiological changes in green stems of *Vitis vinifera* L. cv. Chardonnay in response to esca proper and apoplexy revealed by proteomic and transcriptomic analyses

Spagnolo, A. ; Magnin-Robert, M. ; **Alayi, T. D.** ; Cilindre, C. ; Mercier, L. ; Schaeffer-Reiss, C. ;
Van Dorsselaer, A. ; Clement, C. ; Fontaine, F.

Journal of Proteome Research, 2012, 11:1

Physiological Changes in Green Stems of *Vitis vinifera* L. cv. Chardonnay in Response to Esca Proper and Apoplexy Revealed by Proteomic and Transcriptomic Analyses

Alessandro Spagnolo,^{†,‡} Maryline Magnin-Robert,^{†,‡} Tchilabalo Dilezitoko Alayi,[§] Clara Cilindre,^{‡,||} Laurence Mercier,[†] Christine Schaeffer-Reiss,[§] Alain Van Dorsselaer,[§] Christophe Clément,[‡] and Florence Fontaine^{*,‡}

[†]Université de Reims Champagne-Ardenne, URVVC EA 2069, Laboratoire Stress, Défenses et Reproduction des Plantes, BP 1039, 51687 Reims Cedex 2, France

[§]Université de Strasbourg, IPHC, UMR 7178, Laboratoire de Spectrométrie de Masse Bioorganique, F-67087 Strasbourg, France

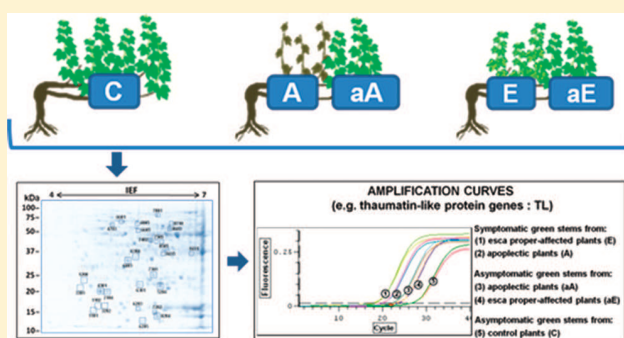
^{||}Université de Reims Champagne-Ardenne, URVVC EA 2069, Laboratoire d'Oenologie et Chimie Appliquée, BP 1039, 51687 Reims Cedex 2, France

[†]Moët & Chandon, 6 rue Croix de Bussy, 51200 Epernay, France

S Supporting Information

ABSTRACT: Among grapevine trunk diseases, esca proper and apoplexy commonly represent a threat for viticulture worldwide. To retrieve further information about the mechanisms activated in apoplectic and esca proper-affected plants, a two-dimensional gel electrophoresis (2-DE) based analysis was conducted on green stems from 26-year-old standing vines. Symptomatic and asymptomatic stems from both apoplectic (A) and esca proper-affected (E) plants compared to control (without visual symptom since 10 years) stems were studied. Thirty-three differentially expressed proteins were identified by nanoLC-MS/MS and included into three groups conceptually defined as proteins involved in (i) metabolism and energy, (ii) stress tolerance, and (iii) defense response. For nine of them, expression of the relative mRNA's was also monitored by qRT-PCR. Proteome variations were specifically related to apoplexy and esca proper but were more similar in asymptomatic stems than in the symptomatic ones. Remarkable quantitative differences were noted for several proteins in symptomatic stems according to the expressed form, A and E. Results further indicate that similar responses are likely activated in asymptomatic stems but a various quantitative expression is triggered upon onset of apoplexy or esca proper symptoms while both kind of plants are infected by the same pathogenic fungi.

KEYWORDS: *Vitis vinifera* L., grapevine trunk diseases, apoplexy, esca proper, green stem, proteomics, PR-5 proteins, qRT-PCR



INTRODUCTION

Esca proper is the term proposed to indicate the coexistence of two different syndromes of the grapevine trunk dieback, named esca and grapevine leaf stripe disease (GLSD), on a same plant.¹ Causal agents of esca are different wood rot fungi represented mainly by *Fomitiporia mediterranea* (Fom) M. Fischer in Europe and the Mediterranean basin.^{1,2} GLSD is a tracheomycotic widespread syndrome; its major causal agents are considered to be *Phaeoconiella chlamydospora* (W. Gams, Crous, M.J. Wingfield and L. Mugnai) P.W. Crous and W. Gams (Pch) and *Phaeoacremonium aleophilum* W. Gams, P.W. Crous, M.J. Wingfield and L. Mugnai (Pal).² Symptoms of GLSD occur both in the woody tissues of trunk and branches, as well as in leaves and berries.² Apoplexy, consisting in the partial or complete sudden wilting of the crown, is regarded as the acute form of GLSD and/or esca

proper.^{2,3} These grapevine trunk diseases, the incidence of which increased in the last 10–15 years,^{4,5} commonly represent a threat for viticulture worldwide.^{6–9}

Several studies have been conducted with the attempt to control GLSD and other diseases of the grapevine trunk by fighting its causal agents or trying to limit its foliar symptoms¹⁰ but no appropriate control is available to date.² The complexity of the wood diseases could explain until now the failure in the development of effective control strategies. As a matter of fact, characterizing the impact of these diseases on grapevine physiology represents a key step toward a better knowledge of symptom

Special Issue: Microbial and Plant Proteomics

Received: September 5, 2011

Published: November 04, 2011

development and consequent possible management strategies. Important steps in understanding response mechanisms of grapevine to GLSD and/or esca have been retrieved using different analytical techniques. Physiological parameters such as gas exchange and especially chlorophyll *a* fluorescence, a reliable and sensitive marker of plant early stress response, have been investigated by easy-to-use and nondestructive methods.^{11–13} Petit et al.¹⁴ studied in detail the perturbation of GLSD on photosynthesis and related decreasing of carbohydrate storage in the woody tissues. In the meantime, molecular analyses using the qRT-PCR technique focusing on the expression of target genes known to be up- or down-regulated following fungal infection have been performed.¹⁵ Using that technique, Letousey et al.³ observed the repression of photosynthesis genes expression and the induction of defense genes in the leaves of the apoplectic plants. Valtaud et al.¹⁶ observed the systemic effect on leaf glutathione metabolism and defense genes expression through qRT-PCR, enzymatic and immunodetection techniques. However, a drawback common to these analytical methods is their target approach.

Physiological and metabolic changes in *Vitis* plants affected by trunk diseases were reported on wood¹⁷ and leaves.⁹ Leaves express visual symptoms and can be easily collected without altering the plant durability and berry production. As far as we know, there is no report on the physiological changes that might occur in the green stem of esca proper-affected plants. However, this organ represents the connection between wood tissues and leaves, and important findings might also be retrieved from that organ. Indeed, pathogens associated to esca proper and other trunk diseases of the grapevine are commonly isolated from woody tissues,^{8,18} while no data of their isolation from green stems have been reported. The most accredited hypothesis to explain foliar symptoms appearance considers the translocation of toxic fungal metabolites from woody tissues to the foliage through the xylem flow.^{2,19} The green stem, an organ devoted to the phloem and xylem flows, is directly exposed to the hypothetical compounds produced in woody tissues by fungi that are found in the xylem.

Proteomics is a powerful molecular tool to monitor global changes in protein expression in tissues and organisms and changes in expression occurring under stress.²⁰ Among the proteomic techniques, two-dimensional polyacrylamide gel electrophoresis (2-DE) represents a reliable and still abundantly used analytical tool for profiling identification and quantitative expression data.^{21–24} In the present study, with the attempt to retrieve further information about the response mechanisms of *Vitis vinifera* to esca proper, a 2-DE based analysis on green stems from standing vines was planned. Moreover, differential protein expression between stems from esca proper-affected vines and those of vines affected by apoplexy was also investigated. A number of selected proteins identified by nanoLC–MS/MS proteomic analysis were confirmed by a qRT-PCR analysis.

MATERIALS AND METHODS

Plant Material

Fifteen standing vines (cv. Chardonnay/41B) were uprooted in 2010 from a 26-year-old vineyard located in the province of Epernay (Champagne-Ardenne region, France) owned by the company Moët & Chandon. Five plants per each target external leaf symptom (GLSD and apoplexy) and five asymptomatic plants were collected. Asymptomatic plants were chosen among

those that had never shown either GLSD or apoplexy symptoms since 2001, thus regarded as visually unaffected plants (control plants). Trunk, branches and one-year-old stems of all the collected plants were inspected internally for the presence of discolorations associated to grapevine trunk diseases and subsequently subjected to fungal isolation as described by Spagnolo et al.²⁵ Green stems were also enclosed in this analysis. Typical wood symptoms of GLSD and esca were noted in all plants examined, included control plants. Thus, plants showing foliar symptoms of GLSD were considered as esca proper-affected plants. Five groups of samples were defined for the study: C (stems from control plants), A and E [symptomatic stems from apoplectic (A) and esca proper (E)-affected plants], and aA and aE [asymptomatic stems from apoplectic (aA) and esca proper (aE)-affected plants]. An asymptomatic stem is a stem with visually healthy leaves. Three biological replicates per group (three stems from three different plants) were carried out. The green stems used for protein extraction, as well as RNA extraction were frozen in the field with liquid nitrogen and subsequently stored at -80°C . Before each analysis, the amount of biological sample needed was ground to a fine powder in liquid nitrogen with a Mixer Mill MM 400 (Retsch, Haan, Germany).

Protein Extraction

Total protein fraction of green stems was isolated using a phenol-based procedure according to Wang et al.²⁶ and Röhrig et al.²⁷ with some modifications. The powdered tissue was placed in microtubes (0.30 ± 0.01 g of powder per 2.0 mL microtubes) and then resuspended in 1.0 mL of cold acetone. After vortexing thoroughly for 30 s, the tubes were centrifuged at $10000 \times g$ for 5 min at 4°C . The resultant pellet was washed once more with cold acetone. The pellet was sequentially rinsed with cold 80% acetone 3 times or until the supernatant was colorless, then resuspended in 1.0 mL of cold 20% (w/v) trichloroacetic acid (TCA)/ H_2O . The suspension was sonicated in a water bath at 4°C for 10 min. After centrifugation, the pellets were sequentially washed twice with 20% (w/v) TCA/ H_2O and twice with 80% (v/v) acetone. This pellet was air-dried and the dry powder was resuspended in 0.7 mL dense sodium dodecyl sulfate (SDS) buffer [30% (w/v) sucrose, 2% (w/v) SDS, 0.1 M tris(hydroxymethyl) amino-methane (Tris) -HCl pH 8.0, 5% (v/v) 2-mercaptoethanol]. Then 0.8 mL of a 90% phenol solution (Sigma-Aldrich, St. Louis, MO) was added, and the resulting mixture was vortexed for 30 s. The phenol phase, recovered by centrifugation at $10000 \times g$ for 5 min at 4°C , was separated in two aliquots. One of 0.7 mL was transferred to a 15 mL Falcon tubes while an aliquot of 0.1 mL was placed in a 1.5 mL microtubes. Further steps were followed in parallel. After addition of 5 volumes of cold 0.1 M ammonium acetate in methanol, proteins were precipitated from the phenol phase overnight at -20°C . The precipitated proteins were recovered by centrifugation, washed twice with cold 0.1 M ammonium acetate in methanol and twice with 80% (v/v) acetone. The final pellet was air-dried and stored at -80°C . The pellet retrieved from the aliquot of 0.1 mL was dissolved in 0.1 mL of 8 M urea for protein quantification using the Pierce 660 nm Protein Assay Kit (Thermo Fisher scientific, Waltham, MA) and bovine serum albumin (BSA) as standard. After quantification, protein samples were solubilized in 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 (2-DE)

For preparative 2-DE analysis, samples containing approximately 40 μg of total protein fraction were diluted in a mixture containing sample buffer and 10% (v/v) glycerol to a final volume of 125 μL . Immobilized pH gradient (IPG) gel strips (ReadyStrip IPG, pH 4–7, 7 cm, Bio-Rad, Hercules, CA) were actively rehydrated for 15 h at 20 °C with the mixture. Isoelectric focusing (IEF) was conducted for 5 h at 20 °C in an IPGphor unit (Amersham Pharmacia, Sweden) as follows: 50 and 250 V for 1 h each followed from a linear increase from 250 to 4000 V, and finally 4000 V to give a total of 10000 V/h. Focused proteins were reduced and subsequently alkylated according to Görg et al.²⁸ IPG strips were then placed on the top of vertical slabs of polyacrylamide (12% T and 2.6% C) and sealed 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) SDS. Electrophoretic migration along the second dimension was performed using a Mini-Protean 3 Cell (Bio-Rad) under a voltage of 30 V for 20 min, followed by 150 V for 1.5 h. After completion of SDS-polyacrylamide gel electrophoresis (PAGE), gels were stained with colloidal Coomassie Brilliant Blue using the PageBlue Protein Staining Solution (Fermentas, Waltham, MA) following the manufacturer's instructions.

Image Analysis

Digitized images at 36.6 μm resolution were obtained using the GS-800 Calibrated Densitometer and Quantity One 4.6.2 software (Bio-Rad). Scan setting for comassie blue stain with red filter and transmissive light path was used. Computerized 2D gel analysis, including spot detection and quantification, was performed using the PDQuest Basic 8.0.1 software (Bio-Rad). The relative molecular mass was calibrated with internal protein markers (Precision Plus Protein Standards, Bio-Rad) after comigration during the second dimension. Quantification of detected protein spots was performed calculating the relative optical density \times area (relative OD \times area) in the gels. Nine gels per group of samples [3 biological repetitions \times 3 replicates (independent protein extractions)] were analyzed to detect qualitative and quantitative differences in protein expression among the five groups. The mean relative OD \times area \pm SD ($n = 9$) values of each group were finally used to estimate relative expression level (relative OD \times area %) of each protein spot among the groups. Differences among the means were evaluated by the Dunn's Multiple Comparison Test after that the null hypothesis (equal means) was rejected in the Kruskal–Wallis test, assuming a significance of $p \leq 0.05$. When the value of Relative OD \times area% of the control samples was not null, the relative expression ratio in the other groups was estimated. Values higher than 1.5 or lower than 0.5 were considered as significant. Coefficients of variation associated to number of spots detected in the biological replicates were 3.94, 5.36, 4.81, 5.18, and 5.10%, for C, A, aA, E, and aE, respectively. Coefficients of variation linked to the technical procedures (technical replicates) ranged between 0.83 and 2.48%.

In-Gel Digestion and NanoLC–MS/MS Analysis

Gel bands of interest were cut out manually. In-gel digestion was then performed with an automated protein digestion system, MassPREP station (Waters, Milford, MA). The gel plugs were washed twice with 50 μL of 25 mM NH_4HCO_3 and 50 μL of acetonitrile. The cysteine residues were reduced by 50 μL of 10 mM dithiothreitol for 30 min at 57 °C and alkylated by 50 μL of 55 mM iodoacetamide for 25 min at room temperature. After dehydration of the gel bands with acetonitrile, proteins were digested in gel with 25 μL of 12.5 ng/ μL modified porcine

trypsin (Promega, Madison, WI) in 25 mM NH_4HCO_3 , at 37 °C overnight. The generated peptides were extracted with 60% acetonitrile in 5% formic acid and analyzed by nanoliquid chromatography coupled to tandem mass spectrometry (nanoLC-MS/MS).

NanoLC–MS/MS were performed using either a nanoACQUITY Ultra-Performance-LC system (Waters) coupled to a Q-TOF SYNAPT G1 mass spectrometer (Waters) or a nanoLC–Chip/MS (Agilent Technologies, Palo Alto, CA) coupled to an ion trap amaZon (Bruker Daltonics, Bremen, Germany).

Ion Trap System. The chip contained a Zorbax 300SB-C18 column (150 mm \times 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 0.1% formic acid in acetonitrile and 2% water (solvent B). Four μL of each sample were loaded into the enrichment column at a flow rate set to 3.75 $\mu\text{L}/\text{min}$ with A. Elution was performed at a flow rate of 300 nL/min with a 8–40% linear gradient (solvent B) in 30 min followed by a 3 min stage at 70% of B before reconditioning the column at 8% of B.

The MS instrument was operated with the following settings: source temperature was set to 135 °C while cone gas flow was at 3 L/min. The nanoelectrospray voltage was set to -1850 V. The MS spectra were acquired in the positive ion mode on the mass range 250 to 1500 m/z using the standard enhanced resolution mode at a scan rate of 8.1 $m/z/s$. The Ion Charge Control was fixed at 200000 with a maximum accumulation time of 200 ms and the number of averages was set to 2. For tandem MS experiments, the system was operated with automatic switching between MS and MS/MS modes. The 6 most abundant peptides were selected on each MS spectrum for further isolation and fragmentation with a preference for doubly charged ions (absolute threshold of 5000 and a relative of 5%). Fragmentation was performed using argon as the collision gas. Ions were excluded after the acquisition of 2 MS/MS spectra and the exclusion was released after 0.12 min. The Smart Parameters Setting option was used for the selected precursor ions. The MS/MS spectra were acquired on the mass range 100–2000 m/z . The Ion Charge Control was fixed at 300000 and 2 scans were averaged to obtain a MS/MS spectrum. The complete system was fully controlled by HyStar 3.2 (Bruker Daltonics) software.

Q-TOF System. Five microliters of each sample were loaded on a Symmetry C18 precolumn (20 \times 0.18 mm, 5 μm particle size, Waters) and the peptides were separated on an ACQUITY UPLC BEH130 C18 column (200 mm \times 75 μm , 1.7 μm particle size, Waters). The solvent system consisted of 0.1% formic acid in water (solvent A) and 0.1% formic acid in acetonitrile (solvent B). Trapping was performed during 3 min at 15 $\mu\text{L}/\text{min}$ with 99% of A and 1% of B. Elution was performed at a flow rate of 450 nL/min, using a 1–35% gradient of B for 9 min at 50 °C followed by a fast rise at 90% of B and back to 1% B.

The Q-TOF G1 system was operating in positive mode, the capillary voltage was set at 2.8 kV and the cone voltage at 35 V. Mass calibration of the TOF was achieved using phosphoric acid (H_3PO_4) on the mass range 50–2000 m/z . Online correction of this calibration was performed with Glu-fibrino-peptide B as the lock-mass. The ion $(M + 2H)^{2+}$ at m/z 785.8426 was used to calibrate MS data and the fragment ion $(M+H)^+$ at m/z 684.3469 was used to calibrate MS/MS data during the analysis.

For tandem MS experiments, the system was operated with automatic switching between MS and MS/MS modes (MS in the range 250–2000 m/z and MS/MS 0.7 s/scan in the range 50–2000 m/z). The 3 most abundant peptides (intensity threshold 15 counts/s), preferably doubly and triply charged ions, were

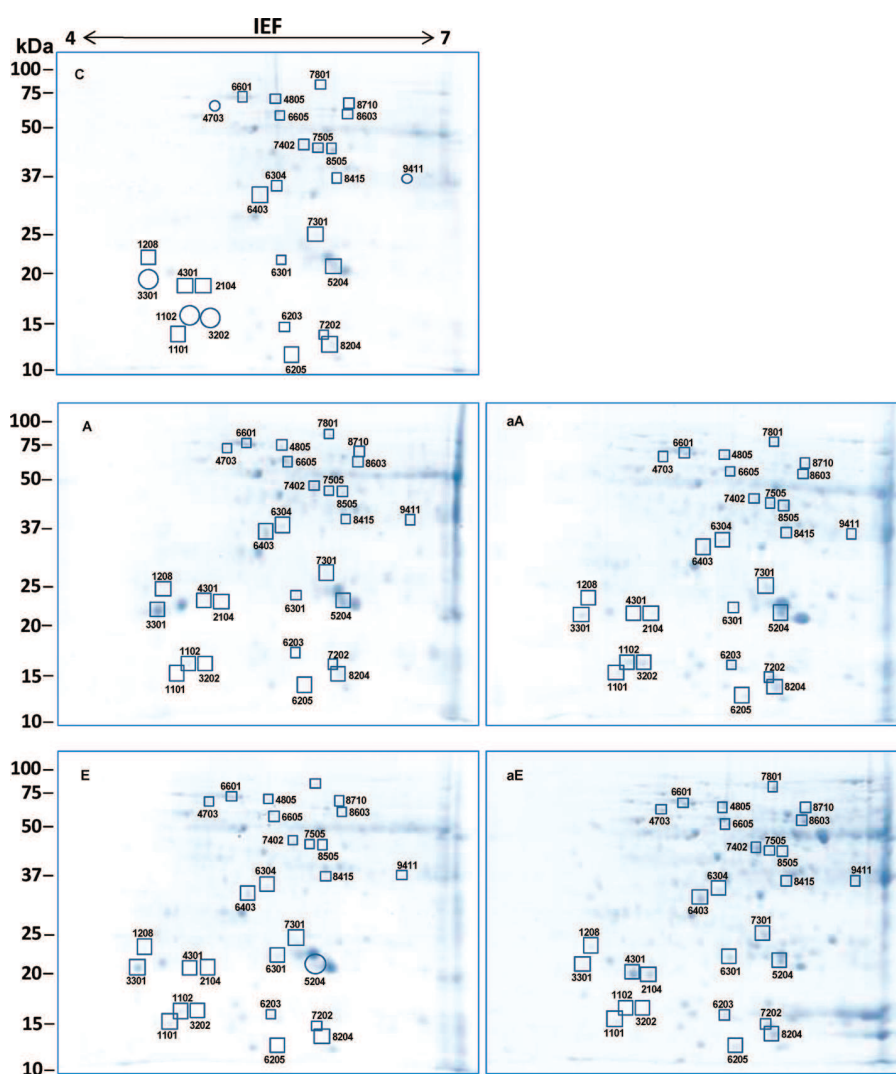


Figure 1. Map of the 28 identified protein spots differentially expressed among the groups of samples: control stems (C); symptomatic and asymptomatic stems of apoplectic (A and aA) and esca proper-affected (E and aE) plants. Isoelectric focusing (IEF) was performed on precast dry polyacrylamide 7 cm length gels ReadyStrip IPG (pH 4–7). The relative molecular mass (kDa) was calibrated with standard protein markers (Prestained SDS-PAGE Standards, Bio-Rad) after cosecond dimensional electrophoresis. Spots detected in at least one replica-gel of each sample are indicated with a square. Spots that were not detected in any gel of the same group are indicated with a circle.

selected on each MS spectrum for further isolation and CID fragmentation with 2 energies set using collision energy profile. Fragmentation was performed using argon as the collision gas. The complete system was fully controlled by MassLynx 4.1 (SCN 639, Waters).

Protein identification. Mass data collected during nanoLC-MS/MS were processed, converted into “.mgf” files (for ion trap analysis) with DataAnalysis 4.0 (Bruker Daltonics) or in “.pkl” (for Q-TOF G1 analysis) with ProteinLynx Global Server 2.4 PLGS (Waters) and interpreted using the MASCOT 2.3.02. algorithm (Matrix Science, London, UK). Searches were performed without any molecular weight, or isoelectric point restrictions against an in-house generated protein database composed of protein sequences of *Viridiplantae* and *Fungi* and known contaminant proteins such as human keratins and trypsin, downloaded from National Center for Biotechnology Information nonredundant database (NCBI nr, January 27, 2011) concatenated with reversed copies of all sequences (total 3640226 entries). Trypsin was selected as enzyme and for

MS/MS data, a parent and fragment mass tolerance of ± 0.25 Da for ion trap data and of 10 ppm on parent and 0.2 Da on fragment for Q-TOF data were set for the searches. A maximum of 2 missed cleavage was allowed and some modifications were taken into account: carbamidomethyl (C), acetyl N-terminus of protein, oxidation (M). The Mascot results were loaded into the Scaffold software (Proteome Software, Portland, OR). To minimize false positive identifications, results were subjected to very stringent filtering criteria as follows: (1) for the identification of proteins with two peptides or more, a Mascot ion score must be higher than 30 and had to be minimum 10 below the 95% Mascot significance threshold (“Identity score”); (2) in the case of single peptide hits, the Mascot score of the unique peptide must be higher than 30 and have a minimal “difference score” of 5 compared to the “Identity score”. The target-decoy database search allowed us to control and estimate the false positive identification rate of our study.²⁹ Thus, the final catalogue of proteins presents an estimated false positive rate below 1%. A list of all identified proteins is provided in Supporting

Table 1. Identified Proteins Differentially Expressed in the 5 Groups, Symptomatic and Asymptomatic Stems of Apoplectic and Esca Proper-affected Vines, and in the Stems of Control Vines

spot ^a	M_r (kDa) ^b	protein matched ^c	no. of unique peptides ^d	coverage % ^e	accession no. ^f	further blast ^g		
						protein	species	accession no. ^f
1101	66.97	hypothetical protein [<i>Vitis vinifera</i>]	1	4	gi 147811887	polyphenol oxidase	<i>Camellia nitidissima</i>	gi 222093457
1101	42.77	S-adenosylmethionine synthase 5 [<i>Vitis vinifera</i>]	1	13	gi 223635289			
1102	67.33	catechol oxidase precursor [<i>Vitis vinifera</i>]	3	7	gi 1076272			
1208	25.96	Proteasome subunit alpha type-5 [<i>Glycine max</i>]	1	8	gi 12229923			
2104	25.01	hypothetical protein [<i>Vitis vinifera</i>]	3	18	gi 225455238	small heat shock protein	<i>Petunia hybrida</i>	gi 232278
3202	67.33	catechol oxidase precursor [<i>Vitis vinifera</i>]	3	7	gi 1076272			
3301	23.84	hypothetical protein [<i>Vitis vinifera</i>]	2	13	gi 225426793	thaumatin-like	<i>Vitis vinifera</i>	gi 33329390
3301	23.86	osmotin-like [<i>Vitis vinifera</i>]	2	9.33	gi 1839046			
4301	25.01	hypothetical protein [<i>Vitis vinifera</i>]	1	8	gi 225455238	small heat shock protein	<i>Petunia hybrida</i>	gi 232278
4703	64.97	hypothetical [<i>Vitis vinifera</i>]	6	17	gi 147799889	PDI-like protein	<i>Arabidopsis thaliana</i>	gi 21592996
4805	61.27	unnamed protein [<i>Vitis vinifera</i>]	6	15	gi 296084626	heat shock protein 70 kDa	<i>Spinacia oleracea</i>	gi 2654210
5204	27.71	hypothetical protein [<i>Vitis vinifera</i>]	4	21	gi 225446775	oxygen-evolving enhancer protein 2, chloroplast precursor, putative	<i>Ricinus communis</i>	gi 255561128
6203	15.94	hypothetical protein [<i>Vitis vinifera</i>]	1	9	gi 225449595	actin depolymerizing factor 1	<i>Populus trichocarpa</i>	gi 224059520
6205	15.26	hypothetical protein isoform 2 [<i>Vitis vinifera</i>]	2	21	gi 225451120	superoxide dismutase	<i>Fagus sylvatica</i>	gi 38228697
6301	24.85	glutathione S-transferase 5 [<i>Vitis vinifera</i>]	2	17	gi 158323772			
6304	39.85	hypothetical protein [<i>Vitis vinifera</i>]	2	10	gi 225439868	diaminopimelate epimerase	<i>Glycine max</i>	gi 255671799
6304	34.27	hypothetical protein [<i>Vitis vinifera</i>]	2	12	gi 147843260	60S acidic ribosomal protein P0	<i>Glycine max</i>	gi 1710587
6403	29.08	hypothetical protein [<i>Vitis vinifera</i>]	2	10	gi 225469270	myo inositol monophosphatase, putative	<i>Ricinus communis</i>	gi 255581603
6601	71.15	similar to HSC70-1 [<i>Vitis vinifera</i>]	8	16	gi 225449497			
6605	55.46	ATP synthase subunit alpha [<i>Gossypium hirsutum</i>]	1	19	gi 110832063			
6605	61.35	GroEL-like chaperone, ATPase [<i>Vitis vinifera</i>]	4	8	gi 225433375			
7202	18.13	hypothetical protein [<i>Vitis vinifera</i>]	1	9	gi 147798422	heat-shock protein, putative	<i>Ricinus communis</i>	gi 25555886
7301	27.97	unnamed [<i>Vitis vinifera</i>]	4	25	gi 297741493	cytosolic ascorbate peroxidase	<i>Vitis vinifera</i>	gi 16177878
7402	39.17	hypothetical protein [<i>Vitis vinifera</i>]	7	29	gi 225460600	peroxidase	<i>Nicotiana tabacum</i>	gi 14031049

Table 1. Continued

spot ^a	M_r (kDa) ^b	protein matched ^c	no. of unique peptides ^d	coverage % ^e	accession no. ^f	further blast ^g		
						protein	species	accession no. ^f
7505	42.91	S-adenosylmethionine synthase 4 [<i>Vitis vinifera</i>]	5	18	gi 223635287			
7801	92.24	hypothetical protein [<i>Vitis vinifera</i>]	8	12	gi 225442981	phospholipase D alpha	<i>Vitis vinifera</i>	gi 84620126
8204	17.32	hypothetical protein [<i>Vitis vinifera</i>]	4	37	gi 225429592	small heat shock protein 17.3 kDa	<i>Vitis vinifera</i>	gi 269838636
8204	17.27	hypothetical protein [<i>Vitis vinifera</i>]	2	37	gi 225429614	small heat shock protein 17.3 kDa	<i>Vitis vinifera</i>	gi 269838636
8415	40.59	predicted: O-acetylserine(thiol)- lyase [<i>Vitis vinifera</i>]	4	13	gi 225434133			
8415	92.94	hypothetical protein [<i>Vitis vinifera</i>]	5	9	gi 14785331	aldo/keto reductase, putative	<i>Ricinus communis</i>	gi 255542314
8505	43.08	hypothetical protein [<i>Vitis vinifera</i>]	10	39	gi 147791974	S-adenosylmethionine synthetase	<i>Gossypium hirsutum</i>	gi 307948774
8505	42.77	S-adenosylmethionine synthase 5 [<i>Vitis vinifera</i>]	2	39	gi 223635289			
8603	62.54	hypothetical protein [<i>Vitis vinifera</i>]	10	21	gi 225428898	phosphoglycerate dehydrogenase	<i>Gossypium hirsutum</i>	gi 211906486
8603	62.41	similar to pyruvate decarboxylase [<i>Vitis vinifera</i>]	2	5	gi 225443847			
8710	76.64	hypothetical protein [<i>Vitis vinifera</i>]	4	8	gi 225430776	succinate dehydrogenase, putative	<i>Ricinus communis</i>	gi 255579273
9411	35.45	cytosolic malate dehydrogenase	1	4	gi 10334493	cytosolic malate dehydrogenase	<i>Cicer arietinum</i>	gi 10334493
9411	37.03	hypothetical protein [<i>Vitis vinifera</i>]	4	15	gi 147790725	aldo/keto reductase	<i>Manihot esculenta</i>	gi 62526573

^a Spot code as reported in Figures 1 and 2. ^b Observed relative molecular mass calibrated with standard protein markers. ^c Protein matched after searching against in house made database from NCBI nr database performed with MASCOT derived data. ^d Number of unique peptides matching the protein sequence. ^e Percentage of the protein sequence covered by the matching peptides. ^f Accession no. of the matched protein as reported in the NCBI database. ^g Results obtained after a further blast against the NCBI database (when no function of the protein matched with MASCOT derived data was reported) by using the accession no.

Information Tables S1 and S2. The spectra of single peptide hits are provided in the Supporting Information (Table S3). The MS data have been submitted to the PRIDE proteomics identification database³⁰ (www.ebi.ac.uk/pride) under accession numbers #19645 for Q-TOF MS data and #19646 for Ion Trap MS data. The data were converted using PRIDE Converter³¹ (<http://code.google.com/p/pride-converter>).

RNA Extraction

Total RNA was isolated from 3 × 50 mg of green stem powder using the Plant RNA Purification Reagent (Invitrogen, Cergy Pontoise, France). The RNA pellet was resuspended in 20 μL of RNase-free water, then treated with RQ1 DNase enzyme (Promega) and quantified by measuring the absorbance at 260 nm following manufacturer's instructions.

Real-time RT-PCR Analysis of Gene Expression

In total, 150 ng of total RNA were reverse-transcribed using the Verso SYBR 2-step QRT ROX enzyme (ABgene, Surrey, U.K.) according to the manufacturer's protocol. PCR conditions were those described by Bézier et al.³² Expression of 9 selected genes selected from proteomic results was tracked by quantitative

Reverse-Transcripts Polymerase Chain Reaction (qRT-PCR) using the primers indicated in Table 2. Reactions were carried out in a real-time PCR detector Chromo 4 apparatus (Bio-Rad) using the following thermal profile: 15 s at 95 °C (denaturation) and 1 min at 60 °C (annealing/extension) for 40 cycles. Efficiency of the primer sets was estimated by performing real-time PCR on several dilutions. For each experiment, PCR reactions were performed in duplicate for two (aA) or three (A, aE and E) independent experiments were analyzed. Results correspond to the means (± standard deviation) of the independent experiments. They were normalized with the *EF1-α* gene³³ and expressed relatively to the control corresponding to a fixed value of 1. Control samples consisted of green stems collected from control plants. The analyzed genes were considered significantly up- or down-regulated when changes of their expression was >2× or <0.5×, respectively.

RESULTS AND DISCUSSION

Isolation of Fungal Species Associated to Grapevine Trunk Diseases

The major causal agents of esca proper (Pch, Pal and Fom) were isolated from discolored woody tissues (trunk and arms) of

Table 2. Genes Analyzed by Real-time Reverse-transcription Polymerase Chain Reaction

gene	primer sequences	Genbank or TC TIGR accession number
<i>EF1-α</i> (elongation factor 1- α , housekeeping gene)	5'-GAACTGGGTGCTTGATAGGC-3' 5'-AACCAAAATATCCGGAGTAAAAGA-3'	GU585871
<i>OEE2p</i> (oxygen-evolving enhancer protein 2)	5'-GCTGACGGAGATGAAGGTGG-3' 5'-TGTCCCCAGCCTGTACCTTG-3'	AY222741
<i>PglyDH</i> (phosphoglycerate deshydrogenase)	5'-CGTCGAAGATGCTCAATGATGA-3' 5'-CCCCACGAGCAACATTAATT-3'	XM_002285322
<i>SAM</i> (<i>S</i> -adenosyl-methionine synthetase)	5'-CCTGAAATCAAAGTTCTCCTTCACA-3' 5'-CCGGCCCTGAAATCAAAGTT-3'	XM_002266322
<i>POX</i> (peroxidase)	5'-GAAGGGTGAAATTCGTGCTAACTG-3' 5'-TTGTCTCTTCCACCACTGTCTT-3'	XM_002263518
<i>cysAPX</i> (cytosolic ascorbate peroxidase)	5'-AAGGTCGCTTGCCTAATGCA-3' 5'-TGGCACCTCCCCAGAGTATG-3'	EU280159
<i>PPO</i> (polyphenol oxidase)	5'-TGGTCTTGTGATAAGCCTAGTGA-3' 5'-TCCACATCCGATCGACATTG-3'	XM_002276076
<i>SOD</i> (superoxide dismutase)	5'-GTGGACCTAATGCAGTGATTGGA-3' 5'-TGCCAGTGGTAAGGCTAAGTTCA-3'	AF056622
<i>GSTphi</i> (glutathione-S-transferase <i>phi</i> class)	5'-GTGATTGCCATGCAGAAGAG-3' 5'-TTTTTGGTAGAAACGCTTTATGTT-3'	XM_002262806.1
<i>TL</i> (thaumatin-like protein)	5'-CCTAACACCTTAGCCGAATTCGC-3' 5'-GGCCATAGGCACATTAATCCATC-3'	AF532965

the three groups of plants analyzed (esca proper- and apoplexy-affected, and control plants). Fom was directly linked to white decay. Other fungi associated to grapevine trunk diseases, as *Botryosphaeriaceae* species²⁵ and *Eutypa lata*,⁸ were also isolated from woody tissues. In the opposite, no fungi were isolated from either nondiscolored wood of trunk and arms or discolored and nondiscolored woody tissues of one-year-old stems, as well as from green stems of control or diseased plants. All these fungal isolates were identified according to the methods of Crous and Gams³⁴ (Pch), Essakhi et al.³⁵ (Pal), Fischer³⁶ (Fom), Phillips³⁷ (*Botryosphaeriaceae*), Crous et al.³⁸ (*Botryosphaeriaceae*) and Carter³⁹ (*Eutypa lata*), respectively.

Differences in Protein Yield and Number of Spots Detected among the Groups of Samples

Protein yield ranged between 0.37 ± 0.03 (A) and 0.49 ± 0.02 (C) μg per mg of fresh weight, while values of 0.38 ± 0.08 , 0.41 ± 0.13 and 0.42 ± 0.08 $\mu\text{g}/\text{mg}$ were registered for E, aE and aA stems, respectively. Different average number of protein spots could be resolved on 2-DE gels among the groups (Figure 1). Highest value was observed for aE (138), followed by A (136), aA (132), E (127) and C (118).

These first results suggest that a slightly lower protein concentration is likely present especially in symptomatic stems (A and E) of apoplectic (A) and esca proper-affected (E) plants than in control stems. Interesting is also the trend observed for asymptomatic stems (aA and aE) of symptomatic plants, which might indicate the occurrence of an altered physiology not only in symptomatic stems but also in the asymptomatic parts of the crown. On the other hand, the average number of detected spots in the pH range used in this study (4–7) seems to indicate that a higher number of proteins are synthesized especially in aE, A and aA.

Identification of Differentially Expressed Protein Spots by nanoLC–MS/MS

Only protein spots detected in at least 2 biological repetitions of a given group were analyzed and compared among all the

groups. Consequently, a total of 123 protein spots were matched and compared by image analysis performed on 45 2-DE gels (3 biological repetitions \times 3 replicates per group) (Figure 1). Among those differentially expressed in at least one group, 28 spots were subjected to in gel trypsin digestion followed by nanoLC–MS/MS analysis which led to the identification of 33 proteins (Table 1). In most of the cases (21 spots), searching against a custom in-house plant and fungi protein database did not allow the attribution of the functional role of these proteins. Thus, an additional bioinformatic analyses was performed to find known biological functions of identified proteins and homologous protein sequences from plant species other than *Vitis vinifera* L. were also taken into account (Table 1). Nevertheless, results of further blast were regarded as reliable only when identities of at least 98% were found. The number of listed proteins (37) higher than that of the analyzed spots (28) is due to the identification of two distinct gene products in 9 spots (1101, 3301, 6304, 6605, 8204, 8415, 8505, 8603 and 9411) (Table 1).

Five proteins were present in more than one spot with slightly different isoelectric point (pI) or different molecular weight (M_r). These observations suggest that these variations in the proteome can be due to different isoforms, which could correspond to protein variants and/or post-translational modifications (PTM) like phosphorylation or glycosylation (pI changes), or to truncation or degradation products, the latter inducing significant differences between theoretical and observed M_r values. For example, catechol oxidase was identified in both spots 1102 and 3202 displaying a slight shift in the pI which could be induced by the presence of different isoforms including some PTMs. However, the low number of peptides identified (3) does not allow to hypothesize whether the 2 proteins represent members of a multigenic family or post-transcriptional forms of the same protein. Another example shows that the same *S*-adenosylmethionine synthase (gi|223635289) was identified in two spots differing both in M_r and pI (1101 and 8505). In the case of the *S*-adenosylmethionine synthase identified in the spot 1101, the observed M_r was much

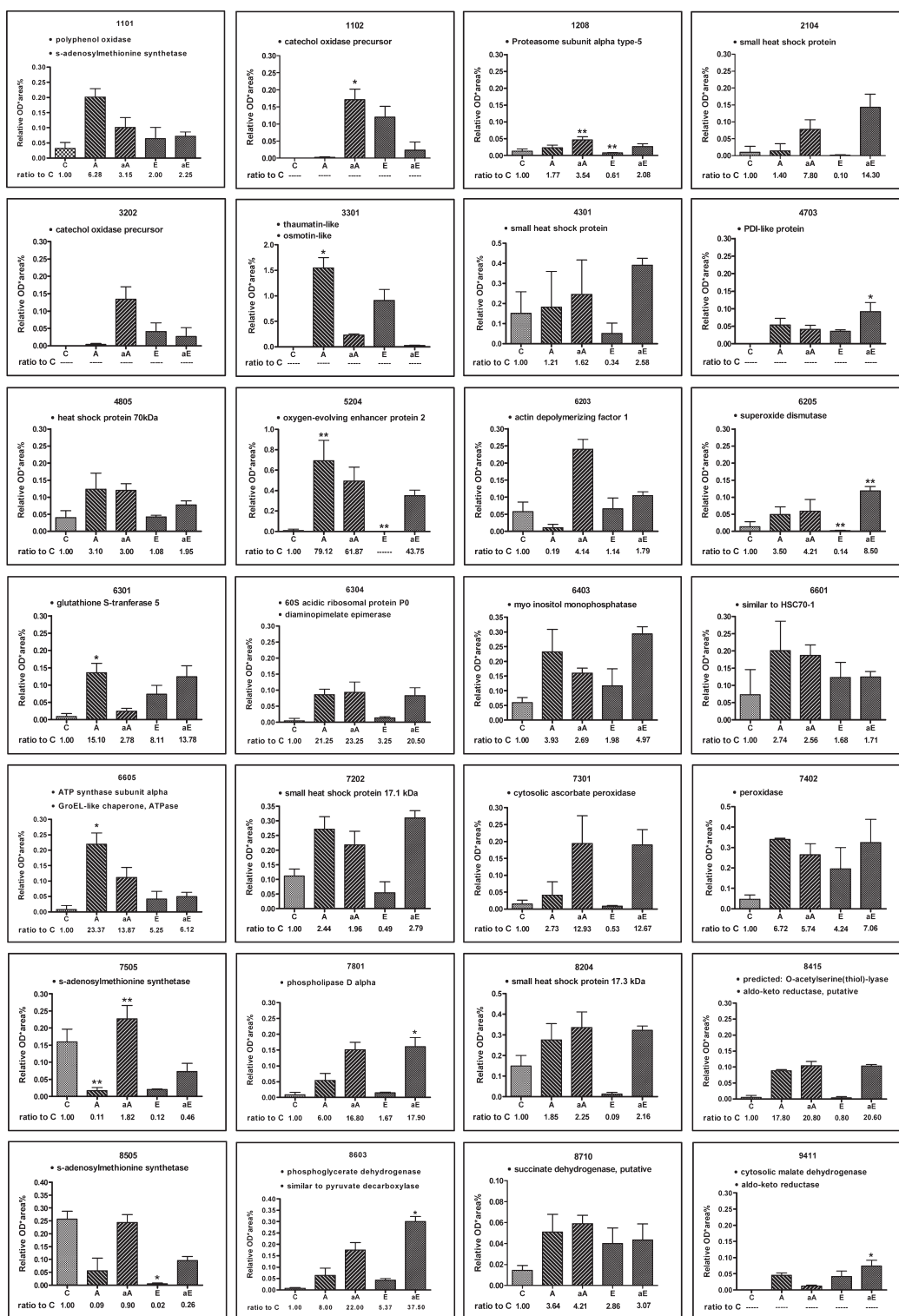


Figure 2. Quantification of differentially expressed proteins extracted from stems of control (C) and from symptomatic (A and E) and asymptomatic (aA and aE) stems of apoplectic (A) and esca-affected (E) 26-year-old standing vines cultivar Chardonnay. Vertical axis represents the relative value of optical density \times area (Relative OD \times area%) of the total value observed in the gel images as a measure of expression. Histograms indicate the mean quantitative values \pm SD calculated from replica gels. Relative mean values \pm SD are from three relative values obtained by dividing the value of each spot to the total density of the corresponding gel. Differences among the means were evaluated by the Dunn's Multiple Comparison Test after that the null hypothesis (equal means) was rejected in the Kruskal–Wallis test, assuming a significance of $P \leq 0.05$. One asterisk indicates a statistically significant difference with the control. Two asterisks indicate a statistically significant difference between the two classes where they were added. When the value of Relative OD \times area% of the control was not null, the relative expression ratio in the other classes was estimated.

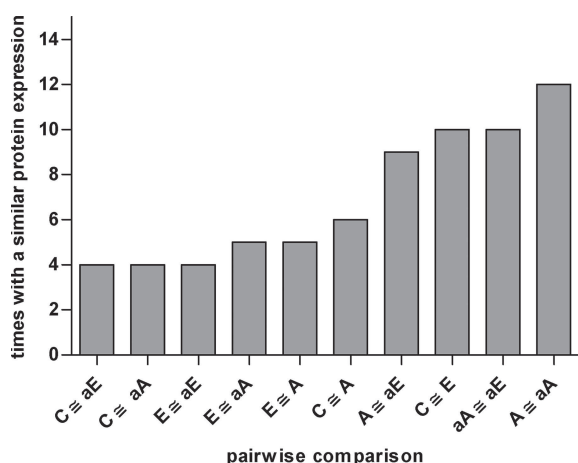


Figure 3. Pairwise comparison between each possible pair of samples from symptomatic (A and E) and asymptomatic (aA and aE) stems of apoplectic (aA and A) and esca proper-affected (aE and E) plants, and from control plants (C). The number of times where their relative expression was considered similar (ratio comprised between 0.5 and 1.5) is reported.

lower than the theoretical one (Figure 1; Table 1). A similar situation was observed for the catechol oxidase identified in both spots 1102 and 3202, the hypothetical protein from *V. vinifera* identified in the spot 8415 (gi|147853311) and, though in a lesser extent, for the hypothetical protein from *V. vinifera* identified in the spot 5204 (gi|225446775). It was suggested that it might be the result of protein degradation.⁴⁰ However, in the case of the spot 1101, searching against in house made database revealed the presence of a hypothetical protein from *V. vinifera* (gi|225466463; Table S1) which showed 99% of identity to gi|222093457, and resulted homologous to the same polyphenol oxidase listed in Table 1. This protein has a theoretical M_r of 19.96 kDa, which is somewhat consistent with the M_r in the gels. This result suggests that these two proteins might be isoforms and therefore no degraded proteins were probably present in the spot 1101.

Differentially Expressed Proteins among the Groups of Samples

Six spots out of the 28 differentially expressed protein spots identified were not detected in all the groups. It was the case of spots 1102, 3202, 3301, 4703, and 9411 in control stems (Figure 1 gel C), and spot 5204 in symptomatic stems of esca proper-affected plants (Figure 1, gel E). Nevertheless, when the relative expression values of the five groups were compared, significant differences between C stems and at least another group were revealed in all spots (Figure 2). In 15 spots (54% of the identified spots), expression in C stems differed from each other group, resulting the lowest in 9 spots. Since many of the identified proteins are known to be induced in response to biotic and/or abiotic stresses, or are more or less directly involved in such a trend likely represents the result of a different protein expression specifically related to esca proper and apoplexy. After a pairwise comparison, the highest levels of similarity in protein expression were observed among all asymptomatic stems (aA and aE) and symptomatic stems of apoplectic plants (A) (Figure 3), while the lowest levels were noted when C and E stems were considered. Interestingly, expression values in E stems showed the highest similarity when compared to those of C stems. Identified proteins were included into three groups conceptually defined as proteins involved in

(i) the primary metabolism and energy, (ii) stress tolerance and (iii) defense response.

Proteins Involved in Metabolism and Energy

On the primary metabolism, our results show that the energy production is affected in diseased plants since protein expressions involved in both photosynthesis and respiration were changed. Indeed, esca proper-affected plants commonly show reduction in leaf chlorophyll content resulting in chlorosis.¹⁴ The perturbations in photosynthetic activity and hence on carbohydrate metabolism have been explored by biochemical and molecular analysis in grapevine leaves.^{14,41} In this study, we identified a putative oxygen evolving enhancer protein 2 precursor (OEE2p) of *Ricinus communis* (gi|255561128) (spot 5204, Table 1). The oxygen evolving enhancer protein 2 (OEE2p, K02717) is a component of the photosynthetic complex of the thylakoid membrane called photosystem II (PSII), which is known to be disturbed by pathogen infection. In agreement with the results previously quoted, strong differences in relative expression of OEE2p were recorded among the stem groups (Figure 2). Whereas it was very few expressed in C and not detected in E stems, values ranging from 44 up to near 80-fold respect to C were noticed in the other groups (aE, aA and A) (Figure 2). Transcript analysis of *OEE2* did not correlate with the level of protein expression in stem (Figure 4). In contrast, the tendency of this gene expression in A stem was a weak repression. Nevertheless, the role of this protein seemed to be important during interaction of the grapevine and the wood diseases. Camps et al.⁴² reported the up-regulation of *OEE2* as a gene that may prevent Eutypa dieback symptom development in leaves, another grapevine trunk disease. Moreover, an interesting study on the response of grapevine shoot to water deficit or salinity indicated that the OEE2 isoforms detected were not affected by these abiotic stresses.⁴³ Similar trend of the OEE2p relative expression was noted for a chloroplastic ATP synthase subunit alpha (gi|110832063) identified in the spot 6605 (Table 1). An increased synthesis of both OEE2p and chloroplastic ATP synthase subunit alpha might indicate a response of photosynthetic apparatus triggered mainly in A and aA stems (Figure 2).

Regarding respiration, two enzymes assigned to the citric acid cycle, namely pyruvate decarboxylase (EC 4.1.1.1) and succinate dehydrogenase (EC 1.3.99.1), were differentially expressed in stem of affected plants. The pyruvate decarboxylase was one of the two proteins identified in the spot 8603 (gi|225443847) (Table 1). This enzyme that participates in the citric acid or glyoxylate cycle,⁴⁴ was overexpressed especially in asymptomatic stems, both aA and aE. Similarly, the succinate dehydrogenase (gi|255579273) identified in the spot 8710 (Table 1) was overexpressed in asymptomatic stems but also in symptomatic stems, A and E (Figure 2). The latter enzyme takes place in both the citric acid cycle and the electron transport chain. The energy production alteration was also suggested by the overexpression of the cytosolic malate dehydrogenase (cyMDH, EC 1.1.1.37) identified in the spot 9411 (gi|10334493) (Table 1). Nearly equal values of relative expression were observed in A and E stems while the highest and lowest levels were registered in aE and aA, respectively (Figure 2). Instead, the cyMDH spot was undetected in C stems. In accordance with our results, Yao et al.⁴⁵ reported that its overexpression might lead to an increased production of ATP. A higher production of ATP might result by an increased supply of malic acid to the citric acid cycle after its import into mitochondria by a malate-OAA shuttle.^{45,46}

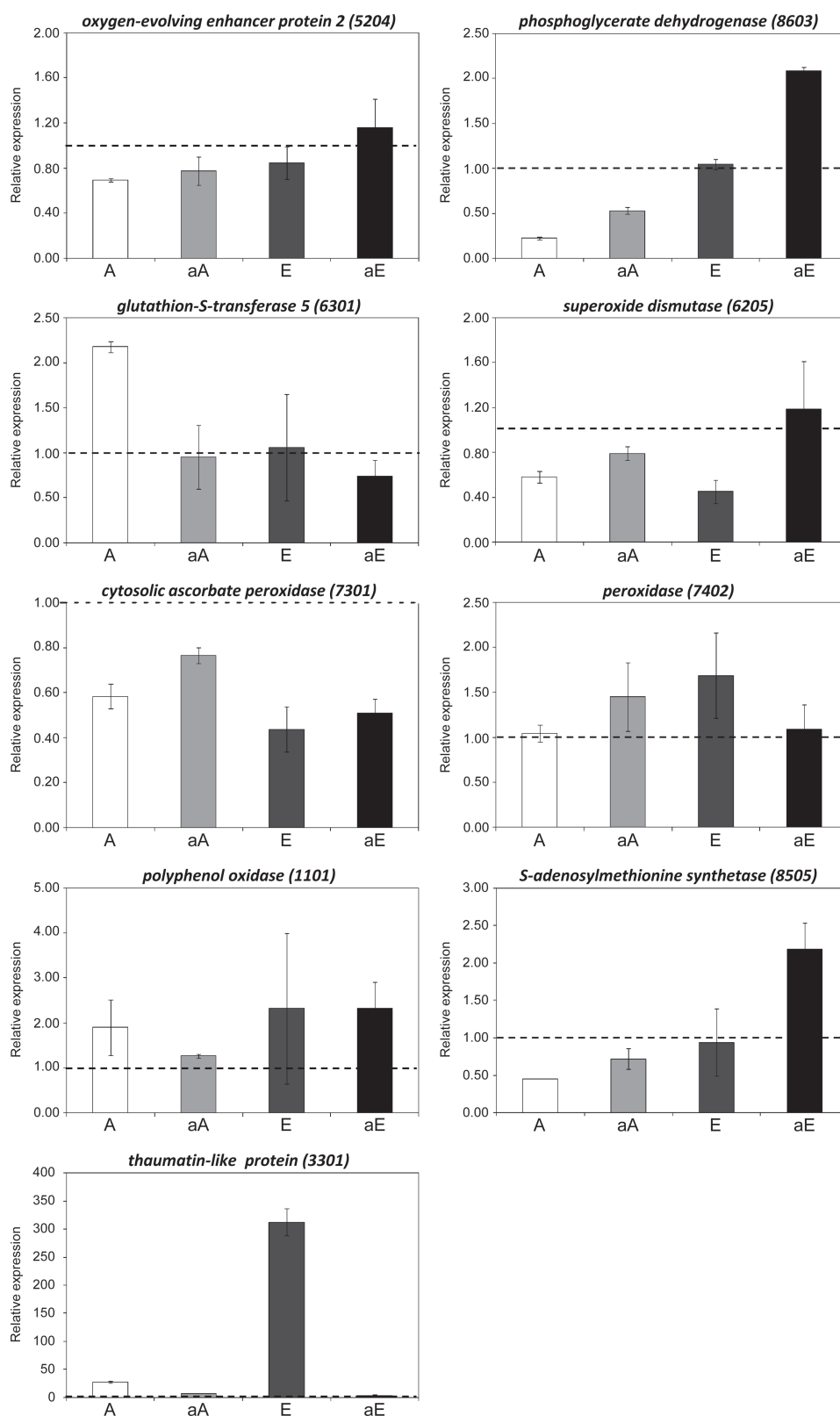


Figure 4. Expression levels of nine selected genes in symptomatic (A and E) and asymptomatic (aA and aE) stems of apoplectic (A) and esca-affected (E) 26-year-old standing vines cultivar Chardonnay by quantitative reverse-transcription polymerase chain reaction. Results correspond to means (\pm standard deviation) of two (A) or three independent experience (aA, E, aE and C). Gene expression was considered as significantly up- or down-regulated to the $1\times$ controls (dotted lines), when changes in relative expression were $>2\times$ or $<0.5\times$, respectively.

Other functions of the primary metabolism seemed to be affected. A 60S acidic ribosomal protein P0 (gi|1710587) was identified in the spot 6304 (Table 1) and was detected in all groups. Relative expression levels in A, aA, E and aE were higher than those observed in C (Figure 2). However, whereas the value in E was 3.25-fold respect to C, values ranging from 20.50 up to 23.25-fold were registered in the remaining groups. The 60S acidic ribosomal proteins are implicated in protein synthesis by interacting with elongation factors, mRNA and tRNA during translation,^{47,48} and have a role in promoting the selective translation of some transcripts in response to various stresses.^{48,49} Indeed, 60S acidic ribosomal protein -P0 and -P1 are known to be up-regulated in the shoots of grapevine under abiotic stress.⁴³

Our study shows that the metabolism of some amino acids was also altered. The diaminoepimelate epimerase (gi|255671799, EC 5.1.1.7, spot 6304) was overexpressed especially in A, aA and aE (Table 1; Figure 2). This enzyme takes place in the lysine biosynthesis pathway (<http://www.genome.jp/kegg/pathway/map/map00300.html>). Two other enzymes, phosphoglycerate dehydrogenase (gi|211906486, EC 1.1.1.95, spot 8603) and O-acetylserine(thiol)-lyase (gi|225434133, EC 2.5.1.49, spot 8415), involved respectively in the biosynthesis of serine and cysteine were identified (Table 1; Figure 2). Except for the value of spot 8415 in E, an up-regulation respect to C was noted for both enzymes. A significant up-regulation of the gene expression *PglyDH* (encoding for phosphoglycerate dehydrogenase identified in spot 8603) was observed for aE, with positive correlation to protein expression result (Figure 4). Another positive correlation was observed between gene expression and protein level for E stems. No correlation was observed in the case of the apoplexy, where protein levels were higher in samples aA and A in comparison to the control while *PglyDH* was down-regulated (significantly for A) in the same samples (Figure 4). The apoplexy appearance is defined by a rapid event,³ then the effect on gene regulation has been rapidly observed in samples and later on protein synthesis. We can hypothesized that if samplings of A and aA green stems had been made late after apoplexy appearance, similar proteomic analyses could have shown a low level of this protein.

Finally, the proteasome complex plays a central role in plant protein degradation and its role in plant-pathogen interactions has been highlighted over the past few years.^{24,50} In this study, the protein spot 1208 showed 100% of identity to the homologous 26S proteasome subunit alpha type-5 of *Glycine max* (gi|12229923) (Table 1). It was more expressed in aA and aE stems and the highest and lowest values of relative expression were observed for aA and E, respectively (Figure 2). Two spots (6605 and 4703) of the protein fate category were identified as a homologous of a GroEL-like chaperone (ATPase, gi|225433375, spot 6605) and a disulfide-isomerase (PDI, gi|21592996, EC 5.3.4.1, spot 4703) (Table 1). Both were overexpressed in asymptomatic and symptomatic stems as compared to the control. For GroEL-like chaperone, the highest value of relative expression was observed in A stems and for the PDI, it was in aE (Figure 2). The latter is commonly known as a molecular chaperone which catalyzes dithiol-disulfide interchange reactions and promotes protein disulfide formation, isomerization, or reduction, depending on the redox potential and nature of the polypeptide substrate.^{51,52} However, the induction of PDI in response to fungal infection has also been reported.⁵³

Proteins Involved in Stress Tolerance

Among identified proteins, actin-depolymerizing factor (ADF, gi|224059520), identified in the spot 6203, is known to play an

important role in regulating actin dynamics in different cellular processes.^{54,55} This protein was overexpressed in aA stems compared to C stems. Furthermore, whereas ADF expression was similar in E and aE stems, it was very different in apoplectic plants, which asymptomatic (aA) and symptomatic (A) stems registered the highest and lowest expression level, respectively. These results suggest that a possible diverse response involving ADF occurred in symptomatic stems of apoplectic (A) and esca proper-affected (E) plants. A putative *myo*-inositol monophosphatase (spot 6403, gi|255581603) was detected in each group but an overexpression ranging from 1.98 (E) up to 4.97 (aE) compared to C stems was noticed (Table 1; Figure 2). *Myo*-inositol monophosphatase (EC 5.5.1.4) lead to inositol (*myo*-inositol) biosynthesis. This molecule and its methylated derivatives are correlated with tolerance to drought and/or salinity.⁵⁶

A phospholipase D (PLD, EC 3.1.4.4) has been identified in spot 7801 (gi|84620126). PLD is linked to the signaling pathway in response to stress conditions: drought, salinity or pathogen attack.^{57,58} In grape berry, PLDs are thought to participate in the early response to low temperature and to be involved in accumulation of SA and thermotolerance induced by temperature acclimation.⁵⁹ The lowest value of relative expression was registered for C, while the highest were observed for both aA and aE, whose values were respectively 16.8- and 17.9-fold respect to C. Our results suggest that increased expression of PLD in asymptomatic stems of both apoplectic and esca proper-affected plants (aA and aE) was related to the presence of symptoms on the plant and/or to the possible imminent symptom development on themselves, but also that it was limited when symptoms directly affected the stem (A and E).

Among heat shock proteins (HSPs), small HSPs (smHSPs) are regarded as molecular chaperone accumulating in response to stresses, especially heat^{60,61} by binding partially denatured proteins for preventing irreversible protein inactivation and aggregation.⁶⁰ Indeed, cloning experiments showed that three genes (*HSG4*, *HSG14* and *HSG19*) encoding for smHSPs might be related to the acquisition of thermotolerance in grapevine.⁶² In this study, four spots (2104, 4301, 7202 and 8204) were identified as smHSP. Spots 2104 and 4301, identified as a smHSP of ~25 kDa (gi|232378), resulted to be identical, while spots 7202 and 8204 were identified as HSP (gi|255558876) and smHSP of 17.3 kDa (gi|269838636), respectively (Table 1). All of them were always detected in C stems. However, the highest values of relative expression were registered for A, aA and aE whereas the lowest were observed for E (Figure 2). A smHSPs, the Hsp21, was known to provide protection and recovery of PSII efficiency (then photosynthesis activity) in grapevine leaves in response to heat stress.⁶³ Unlike smHSPs, high molecular weight HSPs, like 70 kDa HSPs (HSP 70s), play fundamental roles in basal cell functions:⁶⁰ (i) as molecular chaperones involved in protein folding and protein transport across membranes and (ii) in the cell response to a wide range of stresses.⁶⁴ In our study, the protein spots 4805 and 6601 were identified as HSP 70 kDa (gi|2654210) and similar to heat shock cognate (HSC) 70-1 (gi|225449497),⁶⁴ respectively. Both of them were detected in all groups, though the highest values of relative OD × area% were recorded for both A and aA stems (Figure 2). Vincent et al.⁴³ reported that two HSP70s may participate in maintaining protein metabolism shoots under water deficit and salinity in grapevine. Our results confirm and reinforce the information according to the appearance of apoplexy symptoms implicate complex physiological perturbations, but one of them concerned water stress.

Five enzymes taking part in oxido-reduction reactions were found to be perturbed in apoplectic or esca-affected plants. Among them, the aldo-keto reductase is a superfamily encompassing a range of oxidoreductases that could play a role in limiting injuries caused by oxidative stresses.^{65,66} The homologous of an aldo-keto reductase (spot 9411, gi|62526573) (Table 1) was observed in nearly equal values of relative expression in A and E stems, while the highest and lowest levels were registered in aE and aA, respectively (Figure 2). Instead, this protein was undetected in C stems.

In response to an oxidative burst, glutathione represents also an important component of plant stress responses.⁶⁷ A glutathione S-transferase (GST) phi class (spot 6301, gi|158323772) was identified in this study and was mainly expressed in A, aE and E, respectively (Table 1, Figure 2). A significant up-regulation for *GSTphi* gene expression has also been observed in A (Figure 4), with positive correlation to protein expression (Figure 3). No correlation between gene expression and protein levels was revealed for aA, E and aE. An extensive study on the effect of GLSD symptoms revealed an induction of leaf glutathione metabolism (genes of *GST phi* and *tau* class, total GST activity) before symptom appearance.¹⁶ Similar up-regulation of *GST1* (encoding for GST tau) was observed in the visually healthy leaves of GLSD-⁴¹ or apoplexy- affected plants.³ Interestingly, an opposite situation was evident for the same spot (6301) in A and aA, where the relative value in aA was lower than in A, suggesting that a possible different response involving glutathione could be activated in apoplexy. In regard to these results, we can hypothesize a role of the GST system also in the detoxification of fungal metabolites and in the appearance of the syndrome type (apoplexy or esca proper).

Another enzyme taking part in the antioxidant system is the superoxide dismutase (SOD, EC 1.15.1.1), which catalyzes the conversion of O_2^- , in H_2O_2 and O_2 by adding H_2O . SOD is known to be up-regulated in the leaves of Flavescence dorée phytoplasma (FDP)-infected grapevine.⁶⁸ In our study, a SOD with 100% of identity to the homologous of *Fagus sylvatica* was exclusively identified in the spot 6205 (gi|38228697, Table 1). Highest and lowest values of relative OD \times area% were recorded for aE and E, respectively, while similar values were observed for A and aA (Figure 2). The lower relative expression in E respect to C confirmed that SOD is a constitutively synthesized enzyme⁶⁹ but also clearly indicated that a heavy dysfunction of its synthesis might occurred in E. Concerning the *SOD* gene encoding for SOD protein identified in the spot 6205, a significant repression of this gene expression was observed for E stems in comparison to the C samples (Figure 4). This repression may be positively correlated with the low level of SOD protein observed in E stem samples (Figure 3). Moreover, similar repression of *SOD* was observed in pre-apoplectic and drying leaves from apoplectic plants³ and in asymptomatic leaves of GLSD-affected plants.⁴¹ These results reinforce the role of SOD in the perturbation observed in esca-proper affected plants and may indicate a lack of oxidative stress control, which is lethal for the plant. The general tendency of the gene expression (A, aA, aE samples) was a slight repression (not statistically proved) and no correlation was observed with protein expression for A, aA and aE samples. Moreover, we observed strong differences of relative expression for a cytosolic ascorbate peroxidase (APX, gi|161778778, EC 1.11.1.11, spot 7301) (Table 1), where the values in aA and aE were higher than those in C, A and E (Figure 2). Similar up-regulation of APX have been observed in leaves of FDP-infected grapevine.⁶⁸ In our

study, a significant differential down-regulation was observed for *APX* expression in the E stems (Figure 4) with a slight correlation of the low protein level. Considering altogether the relative perturbation of antioxidant system, ROS regulation is really critical and continuously modulated during symptom expression.

Proteins Involved in Defense Response

H_2O_2 is a catalyst of cell-wall reinforcement and implicates the action of the peroxidase (POX, EC 1.11.1.7). A POX (gi|14031049) was identified in the spot 7402. Expression levels were higher in the stems of affected plant (A, aA, E aE) than in the control (C), with the highest level for the A stems (Figure 2). RT-PCR analysis revealed no significant perturbation of *POX* expression in tested samples in comparison to the control (Figure 4). Like POX, polyphenol oxidase (PPO) has shown to be involved in plant resistance⁷⁰ through the production of toxic metabolites (for pathogens), like phytoalexins, phenols and lignins and its wound inducibility.^{71,72} An increased activity of apoplastic PPO (gi|222093457, EC 1.14.18.1, spot 1101) was detected in all the groups but the highest relative expression was recorded in A. Nevertheless, gene analysis revealed no changes in the relative expression of *PPO* (encoding for the protein identified in spot 1101), independently of the sample (Figure 4). It should be noted that there are three types of PPO: laccase, cresolase and catechol oxidase.⁷³ In our study, we also identified two isoforms of a catechol oxidase precursor (gi|1172587) in spots 1102 and 3202 (Table 1). Both spots 1102 and 3202, which showed a similar trend of relative expression levels among the groups, were undetected in C and were very faint in A (Figure 2). In both cases, the highest values of relative expression were recorded for aA. These results confirm the accumulation of phenols observed in leaves with symptoms of esca proper-infected plants.⁹

Finally, the enzyme s-adenosylmethionine (SAM) synthetase (SAMS, EC 2.5.1.6) was identified in the spots 7505 (gi|223635287), 8505 (gi|307948774 and gi|223635289) and 1101 (gi|223635289) (Table 1). Very similar trends of relative expression were observed for all the groups in the spots 7505 and 8505, where highest values for C and aA and lowest for A and E were registered (Figure 2). A significant down-regulation of the *SAM* gene expression was observed only in A (Figure 4) with positive correlation to the protein level (Figure 2). This enzyme (SAMS) produces the SAM, which is the major methyl group donor in plants⁷⁴ and lead to the biosynthesis of polyamines⁷⁵ and ethylene.⁷⁶ Polyamines participate in responses to biotic and abiotic stresses.⁷⁷ Ethylene, a stress hormone, plays a role in plant disease resistance pathway.^{76,78} Figueiredo et al.⁷⁹ reported an up-regulation of SAMS in grapevine leaves in response to cryptogamic fungi. Conversely, our results suggest that a possible ethylene- and/or PAs-mediated response in A, E and aE stems would have been limited by the decreased expression of SAMS in response to wood diseases.

Among the 17 families of pathogenesis related (PR) proteins recognized so far,⁸⁰ thaumatin-like (TL) and osmotin-like proteins form the PR-5 family.²⁴ Members of this family are known to have antifungal⁸¹ and antioomycete⁸² properties, and may also have function in plant resistance.^{80,82} In our study, a thaumatin- (gi|33329390) and an osmotin-like (gi|1839046) proteins were identified in the same spot (3301) (Table 1). Highest and lowest values of relative expression were noticed respectively in symptomatic (A and E) and asymptomatic (aA and aE) stems of both apoplectic and esca proper-affected plants, while the spot was not detected in C stems (Figure 2). Similar responses were observed

concerning the up-regulation of *TL* gene expression in the tested samples in comparison to the control (Figure 4). The higher value of *TL* induction was observed in E, followed by A > aE > aA (Figure 4). In agreement with our results, up-regulation of genes expression encoding for- or accumulation of- thaumatin- and/or osmotin-like proteins were described in leaves of rooted cuttings affected by *Eutypa lata*⁴² and of FDP- or *Plasmopara viticola*-infected grapevine.^{68,83} Relative to their antifungal properties, the increased synthesis of these two proteins in E stems could either be the effect of the translocation of toxic fungal metabolites from woody tissues to the foliage through the xylem flow^{2,19} or mean that fungal propagules are effectively present within symptomatic stems.

CONCLUSIONS

Remarkable quantitative differences were observed for several proteins according to the expressed syndrome (A and E). In contrast, before foliar symptom appearance of either esca proper or apoplexy a similar response strategy is triggered in affected plants since no qualitative differences between asymptomatic stems (aA and aE) were detected. Proteins involved in plant stress response such as OEE2p HSP, GSTphi, PPO were especially over-expressed in apoplectic green stems. On the contrary, plant responses developed in GLSD-affected plant green stems revealed a down-expression of OEE2, smHSPs, SOD and SAMs proteins. It could therefore be speculated that a diverse response (or effect) occurs in terms of protein expression upon onset of these different external symptoms. However, the highest expression values of two PR-5 proteins, thaumatin- and osmotin-like, in both A and E stems might strengthen the most accredited hypothesis that considers the translocation of toxic fungal metabolites from woody tissues to leaves by green stem (Andolfi et al., 2009; Surico et al., 2008) to explain esca proper and apoplectic symptom appearance. Overall, the present study suggests that esca proper and apoplectic symptom expression results in a different effect possibly provoked by the same causal agent(s) on plant metabolisms.

ASSOCIATED CONTENT

Supporting Information

Supplementary tables. This material is available free of charge via the Internet at <http://pubs.acs.org>.

AUTHOR INFORMATION

Corresponding Author

*Florence Fontaine, e-mail: florence.fontaine@univ-reims.fr. Phone: 3 33 26 91 33 18. Fax: 3 33 26 91 33 39.

Author Contributions

[†]These author contributed equally to this manuscript

ACKNOWLEDGMENT

This research was financed by the national program CPER (Contract Projet État-Region) program – task 1E altesca and CASDAR (Compte d’Affectation Spéciale au Développement Agricole et Rural). The company Moët & Chandon is thanked for making available the vineyard that represented the experimental plot of this study.

REFERENCES

- (1) Surico, G. Towards a redefinition of the diseases within the esca complex of grapevine. *Phytopathol. Mediterr.* **2009**, *48*, 5.
- (2) Surico, G.; Mugnai, L.; Marchi, G. The esca disease complex. In *Integrated management of diseases caused by fungi, phytoplasma and bacteria*; Ciancio, A., Mukerji, K. G., Eds.; Springer: Heidelberg, 2008; pp 119–136.
- (3) Letousey, P.; Baillieul, F.; Perrot, G.; Rabenoelina, F.; Boulay, M.; Vaillant-Gaveau, N.; Clément, C.; Fontaine, F. Early events prior to visual symptoms in the apoplectic form of grapevine esca disease. *Phytopathology* **2010**, *100*, 424–431.
- (4) Graniti, A.; Surico, G.; Mugnai, L. Esca of Grapevine: a disease complex or a complex of diseases? *Phytopathol. Mediterr.* **2000**, *39*, 16–20.
- (5) Fussler, L.; Kobes, N.; Bertrand, F.; Mauray, M.; Grosman, J.; Savary, S. A characterization of grapevine trunk diseases in France from data generated by the National Grapevine Wood Diseases Survey. *Phytopathology* **2008**, *98*, 571–579.
- (6) Surico, G. The grapevine and wine production through the ages. *Phytopathol. Mediterr.* **2000**, *39*, 3–10.
- (7) van Niekerk, J. M.; Fourie, P. H.; Halleen, F.; Crous, P. W. *Botryosphaeria* spp. as grapevine trunk disease pathogens. *Phytopathol. Mediterr.* **2006**, *45*, S43–S54.
- (8) Kuntzmann, P.; Villaume, S.; Larignon, P.; Bertsch, C. Esca, BDA and Eutypiosis: foliar symptoms, trunk lesions and fungi observed in diseased vinestocks in two vineyards in Alsace. *Vitis* **2010**, *49*, 71–76.
- (9) Lima, M. R. M.; Felgueiras, M. L.; Graça, G.; Rodrigues, J. E. A.; Barros, A.; Gil, A. M.; Dias, A. C. P. NMR metabolomics of esca disease-affected *Vitis vinifera* cv. Alvarinho leaves. *J. Exp. Bot.* **2010**, *61*, 4033–4042.
- (10) Di Marco, S.; Mazzullo, A.; Calzarano, F.; Cesari, A. The control of esca: status and perspectives. *Phytopathol. Mediterr.* **2000**, *39*, 232–240.
- (11) Genty, B.; Briantais, J. M.; Baker, N. R. The relationship between quantum yield of photosynthetic electron transport and quenching of chlorophyll fluorescence. *Biochim. Biophys. ACTA: Gen. Subjects* **1989**, *990*, 87–92.
- (12) Krause, G. H.; Weis, E. Chlorophyll fluorescence and photosynthesis: the basics. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **1991**, *42*, 313–349.
- (13) Maxwell, K.; Johnson, G. N. Chlorophyll fluorescence: A practical guide. *J. Exp. Bot.* **2000**, *51*, 659–668.
- (14) Petit, A.-N.; Vaillant, N.; Boulay, M.; Clément, C.; Fontaine, F. Alteration of photosynthesis in grapevines affected by esca. *Phytopathology* **2006**, *96*, 1060–1066.
- (15) Glazebrook, J. Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. *Annu. Rev. Phytopathol.* **2005**, *43*, 205–227.
- (16) Valtaud, C.; Foyer, C. H.; Fleurat-Lessard, P.; Bourbouloux, A. Systemic effects on leaf glutathione metabolism and defence protein expression caused by esca infection in grapevines. *Func. Plant Biol.* **2009**, *36*, 260–279.
- (17) Agrelli, D.; Amalfitano, C.; Conte, P.; Mugnai, L. Chemical and spectroscopic characteristics of the wood of *Vitis vinifera* cv. Sangiovese affected by esca disease. *J. Agric. Food Chem.* **2009**, *57*, 11469–11475.
- (18) Urbez-Torres, J. R.; Leavitt, G. M.; Voegel, T. M.; Gubler, W. D. Identification and distribution of *Botryosphaeria* spp. associated with grapevine cankers in California. *Plant Dis.* **2006**, *90*, 1490–1503.
- (19) Andolfi, A.; Cimmino, A.; Evidente, A.; Iannaccone, M.; Capparelli, R.; Mugnai, L.; Surico, G. A new flow cytometry technique to identify *Phaeoaniella chlamydospora* exopolysaccharides and study mechanisms of esca grapevine foliar symptoms. *Plant Dis.* **2009**, *93*, 680–684.
- (20) Komatsu, S.; Konishi, H.; Shen, S.; Yang, G. Rice proteomics: a step toward functional analysis of the rice genome. *Mol. Cell Proteomics* **2003**, *2*, 2–10.
- (21) Castro, A. J.; Carapito, C.; Zorn, N.; Magné, C.; Leize, E.; Van Dorsselaer, A.; Clément, C. Proteomic analysis of grapevine (*Vitis vinifera* L.) tissues subjected to herbicide stress. *J. Exp. Bot.* **2005**, *56*, 2783–2795.

- (22) Giribaldi, M.; Giuffrida, M. G. Heard it through the grapevine: Proteomic perspective on grape and wine. *J. Proteomics* **2010**, *73*, 1647–1655.
- (23) Victor, K. J.; Fennell, A. Y.; Grimplet, J. Proteomic analysis of shoot tissue during photoperiod induced growth cessation in *V. riparia* Michx. grapevines. *Proteome Sci.* **2010**, *8*, 44.
- (24) Afroz, A.; Muhammad, Ali.; Ghulam.; Mir, Asif.; Komatsu, S. Application of proteomics to investigate stress-induced proteins for improvement in crop protection. *Plant Cell Rep.* **2011**, *30*, 745–763.
- (25) Spagnolo, A.; Marchi, G.; Peduto, F.; Phillips, A. J. L.; Surico, G. Detection of Botryosphaeriaceae species within grapevine woody tissues, with particular emphasis on the *Neofusicoccum parvum*/N. ribis complex. *Eur. J. Plant Pathol.* **2011**, *129*, 485–500.
- (26) Wang, W.; Scali, M.; Vignani, R.; Spadafora, A.; Sensi, E.; Mazzuca, S.; Cresti, M. Protein extraction for two-dimensional electrophoresis from olive leaf, a plant tissue containing high levels of interfering compounds. *Electrophoresis* **2003**, *24*, 2369–2375.
- (27) Röhrig, H.; Schmidt, J.; Colby, T.; Bräutigam, A.; Hufnagel, P.; Bartels, D. Desiccation of the resurrection plant *Craterostigma plantagineum* induces dynamic changes in protein phosphorylation. *Plant Cell Environ.* **2006**, *29*, 1606–1617.
- (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*, 122–124.
- (29) Elias, J. E.; Gygi, S. P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* **2007**, *4*, 207–214.
- (30) Vizcaino, J. A.; Côté, R.; Reisinger, F.; Barsnes, H.; Foster, J. M.; Rameseder, J.; Hermjakob, H.; Martens, L. The Proteomics Identifications database: 2010 update. *Nucleic Acids Res.* **2010**, *38*, D736–D742.
- (31) Barsnes, H.; Vizcaino, J.; Eidhammer, I.; Martens, L. PRIDE Converter: making proteomics data-sharing easy. *Nat. Biotechnol.* **2009**, *27*, 598–599.
- (32) Bézier, A.; Lambert, B.; Baillieul, F. Study of defense-related gene expression in grapevine leaves and berries infected with *Botrytis cinerea*. *Eur. J. Plant Pathol.* **2002**, *108*, 111–120.
- (33) Terrier, N.; Glissant, D.; Grimplet, J.; Barrieu, F.; Abbal, P.; Couture, C.; Ageorges, A.; Atanassova, R.; Léon, C.; Renaudin, J. P.; Dédaldéchamp, F.; Romieu, C.; Delrot, S.; Hamdi, S. Isogene specific oligo arrays reveal multifaceted changes in gene expression during grape berry (*Vitis vinifera* L.) development. *Planta* **2005**, *222*, 832–847.
- (34) Crous, P. W.; Gams, W. *Phaeoconiella chlamydospora* gen. et comb. Nov., a causal organism of Petri grapevine decline and esca. *Phytopathol. Mediterr.* **2000**, *39*, 112–118.
- (35) Essakhi, S.; Mugnai, L.; Crous, P. W.; Groenewald, J. Z.; Surico, G. Molecular and phenotypic characterisation of novel *Phaeoacremonium* species isolated from esca diseased grapevines. *Persoonia* **2008**, *21*, 119–134.
- (36) Fischer, M. A new wood-decaying basidiomycete species associated with esca of grapevine: *Fomitiporia mediterranea* (Hymenochaetales). *Mycol. Prog.* **2002**, *1*, 315–324.
- (37) Phillips, A. J. L. *Botryosphaeria* species associated with diseases of grapevines in Portugal. *Phytopathol. Mediterr.* **2002**, *41*, 3–18.
- (38) Crous, P. W.; Slippers, B.; Wingfield, M. J.; Rheeder, J.; Marasas, W. F. O.; Phillips, A. J. L.; Alves, A.; Burgess, T.; Barber, P.; Groenewald, J. Z. Phylogenetic lineages in the Botryosphaeriaceae. *Stud. Mycol.* **2006**, *55*, 235–253.
- (39) Carter, M. V. Wood and root diseases caused by fungi. *Eutypa dieback*. In *Compendium of grape diseases*, 3rd ed.; Pearson, R. C., Goheen, A. C., Eds; APS Press: St. Paul, MN, 1994; pp 32–34.
- (40) Alvarez, S.; Goodger, J. Q. D.; Marsh, E. L.; Chen, S.; Asirvatham, V. S.; Schachtman, D. P. Characterization of the maize xylem sap proteome. *J. Proteome Res.* **2006**, *5*, 963–972.
- (41) Magnin-Robert, M.; Letousey, P.; Spagnolo, A.; Rabenoelina, F.; Jacquens, L.; Mercier, L.; Clément, C.; Fontaine, F. Leaf stripe form of esca induces alteration of photosynthesis and defense reactions in pre-symptomatic leaves. *Func. Plant Biol.* **2011**, *38*, 856–866.
- (42) Camps, C.; Kappel, C.; Lecomte, P.; Léon, C.; Gomès, E.; Coutos-Thévenot, P.; Delrot, S. A transcriptomic study of grapevine (*Vitis vinifera* cv. Cabernet-Sauvignon) interaction with the vascular ascomycete fungus *Eutypa lata*. *J. Exp. Bot.* **2010**, *61*, 1719–1737.
- (43) Vincent, D. A.; Ergül, A.; Bohlman, C.; Tattersall, E. A. R.; Tillett, R. L.; Wheatley, M. D.; Woolsey, R.; Quilici, D. R.; Joets, J.; Schlauch, K.; Schooley, D. A.; Cushman, J. C.; Cramer, G. R. Proteomic analysis reveals differences between *Vitis vinifera* L. cv. Chardonnay and cv. Cabernet Sauvignon and their responses to water deficit and salinity. *J. Exp. Bot.* **2007**, *58*, 1873–1892.
- (44) Nelson, D. L.; Cox, M. M. G. The citric acid cycle. In *Lehninger Principles of Biochemistry*, 4th ed.; W. H. Freeman and Company: New York, 2005; pp 601–630.
- (45) Yao, Y.-X.; Dong, Q.-L.; Zhai, H.; You, C.-X.; Hao, Y.-J. The functions of an apple cytosolic malate dehydrogenase gene in growth and tolerance to cold and salt stresses. *Plant Physiol. Biochem.* **2011**, *49*, 257–264.
- (46) Pastore, D.; Trono, D.; Laus, M. N.; Fonzo, N. D.; Flagella, Z. Possible plant mitochondria involvement in cell adaptation to drought stress. *J. Exp. Bot.* **2007**, *58*, 195–210.
- (47) Montoya-García, L.; Muñoz-Ocoteo, V.; Aguilar, R.; Sánchez de Jiménez, E. Regulation of acidic ribosomal protein expression and phosphorylation in maize. *Biochemistry* **2002**, *41*, 10166–10172.
- (48) Bailey-Serres, J.; Vangala, S.; Szick, K.; Lee, C. H. Acidic phosphoprotein complex of the 60S ribosomal subunit of maize seedling roots. Components and changes in response to flooding. *Plant Physiol.* **1997**, *114*, 1293–1305.
- (49) Li, X.-J.; Yang, M.-F.; Chen, H.; Qu, L.-K.; Chen, F.; Shen, S.-H. Abscisic acid pretreatment enhances salt tolerance of rice seedlings: Proteomic evidence. *Biochim. Biophys. ACTA: Proteins Proteomics* **2010**, *1804*, 929–940.
- (50) Kurepa, J.; Smalle, J. A. Structure, function and regulation of plant proteasomes. *Biochimie* **2008**, *90*, 324–335.
- (51) Freedman, R. B. Protein disulfide isomerase: multiple roles in the modification of nascent secretory proteins. *Cell* **1989**, *57*, 1069–1072.
- (52) Afroz, A.; Khan, M. R.; Ahsan, N.; Komatsu, S. Comparative proteomic analysis of bacterial wilt susceptible and resistant tomato cultivars. *Peptides* **2009**, *30*, 1600–1607.
- (53) Ray, S.; Anderson, J. M.; Urmeev, F. I.; Goodwin, S. B. Rapid induction of a protein disulfide isomerase and defense-related genes in wheat in response to the hemibiotrophic fungal pathogen *Mycosphaerella graminicola*. *Plant. Mol. Biol.* **2003**, *53*, 741–754.
- (54) Tian, M. Y.; Chaudhry, F.; Ruzicka, D. R.; Meagher, R. B.; Staiger, C. J.; Day, B. *Arabidopsis* actin-depolymerizing factor *AtADF4* mediates defense signal transduction triggered by the *Pseudomonas syringae* effector *AvrPphB*. *Plant Physiol.* **2009**, *150*, 815–824.
- (55) Ouellet, F.; Carpentier, É.; Cope, M. J. T. V.; Monroy, A. F.; Sarhan, F. Regulation of a wheat actin-depolymerizing factor during cold acclimation. *Plant Physiol.* **2001**, *125*, 360–368.
- (56) Loewus, F. A.; Murthy, P. P. N. *myo*-Inositol metabolism in plants. *Plant Sci.* **2000**, *150*, 1–19.
- (57) Bargmann, B. O.; Munnik, T. The role of phospholipase D in plant stress responses. *Curr. Opin. Plant Biol.* **2006**, *9*, 515–522.
- (58) de Torres Zabela, M.; Fernandez-Delmond, I.; Niittyla, T.; Sanchez, P.; Grant, M. Differential expression of genes encoding *Arabidopsis* phospholipases after challenge with virulent or avirulent *Pseudomonas* isolates. *Mol. Plant-Microbe Interact.* **2002**, *15*, 808–816.
- (59) Wan, S. B.; Tian, L.; Tian, R. R.; Pan, Q. H.; Zhan, J. C.; Wen, P. F.; Chen, J. Y.; Zhang, P.; Wang, W.; Huang, W. D. Involvement of phospholipase D in the low temperature acclimation-induced thermotolerance in grape berry. *Plant Physiol. Biochem.* **2009**, *47*, 504–510.
- (60) Waters, E. R.; Lee, G. J.; Vierling, E. Evolution, structure and function of the small heat shock proteins in plants. *J. Exp. Bot.* **1996**, *47*, 325–338.
- (61) Yang, L.; Lin, H.; Takahashi, Y.; Chen, F.; Walker, M. A.; Civerolo, E. L. Proteomic analysis of grapevine stem in response to *Xylella fastidiosa* inoculation. *Physiol. Mol. Plant Pathol.* **2011**, *75*, 90–99.

- (62) Kobayashi, M.; Katoh, H.; Takayanagi, T.; Suzuki, S. Characterization of thermotolerance-related genes in grapevine (*Vitis vinifera*). *J. Plant Physiol.* **2010**, *167*, 812–819.
- (63) Wang, L.-J.; Fan, L.; Loescher, W.; Duan, W.; Liu, G.-J.; Cheng, J.-S.; Luo, H.-B.; Li, S.-H. Salicylic acid alleviates decreases in photosynthesis under heat stress and accelerates recovery in grapevine leaves. *BMC Plant Biol.* **2010**, *10*, 34.
- (64) Latijnhouwers, M.; Xu, X.-M.; Geir Møller, S. Arabidopsis stromal 70-kDa heat shock proteins are essential for chloroplast development. *Planta* **2010**, *232*, 567–578.
- (65) Jez, J. M.; Penning, T. M. The aldo-keto reductase (AKR) superfamily: an update. *Chem. Biol. Interact.* **2001**, *130–132*, 499–525.
- (66) Turóczy, Z.; Kis, P.; Török, K.; Cserhádi, M.; Lendvai, A.; Dudits, D.; Horváth, G. V. Overproduction of a rice aldo-keto reductase increases oxidative and heat stress tolerance by malondialdehyde and methylglyoxal detoxification. *Plant Mol. Biol.* **2011**, *75*, 399–412.
- (67) Kmives, T.; Gullner, G.; Kiraly, Z. Role of glutathione and glutathione-related enzymes in response of plants to environmental stress. *Ann. N.Y. Acad. Sci.* **1998**, *851*, 251–258.
- (68) Margaria, P.; Palmano, S. Response of the *Vitis vinifera* L. cv. 'Nebbiolo' proteome to Flavescence dorée phytoplasma infection. *Proteomics* **2011**, *11*, 212–214.
- (69) Alscher, R. G.; Erturk, N.; Heath, L. S. Role of superoxide dismutases (SODs) in controlling oxidative stress in plants. *J. Exp. Bot.* **2002**, *53*, 1331–1341.
- (70) Thipyapong, P.; Stout, M. J.; Attajarusit, J. Functional analysis of polyphenol oxidases by antisense/sense technology. *Molecules* **2007**, *12*, 1569–1595.
- (71) Thipyapong, P.; Joel, D. M.; Steffens, J. C. Systemic wound induction of potato (*Solanum tuberosum*) polyphenol oxidase. *Phytochemistry* **1995**, *18*, 193–215.
- (72) Constabel, C. P.; Bergey, D. R.; Ryan, C. A. Systemin activates synthesis of wounded-inducible tomato leaf polyphenol oxidases via the octadecanoid defense signaling pathway. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 407–411.
- (73) Sanchez-Amat, A.; Solano, F. A pluripotent polyphenol oxidase from the melanogenic marine alteromonas shares catalytic capabilities. *Biochem. Biophys. Res. Commun.* **1997**, *240*, 787–792.
- (74) Deytieux, C.; Geny, L.; Lapaillerie, D.; Claverol, S.; Bonneau, M.; Donèche, B. Proteome analysis of grape skins during ripening. *J. Exp. Bot.* **2007**, *58*, 1851–1862.
- (75) Roje, S. S-Adenosyl-L-methionine: Beyond the universal methyl group donor. *Phytochemistry* **2006**, *67*, 1686–1698.
- (76) Tsunezuka, H.; Fujiwara, M.; Kawasaki, T.; Shimamoto, K. Proteome analysis of programmed cell death and defense signaling using the rice lesion mimic mutant *cdr2*. *Mol. Plant-Microbe Interact.* **2005**, *18*, 52–59.
- (77) Hussain, S. S.; Ali, M.; Ahmad, M.; Siddique, K. H. M. Polyamines: Natural and engineered abiotic and biotic stress tolerance in plants. *Biotechnol. Adv.* **2011**, *29*, 300–311.
- (78) Arimura, G.; Ozawa, R.; Nishioka, T.; Boland, W.; Koch, T.; Kühnemann, F.; Takabayashi, J. Herbivore-induced volatiles induce the emission of ethylene in neighboring lima bean plants. *Plant J.* **2002**, *29*, 87–98.
- (79) Figueiredo, A.; Fortes, M. A.; Ferreira, S.; Sebastiana, M.; Choi, Y. H.; Sousa, L.; Acioli-Santos, B.; Pessoa, F.; Verpoorte, R.; Pais, M. S. Transcriptional and metabolic profiling of grape (*Vitis vinifera* L.) leaves unravel possible innate resistance against pathogenic fungi. *J. Exp. Bot.* **2008**, *59*, 3371–3381.
- (80) van Loon, L. C.; Rep, M.; Pieterse, C. M. J. Significance of inducible defense-related proteins in infected plants. *Ann. Rev. Phytopathol.* **2006**, *44*, 135–162.
- (81) Monteiro, S.; Barakat, M.; Piçarra-Pereira, M. A.; Teixeira, A. R.; Ferreira, R. B. Osmostin and Thaumatin from grape: A putative general defense mechanism against pathogenic fungi. *Phytopathology* **2003**, *93*, 1505–1512.
- (82) Colditz, F.; Niehaus, K.; Krajinski, F. Silencing of PR-10-like proteins in *Medicago truncatula* results in an antagonistic induction of other PR proteins and in an increased tolerance upon infection with the oomycete *Aphanomyces euteiches*. *Planta* **2007**, *226*, 57–71.
- (83) Kortekamp, A. Expression analysis of defence-related genes in grapevine leaves after inoculation with a host and non-host pathogen. *Plant Physiol. Biochem.* **2006**, *44*, 58–67.

III. Etude des troncs de vignes infectés par l'esca

Une partie des résultats obtenus au cours de ce travail sur les troncs de vignes atteints de la maladie de l'esca a fait l'objet d'un article soumis à *Phytopathologia Mediterranea* disponible en annexe 5.

1. Description des échantillons

Trois catégories de troncs de vignes de la variété Chardonnay (Figure 7) ont été étudiées: (1) des troncs de vignes contrôle (C1 et C2, Figure 7), qui ne présentent pas de symptômes visibles caractéristiques de l'esca, mais qui ont un bois blanc (C1) ou bois blanc parsemé de taches brunes (C2) semblables à celles de la maladie de l'esca ; (2) des troncs de vignes atteints de la forme aiguë de l'esca (A1 et A2, Figure 7) portant à la fois le bois blanc (A1) et un bois brun (A2) ; (3) des troncs de vignes atteints de la forme lente de l'esca (E1 et E2, Figure 7) avec les symptômes visibles et portant à la fois un bois blanc (E1) et un bois brun (E2).

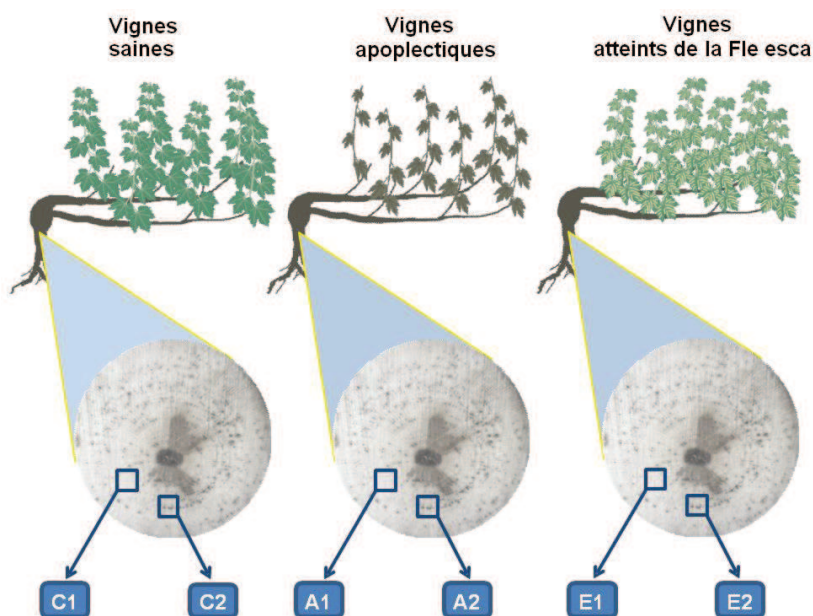


Figure 7. Représentation schématique de l'échantillonnage relatif aux différents symptômes de la maladie de l'esca sur les troncs de vignes.

2. Stratégie d'analyse protéomique des troncs de vignes

a. Stratégie globale de l'analyse protéomique

Les protéines ont été extraites des troncs de vignes de la même manière que dans l'étude précédente¹⁵.

Les extraits protéiques issus de bois de troncs de vignes sont déposés et séparés sur gel 2D (3 réplicats techniques et 3 réplicats biologiques, soit un total de 54 gels 2D).

Après révélation des protéines au bleu de Coomassie, et analyse des images de gels, 30 spots différenciellement exprimés sur les 150 spots (issus des 6 gels) ont été soumis à la stratégie d'analyse protéomique globale schématisée sur la Figure 8.

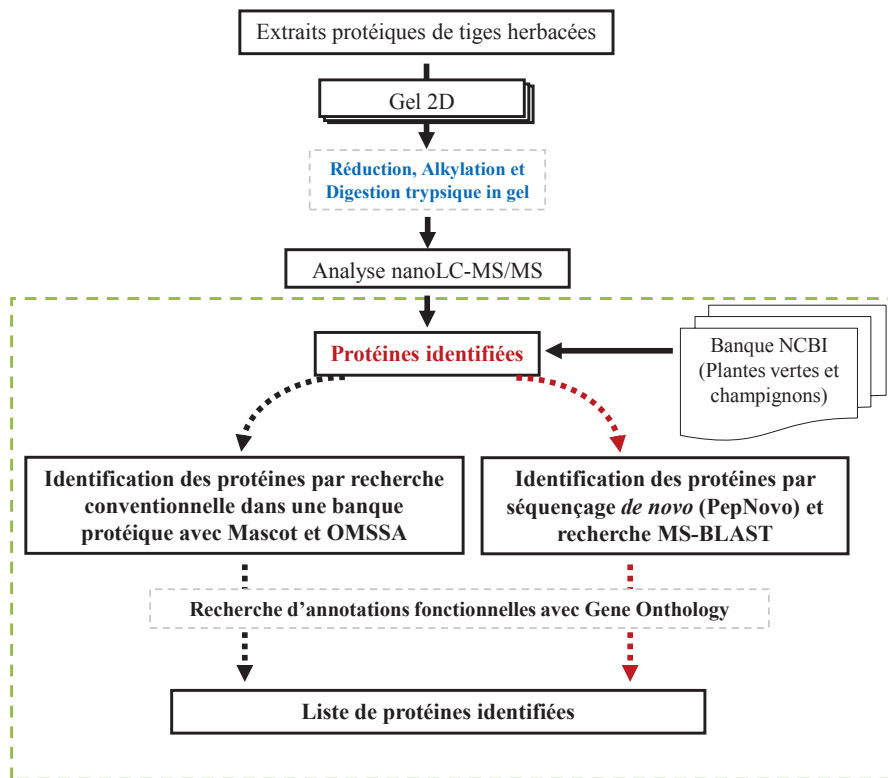


Figure 8. Stratégie globale d'analyse protéomique adoptée pour l'étude des troncs de vignes infectées par les deux formes de l'esca.

A la différence de la première étude sur **les tiges herbacées de vigne**, deux approches complémentaires d'identification de protéines ont été menées en parallèle (Figure 8) afin d'identifier à la fois les protéines de vignes mais aussi des protéines de champignons (dont les génomes ne sont pas séquencés) par homologie.

Les protocoles expérimentaux sont détaillés dans la partie expérimentale de la publication des résultats (Annexe 5). Nous ne détaillerons dans les prochains paragraphes que la partie identification des protéines encadrée en vert sur la Figure 8.

Contrairement aux tiges herbacées de vignes, les troncs de vignes sont le siège des infections par les champignons. L'absence de banques protéiques de ces champignons rend la stratégie d'identification par recherche conventionnelle dans les banques protéiques inadéquate et insuffisante pour cette étude.

Ainsi, une stratégie utilisant le séquençage *de novo* a été développée comme décrit dans la Figure 8. Le protéome de tiges infectées par les champignons comporte à la fois les protéines de vignes et des protéines de champignons. La protéolyse enzymatique va générer des peptides issus des protéines de différentes espèces qui vont être analysés par MS. Ces peptides vont être fragmentés pour donner des spectres MS/MS au cours des analyses LC-MS/MS, et ces spectres vont être interprétés pour identifier les protéines dont ils proviennent. La fragmentation MS/MS permet d'obtenir des informations de séquence d'acides aminés contenues dans les spectres MS/MS sous forme de listes de masses d'ions parents (m/z sélectionnés pour l'expérience MS/MS) et ions fragments (les m/z de chaque ion fils contenus dans les spectres de fragmentation MS/MS) (Figure 9).

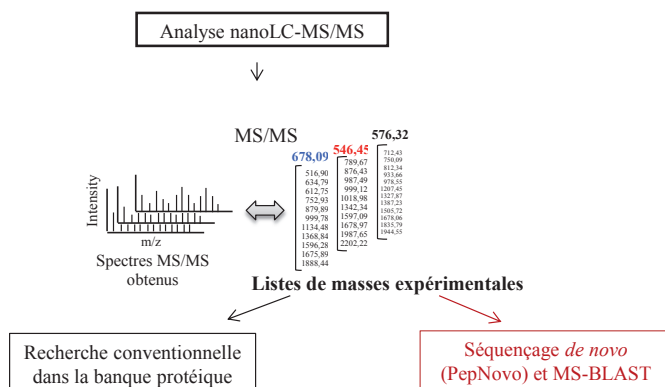


Figure 9. La première étape de la stratégie mise en place.

La recherche conventionnelle dans une banque protéique (PFF, Peptide Fragment Fingerprinting), repose sur la comparaison des listes de masses expérimentales d'ions parents et fragments obtenues au cours d'une expérience LC-MS/MS avec des listes de masses théoriques d'ions parents et fragments théoriques générés *in silico* à partir de banques protéiques. Les peptides ayant les mêmes ions parents et fragments que ceux contenus dans les listes de masses théoriques seront identifiés lors de cette recherche à l'aide d'un moteur de recherche comme Mascot ou OMSSA. Les protéines présentant une très forte homologie de séquence avec des protéines issues d'autres organismes séquencés de plantes vertes contenues dans la banque, seront identifiées grâce à des peptides tryptiques communs (Figure 10, (i) la protéine B a été identifiée avec les peptides **VFGNDFHEGR** et **FNCHDSFAQRI** qui existent dans la banque protéique).

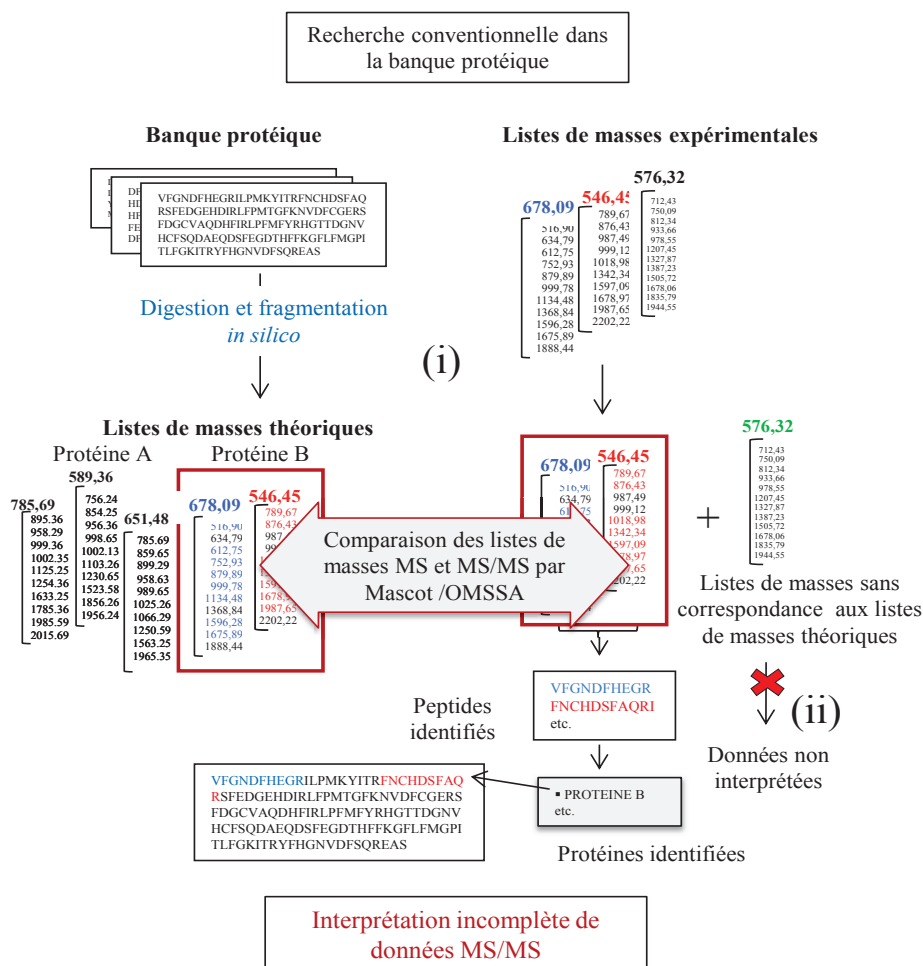


Figure 10. Interprétation de données MS/MS basée sur la recherche dans une banque protéique : (i) identification de protéines dont les séquences en acides aminés sont identiques aux peptides issus de la digestion *in silico* d'une banque protéique. (ii) Donnée non interprétée car elle ne correspond à aucune donnée théoriques issus de la banque protéique utilisée. Cette approche d'interprétation de données MS/MS est incomplète pour des organismes non séquencés.

Cependant, les peptides spécifiques aux protéines de champignons par exemple qui ne sont pas communs aux protéines contenues dans la banque, ne sont pas identifiés avec cette approche (Figure 10, (ii)). En effet pour identifier un peptide par cette approche, les masses de fragments sur les spectres MS/MS (donnée expérimentale) doivent être strictement identiques avec une liste de masse théoriques d'un peptide d'une ou des protéines d'organismes séquencés contenues dans la banque utilisée. Par exemple les peptides différents d'un ou plusieurs acides aminés dans leurs séquences ne sont pas identifiés puisqu'ils ne correspondent à aucune donnée théorique de la banque protéique utilisée. L'interprétation des données est donc incomplète dans ce cas.

Pour permettre d'identifier les séquences de ces peptides de protéines dont les données MS/MS sont de bonnes qualités et qui n'ont pas été interprétées lors de la première approche basée sur la banque protéique, nous avons utilisé l'approche par séquençage *de novo* suivi de MS-BLAST, une identification par homologie de séquence.

Le séquençage *de novo* est une interprétation des spectres MS/MS indépendantes des banques protéiques. En effet, le séquençage *de novo* permet d'obtenir les tags d'acides aminés (fragments de séquences

d'acides aminés composant un peptide), directement déduits des différences de masses entre les ions fragments successifs des spectres MS/MS (Figure 11).

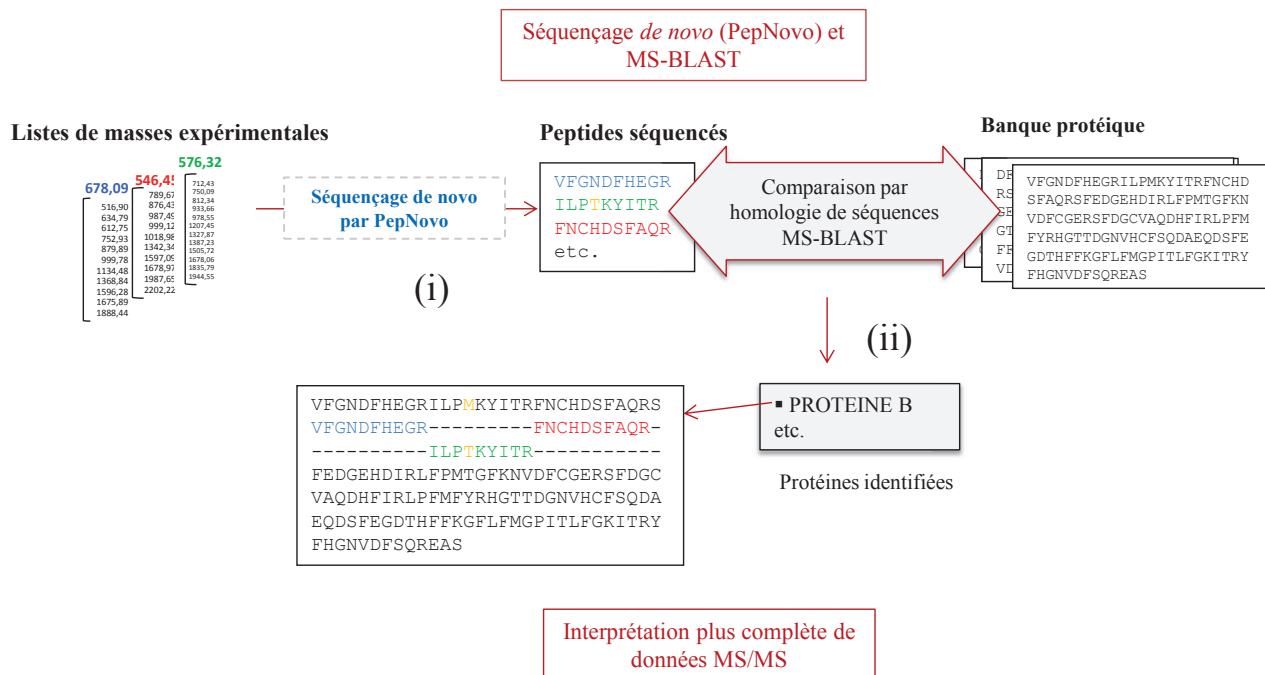


Figure 11. Interprétation de données MS/MS basée sur le séquençage *de novo* suivi de MS-Blast : une recherche par homologie de séquence. (i) les spectres MS/MS sont interprétés par séquençage *de novo*. (ii) Les résultats de séquençage (peptides séquencés) sont soumis à une recherche par homologie de séquence des protéines contenant dans une banque protéique pour identifier les protéines homologues. Cette approche d'interprétation de données MS/MS est plus complète que l'identification par recherche conventionnelle basée sur une banque protéique.

Puisque la qualité de spectre de fragmentation est importante pour reconstituer la séquence d'acides aminés des peptides fragmentés, il est donc indispensable d'avoir des spectres de bonne qualité contenant le maximum de fragments pour un peptide. En effet, selon la qualité du spectre de fragmentation MS/MS d'un peptide, on va obtenir une séquence d'acides aminés plus ou moins complète de ce peptide permettant de l'identifier. De plus le séquençage *de novo*, permet à la fois l'interprétation des spectres des peptides des organismes séquencés et non séquencés, ce qui n'est pas le cas avec l'approche conventionnelle d'interprétation de données MS/MS basée les banques protéiques. Comme le montre la Figure 11 (i), les peptides VFGNDFHEGR et FNCHDSFAQRI, existant dans la banque protéique et le peptide ILPTKYITR (homologue de ILPMKYITR) qui n'existe pas dans la banque protéique à cause de la différence entre T et M sont tous séquencés par *de novo*. Par la suite, une recherche par homologie de séquences est faite pour identifier les protéines homologues (Figure 11, (ii)).

Cette interprétation des données est plus complète que la recherche conventionnelle dans une banque de données protéique classiquement utilisée en analyse protéomique. En effet, elle permet d'interpréter plus données MS/MS notamment les données MS/MS des peptides appartenant aux protéines contenues dans la banque de recherche et d'autres protéines non séquencés. Ainsi cette interprétation offre la possibilité d'identifier les protéines d'organismes non séquencés, par homologie de séquences aux protéines homologues d'autres espèces dont les génomes sont séquencés.

b. Identification des protéines des troncs de vigne par recherche conventionnelle dans une banque protéique avec les algorithmes de recherche Mascot et OMSSA

L'identification des protéines a été faite de la même façon que lors de l'étude des tiges herbacées. A la différence de la première étude, deux algorithmes de recherche Mascot et Open Mass Spectrometry search Algorithm OMSSA (algorithme de recherche en accès libre) ont été utilisés afin d'augmenter la fiabilité dans nos identifications^{20, 21}.

La banque protéique était composée de séquences protéiques de *vitis et fungi*, des kératines humaines et de la trypsine. Toutes les séquences protéiques ont été extraites de la version non redondante de NCBI (datée du 03 septembre 2012) et additionnées des copies inversées de toutes les séquences (au total : 3409536 entrées).

L'identification et la validation des résultats ont été faites selon la procédure détaillée dans la partie expérimentale de la publication de l'annexe 5.

c. Identification des protéines des troncs de vigne par séquençage *de novo* (PepNovo) et recherche MS-BLAST.

Nous avons utilisé PepNovo qui est un algorithme de séquençage *de novo* automatique et qui permet de traiter les centaines de spectres MS/MS générés lors d'une analyse LC-MS/MS.

PepNovo utilise un raisonnement probabiliste pour modéliser les événements de fragmentation de peptides dans un spectromètre de masse ²². Dans différentes expériences indépendantes comparant PepNovo et d'autres algorithmes de séquençage *de novo*, PepNovo a donné de meilleurs résultats en terme de peptides correctement séquencés par rapport à plusieurs autres algorithmes de séquençage *de novo* ^{23, 24}.

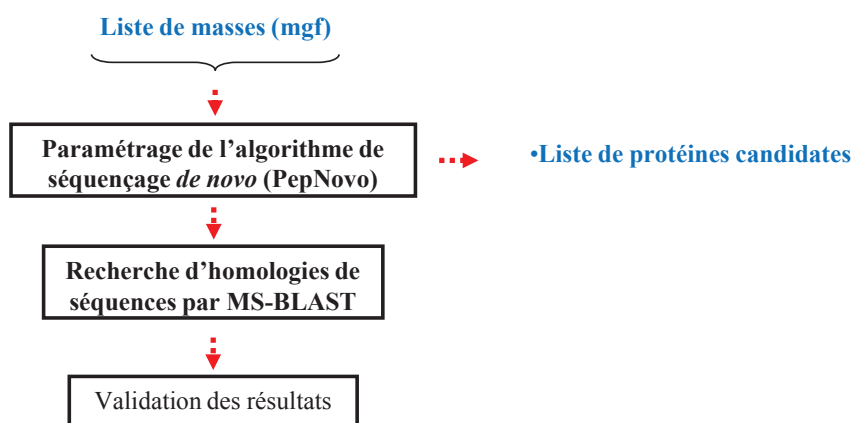
Pour évaluer le séquençage *de novo* par PepNovo, nous avons effectué un test de comparaison de séquençage *de novo* entre PEAKS et PepNovo sur des données LC-MS/MS du digeste de BSA analysé sur la trappe ionique amaZon (Bruker Daltonics). Nous avons sélectionnés 25 spectres MS/MS de bonne qualité avec des scores Mascot (ion score) supérieurs à 37. Les 25 spectres ont été soumis à un séquençage *de novo* automatisé par PEAKS et PepNovo avec les paramètres suivants : oxydation (M), carbamidométhylation (C), acétylation N-terminal, et les erreurs de masse ions parents et fils de 0,25 Da. Le résultat de cette comparaison est présenté dans le Tableau 3.

Comme résultats, 225 acides aminés contre 214 acides aminés sur 309 ont été correctement séquencés respectivement par PepNovo et PEAKS. Ce résultat montre que le logiciel PepNovo présente des performances supérieures à celles de PEAKS.

Tableau 3. Résultats de comparaison entre PepNovo et PEAKS : en rouge les séquences acides aminés (A. A) corrects de peptides de la BSA qui ont été identifiés par Mascot, PEAKS et PepNovo.

m/z (z)	Recherche Mascot		Séquençage <i>de novo</i> par PEAKS		Séquençage <i>de novo</i> par PepNovo	
	Peptides assignés	Score Mascot	Séquences identifiées	A. A corrects	Séquences identifiées	A. A corrects
435.9 (3+)	HLVDEPQNLIK	39	EFFKPNLLK	6/11	BXXXXXXEPQNLLK	7/11
461.7 (2+)	AEFVEVTK	39	AEFVEVTK	8/8	BAEFVEVTK	8/8
464.3 (2+)	YLYEIAR	39	YLYELAR	7/7	BYLYEXXR	5/7
480.6 (3+)	RHPEYAVSVLLR	62	WWMYAVSVLLR	8/12	BNHLLYAVSVLLR	8/12
493.9 (3+)	LGEYGFQNALIVR	61	WSLLANNYMLVR	3/13	BKEYGXXXNALLVR	6/13
547.3 (3+)	KVPQVSTPTLVEVSR	63	WCYLTPTLVEVSR	9/15	BXXXXXXLPTLVEVSR	7/15
569.8 (2+)	CCTESLVNR	51	CCTESLVNR	9/9	BDYTESLVNR	7/9
571.8 (2+)	KQTALVELLK	56	KQTALVELLK	10/10	BKQTALVELLK	10/10
582.3 (2+)	LVNELTEFAK	63	LVNELTEFAK	10/10	BLVNELTEFAK	10/10
625.3 (2+)	FKDLGEEHFK	50	QFDLGEEHFK	10/10	BKFDLGEEHFK	10/10
643.3 (3+)	CCAADDKEACFAVEGPK	73	MFAWDKEACGFAVEGPK	7/17	BXXXXADDXXXAQLHVEGPK	8/17
700.4 (2+)	TVMENFVAFVDK	71	SLMENFVAFVDK	10/12	BSLMENFVAFVDK	10/12
710.4 (2+)	SLHTLFGDELCK	67	SLHTLFGDELCK	12/12	BSLHTLFGDELCK	12/12
722.4 (2+)	YICDNQDTISSK	72	YLCDDDELTLSSK	12/12	BYLCDNQDTLSSK	12/12
739.9 (2+)	ETYGDMADCCCK	73	STYGDMADCCCK	11/12	BETYGDMADYXXXXR	8/12
751.9 (2+)	EYEATLEECCA	69	YEATLEECCA	10/12	BHLEATLESSAXXXXXK	5/12
756.4 (2+)	VPQVSTPTLVEVSR	71	VPQVSTACHVEVSR	14/14	BVPQVSTPTLVEWR	11/14
777.9 (2+)	DDPHACYSTVFDK	48	ERHACGSHHGFDK	3/13	BXXXXHACYSTVXXK	8/13
784.4 (2+)	DAFLGSFLYEYSR	76	WFLGSFLYEYSR	11/13	BWFLGSFLYEYDK	9/13
788.9 (2+)	LKPDPTLCLDEFK	53	KLPPDPTLCLDEFK	11/13	BLKPDPTLCLXXXEFK	12/13
831.4 (3+)	GLVLIAFSQYLQCPDEHVK	80	WRYWYCQKLYFFQANFR	3/21	BNLLLAFSQYLQ	10/21
862.9 (2+)	MPCTEDYLSLLNR	109	MPCTEDYLSLLNR	14/14	BXXXXCTEDYLSLLNR	12/14
874.4 (2+)	YNGVFQECQAEDK	65	YNGVMQECQAEDK	9/14	BYNGVFQEXXXTADK	11/14
886.5 (2+)	DDSPDLPK	37	DDSGVLHK	3/8	BDDSPDLPK	8/8
964.5 (2+)	CCAADDKEACFAVEGPK	79	CCAWDQEACGFAVEGPK	4/17	BXXXADDQEACFAVGLLK	11/17
Total			214 / 309		225 / 309	

Afin de faciliter l'utilisation de notre stratégie comprenant les 2 principales étapes: séquençage *de novo* et recherche d'homologies de séquences, nous avons implémenté PepNovo et MS-BLAST²⁵ l'algorithme de recherche d'homologies de séquences peptidiques, dans une suite logicielle mis en place au laboratoire (<https://msda.unistra.fr>). Ce schéma de fonctionnement a été mise en place pour que les peptides séquencés par PepNovo soient automatiquement pris en compte pour une recherche d'homologie par MS-BLAST en une seule étape. Le séquençage des peptides et la recherche MS-BLAST est ainsi entièrement automatisé contrairement au séquençage *de novo* manuel et la recherche BLAST peptide par peptide. De plus, une interface beaucoup plus conviviale a été développée pour le paramétrage de PepNovo qui à l'origine est programmé en ligne de commande. Le séquençage *de novo* PepNovo/MS-BLAST fonctionne selon le schéma de la figure 12.

**Figure 12.** Représentation schématique des étapes de fonctionnement de la plate forme de séquençage *de novo* PepNovo/MS-BLAST des données des analyses nanoLC-MS/MS.

Successivement, les 3 étapes de la figure 12 ont été suivies lors de cette recherche d'identification par homologies de séquences *de novo* PepNovo/MS-BLAST.

i. Paramétrage de l'algorithme de séquençage *de novo* PepNovo

Les **listes de masses** obtenues à partir des données d'analyse nanoLC-MS/MS sont traités par l'algorithme de séquençage *de novo* PepNovo avec les paramètres suivants:

- erreurs de masse sur les ions parents et ions fils: 0,25 Da pour les deux types d'ions ;
- l'enzyme utilisée: trypsine ;
- les possibles modifications: la carbamidomethyl (C), acétylation N-terminale de la protéine, oxydation (M) ;
- le mode de fragmentation: mode CID avec coupure enzymatique tryptique ;
- tous les ions des spectres MS/MS sont considérés pour le séquençage, même ceux dont l'état de charge n'a pas pu être déterminé au moment de la création de la liste de masse ;
- critère de score de qualité de spectre : 0,01 (valeur minimale). Cette valeur correspond à la valeur seuil minimale à partir de laquelle un spectre MS/MS est considéré suffisamment informatif (listes des ions fragments) donc de qualité acceptable pour un séquençage *de novo*.

NB : A cette précision de masse les acides aminés Q/K et I/L sont considérés comme isobariques.

Grâce ce dernier critère, les spectres MS/MS de mauvaise qualité (de faible rapport signal/bruit de fond (S/N), faible nombre d'ions fragment, etc.) sont automatiquement éliminés de la procédure du séquençage *de novo* et à la fin seuls 5 à 35% des spectres MS/MS par analyse donnent des peptides avec un score de qualité PepNovo supérieur à 0,01. Ce qui représente entre 50 et 300 spectres de qualité sur 1200 spectres MS/MS en moyenne que compte chaque analyse de spots de gel 2D. Au total 1609 peptides ont été séquencés pour les 30 spots de gel 2D.

ii. Recherche d'homologie de séquences par MS-BLAST

Les peptides issus du séquençage *de novo* ayant passés les critères spécifiés dans PepNovo sont automatiquement confrontés dans une recherche d'homologie de séquence MS-BLAST à une banque protéique (protéines de vigne et champignons, kératines et trypsine). L'alignement automatique de tous les peptides avec toutes les protéines de cette banque permet de, déterminer celles qui ont un meilleur score d'homologie de séquences "positive hits" (couverture de séquence peptidique > 50%) tandis que celles de faible scores "negative hits"(couverture de séquence peptidique <50%) ne seront pas considérés. Cet alignement de séquence va conduire à identifier pour chaque peptide issu du séquençage *de novo* les protéines d'organismes différents dans les quelles on retrouve des peptides quasi identiques à quelque acides aminés près.

La Figure 13 présente 8 des 220 protéines identifiées dont les peptides sont homologues au peptide BSLLMELLNNVAK séquencé par PepNovo avec les scores d'homologie supérieurs à 50 %. Ces protéines, bien que d'organismes différents, sont majoritairement des "ATP synthase". En effet, 81% de ces protéines qui ont été identifiées avec ce peptide par homologies de séquences sont des "ATP synthase" contre 19% qui sont protéines hypothétiques. La séquence d'acides aminés du peptide BSLLMELLNNVAK (Figure 13, "Query") est identique à quelques acides aminés près aux peptides homologues (Figure 13, "Sbjet"). Entre chaque peptide homologue et le peptide séquencé par *de novo* on retrouve les acides aminés communs entre les deux séquences. Les espaces vides entre les acides aminés signifient que les acides aminés respectifs au peptide l'homologue et au peptide séquencé par *de novo* sont complètement différents. Par contre un (+) désigne les acides aminés isobariques (I/L) ou (K/R) en amont d'un peptide tryptique.

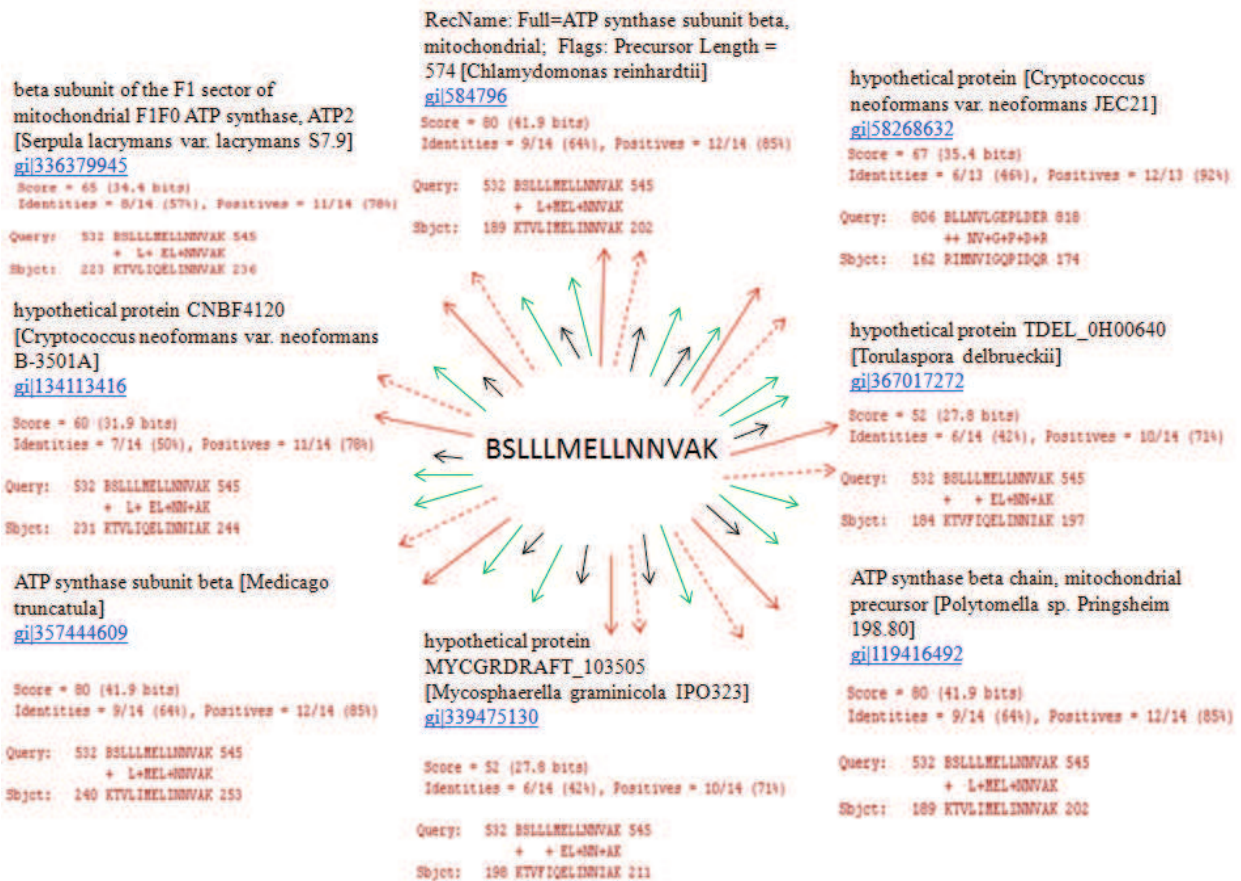


Figure 13. Représentation schématique de 8 des 220 possibilités d'homologie de séquence pour le peptide candidat BSLLMELLNNVAK obtenu par séquençage *de novo*. Ce peptide a conduit à l'identification de 220 protéines dont 162 "ATP synthase" de différents organismes avec une homologie de séquence peptidique > 50 %. Pour chaque protéine on retrouve l'alignement de séquence entre les peptides homologues des protéines identifiées et BSLLMELLNNVAK. Les protéines ont été identifiées avec un nombre de peptide compris entre 2 et 18 dont le peptide BSLLMELLNNVAK.

Dans le cas de notre étude, la même banque (composée des séquences protéiques de *vitis* et *fungi*, kératines humaines, trypsine et copies des séquences inversées) que celle utilisée lors de la recherche conventionnelle, a été utilisée.

iii. Validation des résultats de recherche PepNovo/MS-BLAST

A la fin de la recherche PepNovo/MS-BLAST, une liste de centaines de protéines (protéines de champignons et protéines de vigne, trypsine ou kératines), dont les peptides sont homologues à ceux séquencés par PepNovo est automatiquement extraite de la banque de données protéiques utilisée pour la recherche d'homologie par MS-BLAST. Ainsi tous les peptides candidats séquencés par *de novo* et identifiés correspondant à chaque protéine par analyse se retrouvent répertoriés dans une feuille de résultats EXCEL beaucoup plus facile à manipuler vu la quantité de protéines homologues identifiées. En effet, 364916 lignes EXCEL correspondant à 16003 protéines ont été obtenues après la recherche par homologie de séquences sur les 117 spots de gels 2D analysés. Ce qui constitue une quantité très importante de données qui ne sont pas prises en charge par la plupart des outils de gestion de données d'interprétation de résultats de spectrométrie de masse comme Scaffold par exemple.

On retrouve dans cette feuille de résultats pour chaque analyse, les protéines identifiées avec leur numéro d'accèsion (gi) NCBI, leur annotation fonctionnelle, leurs poids moléculaires (Mr), leurs scores MS-BLAST totaux (somme des couvertures de séquence peptidique des "positive hits" identifiés pour chaque protéine), l'espèce à laquelle appartient chaque protéine identifiée, le nombre de peptides identifiés pour chaque protéine, les séquences peptidiques correspondantes, et le numéro de spectre qui a permis de séquencer chaque peptide par analyse (Figure 14).

N° d'Analyse	GI	Nom de la protéine	Organism	Mr	Score MS Blast	Nbr Peptide	Séquence	Cmpd
F028316TA	gi 359496362	PREDICTED: uncharact	[Vitis vinifera]	18980	421	7	VTTTGSFVPCA	222
F028316TA	gi 359496362	PREDICTED: uncharact	[Vitis vinifera]	18980	421	7	LFEPGHN	143
F028316TA	gi 359496362	PREDICTED: uncharact	[Vitis vinifera]	18980	421	7	PACAL	278
F028316TA	gi 359496362	PREDICTED: uncharact	[Vitis vinifera]	18980	421	7	LATDLSPFELTALR	929
F028316TA	gi 359496362	PREDICTED: uncharact	[Vitis vinifera]	18980	421	7	EVTTTGPHV	276
F028316TA	gi 359496362	PREDICTED: uncharact	[Vitis vinifera]	18980	421	7	BGNLDFSGR	355
F028316TA	gi 359496362	PREDICTED: uncharact	[Vitis vinifera]	18980	421	7	BGGTDSLVSRLR	250
F028316TA	gi 136429	Trypsin precursor	[Sus scrofa]	24410	263	4	BVATVS	225
F028316TA	gi 136429	Trypsin precursor	[Sus scrofa]	24410	263	4	FNGNTLDNDL	627
F028316TA	gi 136429	Trypsin precursor	[Sus scrofa]	24410	263	4	BLSSPATLNSR	135
F028316TA	gi 136429	Trypsin precursor	[Sus scrofa]	24410	263	4	GNTLDNDLMLLK	618

Figure 14. Aperçu de tableau de résultats EXCEL de la recherche par homologie de séquences PepNovo/MS-BLAST

La principale difficulté face à cette très grande quantité de données va être de parvenir à distinguer entre toutes ces centaines de protéines, celles à valider en déterminant pour chaque d'elles les peptides qui lui sont spécifique spécifiques. En effet les peptides spécifiques d'une protéine sont les peptides qui sont uniquement attribués à cette protéine et qui permettent donc d'identifier spécifiquement une protéine. Pour valider donc les protéines identifiées par homologie de séquence, seules les protéines ayant des peptides spécifiques identifiés ont été considérées dans la liste des protéines identifiées avec un score MS-BLAST supérieur à 50 %. Dans la pratique, cela revient donc à exclure de la validation, les peptides communs (partagés) à la trypsine, la kératine, les protéines de champignon et les protéines de vigne pour obtenir les peptides spécifiques à chaque protéine par analyse.

Ainsi PepNovo a permis de séquencer au total 1609 peptides (pour les 30 spots de gels 2D) dont plus de la moitié (62%) correspond à des protéines de vigne (PV) et 38% aux protéines d'autres organismes (PnV). Parmi les 62 % de peptides attribués aux protéines de vigne, 45% sont des peptides qui leur sont spécifiques (PVs) (210 protéine uniques) et 17% sont des peptides partagés avec des protéines d'autres espèces (PVP) (Figure 15).

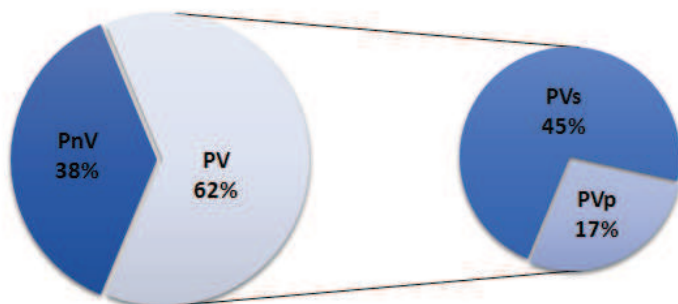


Figure 15. Diagramme de répartition de peptides séquencés par PepNovo : 62% des peptides séquencés correspondent aux peptides de protéines de vigne (PV) et sont répartis en peptides spécifiques aux protéines de vigne (PVs, 45%) et en peptides partagés (PVP, 17%) avec d'autres protéines d'autres espèces. PnV, représente les 38% des peptides appartenant aux protéines d'autres organismes.

La recherche de peptides spécifiquement attribués aux protéines de champignons a permis de mettre en évidence 148 peptides spécifiques (PCs), ce qui représente 9% de tous les peptides séquencés.

Pour identifier les protéines de champignons, les peptides spécifiques aux protéines fongiques ont été validés avec les critères supplémentaires suivants:

- posséder au moins 5 acides aminés consécutifs séquencés par PepNovo ;
- posséder un maximum d'acides aminés consécutifs non séquencés (X) de 4 ;
- la protéine retenue dans la liste finale après validation des peptides spécifiques est celle ayant le score MS-BLAST le plus élevé parmi tous les homologues.

Les critères supplémentaires ont permis de valider 70% (PCsv) des peptides spécifiques aux protéines de champignons (PCs). Les peptides spécifiques de champignons validés (PCsv) sont repartis en peptides partagés entre protéines de champignons (PCsvp, 43%) (42 protéines uniques), et en peptides non partagés (PCsvnp, 27%), c'est-à-dire des peptides appartenant uniquement à une protéine de champignons par analyse considérée (Figure 16).

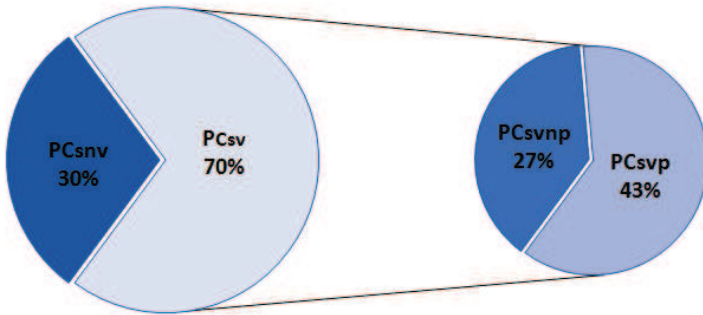


Figure 16. Diagramme de répartition de peptides séquencés par PepNovo : 70% des peptides spécifiques aux protéines fongiques (PCs) ont été validés et sont répartis en peptides spécifiques validés non partagés PCsvnp (43%) et 27% des peptides partagés validés partagés (PVp) entre protéines de vigne et protéines d'autres espèces (champignons, kératine et trypsine). PCsv représente les 30% des peptides spécifiques aux protéines de champignons qui n'ont pas été validé à cause de leurs séquences peptidiques inférieures à 5 acides aminés consécutifs.

A titre d'exemple, le peptide LTSGFYLGELVR ($m/z = 677$ (2+)) séquencé par *de novo* a été identifié comme peptide spécifique aux protéines fongiques de type hexokinase, alors que dans la même analyse, son homologue LTSGMYLGELVR est commun à l'hexokinase de vigne et de champignons. En effet les peptides LTSGMYLGELVR et TISGMYLGELVR existent dans les séquences protéiques d'héxokinase dans la banque et ont été identifiés lors de la recherche conventionnelle à partir d'un même spectre avec un bon score de qualité "Mascot ion score" de 61 (Figure 17).

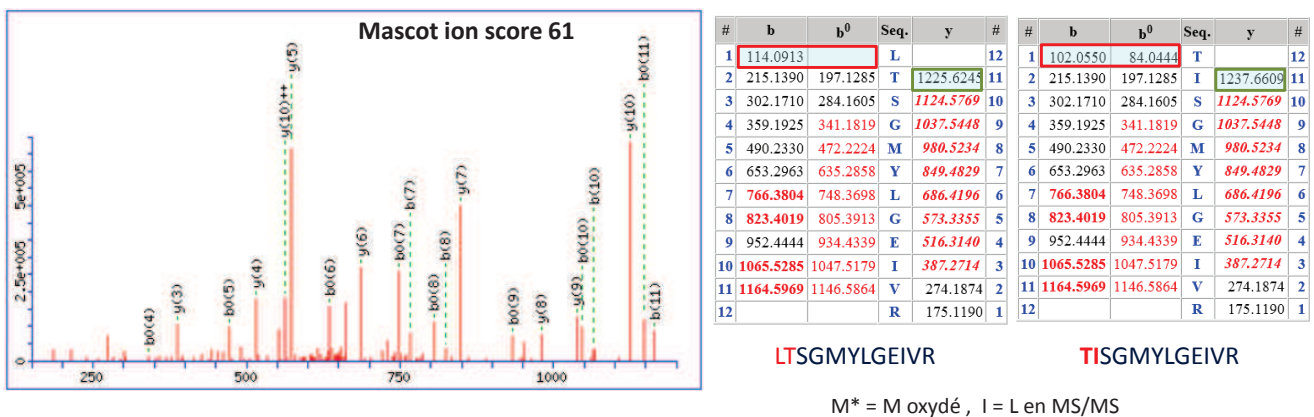


Figure 17. Double interprétation d'un spectre MS/MS. Le spectre a donné deux interprétations peptidiques (TISGMYLGEIVR et LTSGMYLGEIV) lors de la recherche dans la banque alors le séquençage *de novo* propose une interprétation quasi identique (LTSGMYLGEIV). b et y sont les ions fragments et b⁰ sont les ions b-H₂O.

Deux interprétations peptidiques de ce spectre MS/MS ont été possibles puisque les deux peptides (LTSGMYLGEIVR et TISGMYLGEIVR) ont le même $m/z = 669,8$ (2+) et ont les mêmes ions fragments à l'exception des ions b1, b2 et y11. L'identification des trois ions sur le spectre permettrait de distinguer entre les deux peptides. Or ce n'est pas toujours le cas. En effet le côté N-terminale de peptides se fragmentent moins bien que le côté C-terminal à cause de la charge prépondérante que porte l'amine libre radicale de l'arginine (R) et lysine (K) C-terminale du peptide tryptique.

Pour vérifier dans un premier temps si le peptide LTSGFYLGELVR n'est pas un faux positif du peptide LTSGM*YLGEIVR ou TISGM*YLGEIVR (M*, M oxydé), et dans un second temps avoir une idée sur le score de qualité "Mascot ion score" du spectre qui a permis de séquencer le peptide nous avons fait une simulation d'interprétation des données MS/MS avec le moteur de recherche Mascot. Ainsi une nouvelle recherche conventionnelle a été refaite dans une banque protéique quasi identique à celle utilisée lors de la recherche conventionnelle initiale avec comme seule différence, le remplacement de M par F dans les séquences normale et inverse de l'hexokinase de vigne. Comme le montre la figure 18, le peptide LTSGFYLGELVR a bien été identifié avec plus de fragments que LTSGM*YLGEIVR ou TISGM*YLGEIVR avec un score Mascot de 49 largement au dessus du score minimum 30 utilisé lors de la validation des résultats de la recherche conventionnelle initiale.

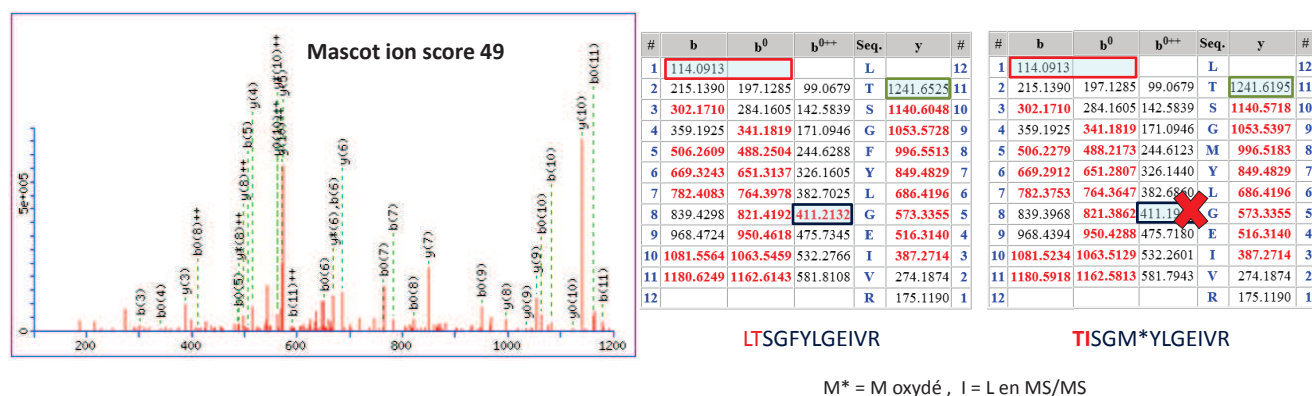


Figure 18. Vérification des résultats du séquençage *de novo* comparé à la recherche conventionnelle. Le séquençage *de novo* distingue entre faux positifs (TISGM*YLGEIVR, M *oxydé) et séquence peptidique plus probable (LTSGFYLGELVR) en identifiant plus d'ions corrects sur le spectre MS/MS. Les ions qui ont été identifiés sont marqués en rouge dans les tableaux. b et y sont les ions fragments et b⁰ sont les ions b-H₂O.

Ce résultat montre bien que le peptide LTSGFYLGELVR n'est pas un faux positif. Ce peptide n'a pas été identifié lors de la recherche conventionnelle parce qu'il n'était pas dans la banque protéique utilisée. Ce peptide (LTSGFLGELVR) pourrait potentiellement donc appartenir à un variant d'une hexokinase de vigne ou à une hexokinase de champignon non séquencé.

3. Résultats

a. Résultats de la recherche conventionnelle dans la banque protéiques avec les algorithmes de recherche Mascot et OMSSA.

La recherche conventionnelle a permis d'identifier 164 protéines uniques dans les 30 spots de gel 2D différentiels. Parmi les 164 protéines, une seule protéine d'espèce fongique non pathogène a été identifiée. Aucune protéine de *Fomitiporia mediterranea*, la seule espèce fongique de génome partiellement séquencé n'a été identifiée. Grâce à la recherche d'annotations fonctionnelles par Gene Ontology, les protéines de vigne ont pu être réparties en trois groupes de protéines suivants (Figure 19):

- (i) protéines impliquées dans le métabolisme et l'énergie (ME) ;
- (ii) protéines impliquées dans la tolérance aux stress (TS) ;
- (iii) protéines imputables à la réponse de défense de la vigne (RD).

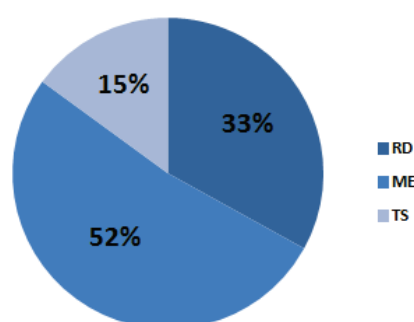


Figure 19. Pourcentages de protéines identifiées réparties entre les trois groupes. (ME) Métabolisme et Energie, (RD) Réponse de Défense, (TS) Tolérance au Stress.

Les différentes fonctions de ces protéines ont été discutées dans l'article dans l'annexe 5 à la fin de ce chapitre.

b. Résultats de la recherche par séquençage *de novo* PepNovo/MS-BLAST

Grâce à l'approche d'identification basée sur le séquençage *de novo* PepNovo/MS-BLAST, 120 protéines sur 210 protéines de vigne identifiées ont été validées et 42 protéines (Tableau 4) sur 94 protéines de champignons identifiées ont pu être validées. Près de 50% des protéines identifiées par la recherche *de novo* PepNovo/MS-BLAST n'ont pas été validées à cause du manque de spécificité des peptides (par exemple, les peptides de séquences < 5 acides aminés ou contenant plus de 4 (X)). L'interprétation des spectres par séquençage *de novo* (PepNovo) donne des tags parfois très courts en nombre d'acides aminés à cause peptide qui se fragmentent par fois mal et qui donnent des spectres de fragmentations souvent incomplets. Ces spectres conduisent par conséquent à de mauvais scores MS-BLAST (<50) lors de la recherche d'homologies et donc à la non validation de ces résultats.

Parmi les protéines de champignons identifiées, près de la moitié n'a pas de fonction connue. Le tableau 3 récapitule toutes les protéines de champignons identifiées par *de novo* PepNovo/MS-BLAST. Par exemple, la protéine beta-tubulin, partial (gi|63333445) a été identifiée avec un score MS-BLAST de 313 avec 7 peptides issus du séquençage *de novo*. Parmi les 7 peptides obtenus par séquençage *de novo*, 6 sont partagés avec d'autres protéines d'espèces différentes, et un peptide spécifique unique à cette protéine a permis de valider la protéine identifiée et non ses homologues (Figure 20). Le peptide spécifique unique à cette protéine a été identifiée avec un score d'homologie de 55 avec le peptide homologue de la beta-tubuline soit 6 acides aminés sur 11 acides aminés séquencés.

Tableau 4. Liste des protéines fongiques identifiées dans les troncs de vignes de la variété Chardonnay par séquençage *de novo* PepNovo /MS-BLAST. A chaque protéine sont associés : le numéro d'accèsion de celle-ci dans la banque NCBI, le score d'homologie MS-BLAST, le nombre total de peptides identifiés et la séquence des peptides spécifiques validés manuellement. (B) acide aminés préfixés potentiellement pouvant correspondre à un site de clivage (K/R). X acides aminés non séquencés.

Protéines fongiques identifiées	N° d'accèsion NCBI	Score MS- Blast	Nombre total de peptides	Séquence des peptides spécifiques uniques validés
Protéines de fonction connues				
alpha-tubulin	gi 30088536	1094	22	ALAVAEEV-APVLATSR
alpha-tubulin, partial	gi 1755092	798	17	PTVLDD
ATP synthase F1-sector subunit beta	gi 398407561	334	7	BNLVADLE
beta-tubulin, partial	gi 63333445	313	7	DNTAFAEVFSR
Chain A, Multiple Distinct Assemblies Reveal Conformational Flexibility In The Small Heat Shock Protein Hsp26	gi 112491087	131	2	BXXXQDGVLTVTVEK
chromatin binding protein	gi 299743222	116	2	DEDLGTYD
DEHA2A08646p	gi 294654645	113	2	BXXXXTLVYGD
DEHA2C07326p	gi 294655806	255	5	SNNTAFAEVFDK
D-P-binding protein	gi 392558542	147	3	TPASSLHA
flavodoxin, putative	gi 320580376	110	2	GDLYHK
Glucokinase	gi 315055849	244	5	SPATLNDK
Glucokinase	gi 15072475	247	5	LXXXXLELLALDE
heat shock protein 22	gi 88766401	384	7	BVLDVELK
heat shock protein 22	gi 88766401	510	10	BDGVLXXSLPK
Hexokinase	gi 150866927	200	4	QELVLQT
methionine adenosyltransferase	gi 401886856	540	10	BNLVDEDNVL
methionine adenosyltransferase	gi 388581252	460	8	BXXXXPADALD
NAP family protein	gi 326480535	108	2	DDLGDGEE
predicted protein	gi 340520731	522	10	LNLNQALDEL
probable translation initiation factor eIF-4A, exon junction complex	gi 353244300	419	8	VTLQDV
60S ribosomal protein L10	gi 51701820	108	2	FSGAFDALNQS
ribosomal protein L12	gi 387597172	198	4	GEVTLA
RNA-binding domain-containing protein	gi 389748310	439	8	LSSPATLND
S-adenosylmethionine synthetase	gi 148469556	397	8	VFVFGELTTK
Protéines de fonction inconnues				
hypothetical protein	gi 336464269	419	8	HVALVK
hypothetical protein	gi 392580023	486	9	WSSDPGXSSSYK
hypothetical protein	gi 67537824	250	5	SSLTSL
hypothetical protein	gi 354546054	253	5	DVLDLVQ
hypothetical protein	gi 402217506	298	6	NDLMLVR
hypothetical protein	gi 358060929	250	5	LTDNDLMLVR-LQTQLQAA
hypothetical protein	gi 328848960	247	5	BXXXAFAEVFDK
hypothetical protein	gi 389746976	203	4	ALLDEAP
hypothetical protein	gi 345565802	114	2	YEDPQD
hypothetical protein	gi 328768374	114	2	DLNDLML
hypothetical protein	gi 342879144	295	5	PGDLGYER
hypothetical protein	gi 302697115	248	5	EGMDGQN
hypothetical protein	gi 302680130	120	2	TPQSSAATLPK
hypothetical protein	gi 392578419	542	10	BNLTVNEAN
hypothetical protein	gi 270056447	194	4	BQAYEDLXXXX
hypothetical protein	gi 387511448	512	10	DEDLLR
hypothetical protein	gi 169608159	112	2	AATLLATS
hypothetical protein	gi 358057430	190	4	LHVDALDE
conserved hypothetical protein	gi 212542765	183	4	DLDSEV

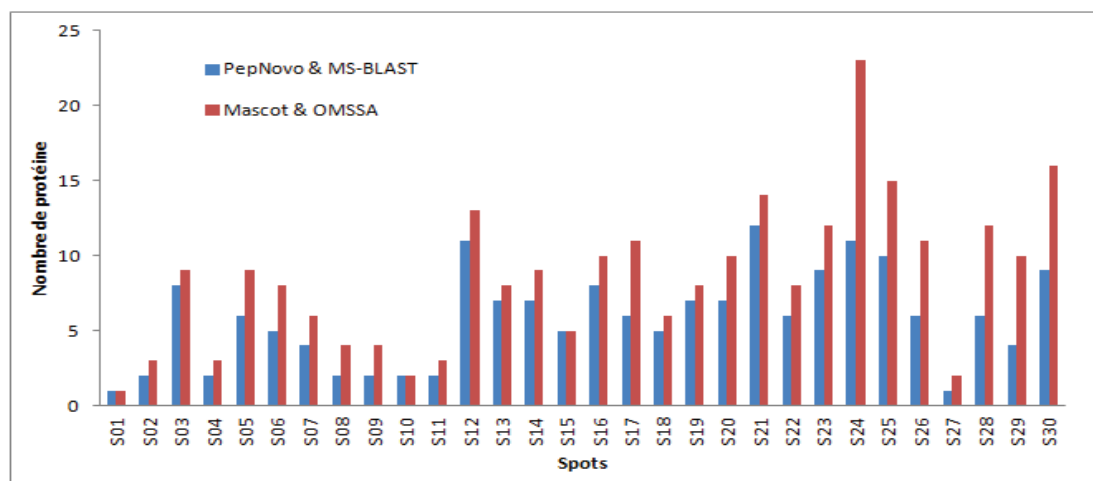


Figure 21. Comparaison du nombre de protéines de vigne identifiées par les deux approches sur les troncs de vignes de la variété Chardonnay. En bleu approche par séquençage *de novo* PepNovo/MS-BLAST, et rouge identification par recherche dans les banques de données.

Au total, près de 75% de protéines uniques de vigne ont été confirmées par l'approche *de novo* PepNovo/MS-BLAST (Figure 22). A l'inverse, seulement 2% des protéines de champignons ont pu être identifiées par l'approche conventionnelle (Figure 22).

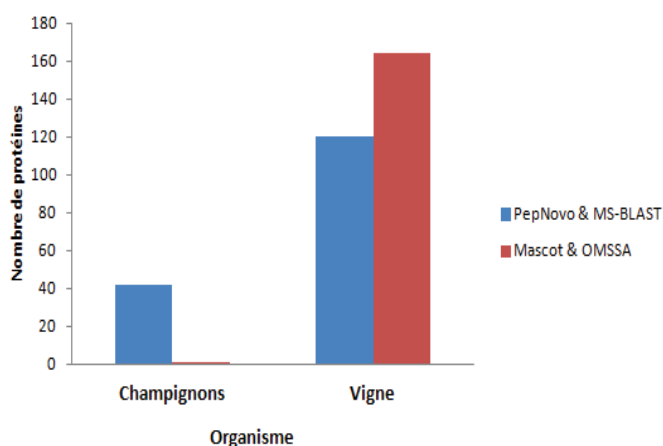


Figure 22. Comparaison globale de protéines de champignons et de vignes identifiées par les deux approches d'identifications dans les troncs de vigne de la variété chardonnay.

4. Application des deux stratégies d'identification de protéines à d'autres variétés de vignes infectées par la maladie de l'esca

Les mêmes expériences d'analyse protéomique basée sur une approche gel 2D ont été réalisées sur des troncs de deux autres variétés de vignes (Mourvèdre et le Gewurztraminer) infectées par la maladie de l'esca. Au total 27 et 24 spots se sont révélés différentiels respectivement pour les troncs de vignes Mourvèdre et les troncs de vigne Gewurztraminer.

a. Résultats des troncs de vigne de la variété Mourvèdre

Au total, 77 protéines uniques de vigne ont pu être identifiées par l'approche de recherche conventionnelle. L'application de l'approche de recherche par séquençage *de novo* PepNovo/MS-BLAST a permis de confirmer 60 de ces protéines de vigne et d'identifier 28 protéines de champignons (Tableau 5).

Dans chaque spot, 40 à 100% des protéines de vigne identifiées par recherche conventionnelle ont été confirmées lors de l'identification par l'approche *de novo* PepNovo/MS-BLAST. Le tiers des protéines fongiques identifiées sont des protéines de fonction inconnue. La majorité des protéines de fonction connue sont des protéines homologues d'énolase, de malate déshydrogénase qui sont respectivement des enzymes de glycolyse et de catalyse oxydative.

Tableau 5. Liste des protéines fongiques identifiées dans les troncs de vigne de la variété Mourvèdre par séquençage *de novo* PepNovo/MS-BLAST. A chaque protéine sont associés : le numéro d'accèsion de celle-ci dans la banque NCBI, le score d'homologie MS-BLAST, le nombre total de peptides identifiés et la séquence des peptides spécifiques validés manuellement. (B) acide aminés préfixés potentiellement pouvant correspondre à un site de clivage (K/R). X acides aminés non séquencés.

Protéines fongiques identifiées	N° d'accèsion NCBI	Score MS-Blast	Nombre total de peptides	Séquence des peptides uniques validés
Protéines de fonction connues				
Enolase	gi 169845435	793	17	GSAAVYQA
Enolase	gi 402222471	567	12	VVSVDSA
Enolase	gi 331241667	834	17	LLSSGL
Enolase	gi 331241667	262	6	VALSQTY
Enolase	gi 384488581	912	18	LLXXXLLVTNPK
Enolase	gi 535442	736	15	BVLEXXXVTESLEAV
enolase 1	gi 146415384	800	16	QLGSVTESL
Enolase I, a phosphopyruvate hydratase that catalyzes the conversion of 2-phosphoglycerate to phosph	gi 254570367	639	14	ALDTLLNK
heat shock protein 22	gi 88766401	405	7	PFGLLDPFBS-BLALDLXXXXLLDP
malate dehydrogenase, mitochondrial precursor	gi 255725576	296	6	BDYTSLNHR
malate dehydrogenase, mitochondrial precursor	gi 116191823	113	2	AALDGLK
malate dehydrogenase, mitochondrial precursor	gi 149236698	283	5	DPVLEH
malate dehydrogenase, NAD-dependent	gi 212546203	344	7	ETTLDVV
mating locus 1-1	gi 295663064	74	1	VGNEVSALQQAF
2-phosphoglycerate dehydratase 1	gi 74659376	584	12	BXXXXXALLAVSLAVCK
2-phosphoglycerate dehydratase 1-2	gi 29839366	459	10	LGFVVAASEFYDNK
2-phosphoglycerate dehydratase 2	gi 74659032	863	18	LLAVSLST
2-phosphoglycerate dehydratase	gi 74659668	657	14	DVAASEFNSE
TPA: malate dehydrogenase (Eurofung)	gi 259480305	309	6	EPSESAL
translation elongation factor 1 alpha, partial	gi 170672809	240	5	PLGSASF
ZYRO0C11814p	gi 238939049	723	15	LALGAEV
Protéines de fonction inconnues				
hypothetical protein	gi 50293181	279	6	ALLSGFA
hypothetical protein	gi 345565802	196	4	QEVEXXXEVGHR-HQEADXXDDE
hypothetical protein	gi 328771058	291	6	TLSQFLNAAK
hypothetical protein	gi 328770238	247	5	GTGSELELA
hypothetical protein	gi 358057804	384	7	LVDAAF
hypothetical protein	gi 372463423	412	8	DLELAQ
hypothetical protein	gi 384500173	207	4	STLDDLDXXLK-YEECLK -BCLELDPTF
hypothetical protein	gi 387512298	234	5	VFGVTTLDXR

b. Résultats des troncs de vigne de la variété Gewurztraminer

Au total, 101 protéines uniques de vigne et 6 protéines de champignons ont pu être identifiées par la recherche conventionnelle. L'approche d'identification par séquençage *de novo* PepNovo/MS-BLAST a permis de confirmer 84 des protéines de vigne et d'identifier 34 protéines de champignons.

Dans chaque spot, 50-100% des protéines de vigne identifiées par l'approche conventionnelle ont été confirmées par l'approche *de novo* PepNovo/MS-BLAST. Près d'un tiers des protéines fongiques identifiées n'ont pas de fonction connue (Tableau 6). Parmi les 2 autres tiers, on distingue majoritairement, les tubilines, les adenosine kinase, transketolase qui sont respectivement des protéines du cytosquelette et de surface d'échange intracellulaire et, des protéines qui catalysent la production d'énergie sous forme d'ATP ou qui catalysent la voie des pentoses phosphates important pour la production d'énergie.

Tableau 6. Liste des protéines fongiques identifiées dans les troncs de vigne de la variété Mourvèdre par séquençage *de novo* PepNovo/MS-BLAST. A chaque protéine sont associés : le numéro d'accèsion de celle-ci dans la banque NCBI, le score d'homologie MS-BLAST, le nombre total de peptides identifiés et la séquence des peptides spécifiques validés manuellement. (B) acide aminés préfixes potentiellement pouvant correspondre à un site de clivage (K/R). X acides aminés non séquencés.

Protéines fongiques identifiées	N° d'accèsion NCBI	Score MS-Blast	Nombre total de peptides	Séquence des peptides uniques validés
Protéines de fonction connues				
14-3-3 family protein epsilon	gi 384485987	591	11	TSDFQDDS
14-3-3 protein	gi 346322623	466	9	AVEQVSL
20S proteasome subunit	gi 401886830	197	4	GGTASLGLK -LVYSGFGPDSR
adenosine kinase	gi 302694619	270	6	BXXXXENGELVD
adenosine kinase	gi 346971325	383	8	LSLLDDLLT
alpha-tubulin	gi 30088546	957	20	EPTVLGSGA
alpha-tubulin, partial	gi 283139119	728	14	BPLYTNLNR
beta tubulin	gi 6424969	602	13	NTYVVVGNP
beta-tubulin	gi 299750193	464	10	QFVDDANL -YTEAVSSTNSL
beta-tubulin 2, partial	gi 6934203	713	15	YVLSLTLASLR
beta-tubulin, partial	gi 383385970	281	6	LPALTAG
beta-tubulin, partial	gi 63333553	436	10	LENELGATY
beta-tubulin, partial	gi 986952	270	6	SAVAYGXXXXLR
DEHA2C07326p	gi 294655806	274	6	DGSYVFN
glyceraldehyde 3-phosphate dehydrogenase	gi 344234248	352	7	VSAHEASV
probable adenosine kinase	gi 323508353	428	9	LTXXXDVPVVVAEDGK
probable tubulin beta chain	gi 388856679	421	9	BAPSGFAE
putative S-adenosylmethionine synthase	gi 361130667	858	16	DELADALK
Beta-2-tubulin	gi 1174597	482	10	EADNDLEN
S-adenosylmethionine synthetase	gi 148469556	807	17	BNLGLDLSLY
Transketolase	gi 384489750	319	6	BNLSEQCLNALA
Transketolase	gi 384489750	344	7	BNLSEQCLNALA
Transketolase	gi 213408571	370	8	BXXXXSVHGSALGA
tubulin gamma chain	gi 388582054	232	5	LSSLTASL
YALI0E26961p	gi 50554075	446	10	GFEEGEFSE
Protéines de fonction inconnues				
hypothetical protein	gi 345566903	239	5	GAESAYASEQ
hypothetical protein	gi 238879720	196	4	DAVGLAAT
hypothetical protein	gi 328849984	510	10	BXXXLTQGAPVDVVA
hypothetical protein	gi 164657866	1041	19	AFSGHFGR
hypothetical protein	gi 366994920	214	4	DDFGVQLT
hypothetical protein	gi 330928435	229	5	EGVLVGS
hypothetical protein	gi 302685025	338	8	EEVNSVVEAXXVE
hypothetical protein	gi 169602531	962	19	BEDAADNFV
hypothetical protein	gi 392580099	1052	18	BNDVTVEYXXAP
Hypothetical protein	gi 358389942	149	3	EDLSNDDSR

Conclusion

La combinaison de stratégies d'analyse protéomique basées sur les gels 2D et l'identification des protéines par combinaison de la recherche conventionnelle et séquençage *de novo* PepNovo/MS-BLAST, a permis d'identifier les protéines différentiellement exprimées dans les troncs de vigne infectés par l'esca.

La recherche par séquençage *de novo* PepNovo/MS-BLAST a permis de confirmer environ 80% des protéines identifiées par la recherche conventionnelle. Parmi les protéines de vignes identifiées par analyse protéomique, quelques une ont été suivies par RT-PCR.

Sur l'ensemble des protéines de champignons identifiées, moins de 10% l'ont été par la recherche conventionnelle tandis que la recherche par séquençage *de novo* PepNovo/MS-BLAST a permis de mettre en évidence plus de 90 % des protéines de champignons malgré l'absence des génomes séquencés de ces champignons.

Annexe 5: Article soumis pour publication

Proteomic insights into changes in trunk wood of *Vitis vinifera* L. cv Chardonnay in response to esca proper and apoplexy.

Magnin-Robert; M. ; Spagnolo, A.; **Alayi, T. D.** ; Cilindre, C. ; Mercier, L. ; Schaeffer-Reiss, C. ; Van Dorsselaer, A. ; Clement, C. ; Fontaine, F.

Phytopathologia Mediterranea

1 **Proteomic insights into changes in trunk wood of *Vitis vinifera* L. cv Chardonnay in response to esca proper**
2 **and apoplexy**

3
4 **Maryline Magnin-Robert^{1*}, Alessandro Spagnolo^{1*}, Tchilabalo Dilezitoko Alayi², Clara Cilindre^{1,3},**
5 **Laurence Mercier⁴, Christine Schaeffer-Reiss², Alain Van Dorsselaer², Christophe Clément¹ and Florence**
6 **Fontaine^{1,**}**

7
8 * Co-first authors.

9 ** To whom the correspondence should be addressed. Florence FONTAINE, E-mail: florence.fontaine@univ-
10 reims.fr, Phone: +3 33 26 91 33 18, Fax: +3 33 26 91 33 39

11 ¹ *Université de Reims Champagne-Ardenne, URVVC EA 4707, Laboratoire Stress, Défenses et Reproduction*
12 *des Plantes, BP 1039, 51687 Reims Cedex 2, France.*

13 ² *Université de Strasbourg, IPHC, UMR 7178, Laboratoire de Spectrométrie de Masse Bioorganique, 67087*
14 *Strasbourg, France.*

15 ³ *Université de Reims Champagne-Ardenne, URVVC EA 4707, Laboratoire d'Œnologie et Chimie Appliquée,*
16 *BP 1039, 51687 Reims Cedex 2, France.*

17 ⁴ *Moët & Chandon, 6 rue Croix de Bussy, 51200 Epernay, France.*
18

19 **Abstract**

20

21 Among grapevine trunk diseases, esca proper (E) and apoplexy (A), pose a significant threat for viticulture. Two
22 dimensional gel electrophoresis coupled to mass spectrometry analysis was used, for the first time, to examine the
23 global changes associated with woody tissues of E and A field-grown vines. Asymptomatic and black streaked
24 trunk (BST) wood from E and A plants were compared to asymptomatic and BST wood of control plants (C).
25 Thirty differentially expressed protein spots were identified. For eleven targeted proteins, expression of the relative
26 transcript was also monitored by qRT-PCR. Hierarchical tree clustering revealed differences in the distribution of
27 spots containing carbohydrate metabolism proteins and heat shock proteins between asymptomatic- and BST-wood
28 of grapevine, irrespective of the type of plant (control or diseased grapevine). Asymptomatic wood was mainly
29 characterized by down-expression of proteins involved in cell growth and defense response. The proteome of BST
30 wood, characterized by an extensive presence of grapevine trunk disease agents, revealed over-expression of
31 proteins involved in defense. Proteome variations related to the nature of samples, i.e., asymptomatic or BST wood,
32 were more pronounced than those related to the expressed foliar symptom form, suggesting the involvement of
33 different signals between the two types of samples.

34

35 **Key-words:** grapevine, proteomics, qRT-PCR, trunk diseases.

36

37

38

39

40 **Introduction**

41 Grapevine trunk diseases have increased in incidence over the last 15 years¹ and now pose a serious problem for
42 viticulture worldwide.² They are caused by fungal pathogens that attack the woody perennial organs of the vine and
43 ultimately lead to the death of the plant. The predominant grapevine trunk diseases are eutypa dieback,
44 botryosphaeria dieback and esca disease complex. For the latter, five syndromes have been described: brown
45 streaking of the wood (mostly affecting rooted cuttings), Petri disease, grapevine leaf stripe disease (GLSD), esca
46 and esca proper.^{3,4} GLSD is a tracheomycotic widespread syndrome for which the major causal agents are
47 *Phaeoconiella chlamydospora* (W. Gams, Crous, M.J. Wingfield & L. Mugnai) P.W. Crous & W. Gams (Pch) and
48 *Phaeoacremonium aleophilum* W. Gams, P.W. Crous, M.J. Wingfield & L. Mugnai (Pal).³ Esca is a typical white
49 decay occurring in the trunk and branches of standing vines and for which the causal agents are different
50 Basidiomycetes species represented mainly by *Fomitiporia mediterranea* (Fom) M. Fischer in Europe and the
51 Mediterranean basin.^{3,4} The coexistence of GLSD and esca on the same plant is called esca proper.⁴ Esca proper
52 affected plants are thus characterized by wood symptoms comprising several forms of discoloration, among which
53 black streaking involving single or several xylem vessels and areas with darkened or brown necrosis
54 circumscribing the pith are most commonly observed.

55 Additionally, symptoms on the leaves can emerge. These are typically characterized by spots appearing between
56 the leaf veins or along the edges, which expand and become confluent, finally resulting in chlorotic and necrotic
57 strips with only a narrow green stripe along the midrib. In most cases, the affected leaf finally assumes a “tiger
58 stripe” appearance. Symptoms on the berries include the appearance of small brown or purple spots particularly on
59 table grapes.³ Apoplexy, consisting of the partial or complete sudden wilting of the crown, is regarded as the severe
60 form of GLSD and/or esca proper.^{3,5} Thus, affected plants die within a few years.⁶ To understand the mechanisms
61 of these decline-associated trunk diseases and their symptoms further, several studies have been conducted on
62 related grapevine responses. A perturbation of photosynthesis revealed by a decrease of both gas exchange and
63 photosystem II activity has been reported in leaves of apoplectic and esca proper affected plants.^{5,7,8} Photosynthesis
64 disruption may originate from damaged intracellular structures.⁹ In the leaves of apoplectic grapevines, drastic
65 physiological alterations of photosynthesis are coincident with a decline in water use efficiency and an activation of
66 stress responses within at least seven days preceding the appearance of visible symptoms.⁵ To obtain further
67 information about the mechanisms activated in apoplectic and esca proper affected plants, Spagnolo *et al.*¹⁰ focused
68 on the proteome variations of their green stems. Their results indicated that similar responses are likely activated in
69 asymptomatic stems, but various quantitative expression is triggered upon the onset of apoplexy or esca proper
70 symptoms, while both kinds of plants are infected by the same pathogenic fungi. Both leaves and green stems of
71 *Vitis vinifera* L. plants affected by trunk diseases showed physiological and metabolic changes related to the
72 external symptoms, although no pathogens associated with esca proper or other trunk diseases have been isolated
73 from these organs.^{8,10} It was hypothesized that these external symptoms are caused by toxins produced by fungi in
74 the woody tissues and then translocated to the leaves via the transpiration stream.¹¹ On the other hand, responses
75 are still poorly described in woody tissues commonly infected by these pathogens.

76 The present work describes a two dimensional gel electrophoresis (2-DE)-based proteomic study of grapevine
77 response in woody tissues of apoplectic (A) and esca proper (E) affected vines. The main aims were to monitor
78 global changes associated with wood symptoms and identify those proteins differentially expressed in the
79 discolored or asymptomatic wood of apoplectic and esca proper affected plants. For 11 selected proteins identified
80 by nanoLC-MS/MS, the expression of related genes was analysed by a qRT-PCR analysis.

81

82

83 **Experimental procedures**

84

85 **Plant material**

86 Fifteen standing vines (cv. Chardonnay/41B) were uprooted in 2010 from a 26-year-old vineyard located in the
87 Champagne-Ardenne region (France) owned by the company Moët & Chandon. Three plants per target external
88 leaf symptom (GLSD or apoplexy) and three asymptomatic plants were collected. Asymptomatic plants were
89 chosen from among those that had not shown either GLSD or apoplexy symptoms since 2001 and were thus
90 regarded as visually unaffected plants (control plants, C). Trunk and branches of all the collected plants were
91 inspected internally for the presence of discolorations associated with grapevine trunk diseases and subsequently

92 subjected to fungal isolation as described by Spagnolo *et al.*¹². Typical wood symptoms of GLSD and esca were
93 recorded in all plants examined, including control plants. Consequently, plants showing foliar symptoms of GLSD
94 were considered as esca proper (E) affected plants. Two kinds of wood samples from trunk were considered for the
95 proteomic and transcript analysis: asymptomatic and black streaked trunk (BST) wood. Black streaking, which
96 consists of single or more xylem vessels gathered into a blackish brown bundle³ or scattered through the whole
97 section of the organ, was the most representative symptom observed in our sampling. Therefore, six groups of
98 samples were defined for the study: asymptomatic trunk wood of control (C1), apoplectic (A1) as well as esca
99 proper affected plants (E1); and black streaked trunk wood of control (C2), apoplectic (A2) and esca proper
100 affected plants (E2) (see figure of sampling strategy provided in Figure S1, Supporting Information). Three
101 biological replicates per group (i.e., three wood samples from three different plants) were performed. Woody
102 tissues used for protein extraction and RNA extraction were frozen in the field with liquid nitrogen and
103 subsequently stored at -80°C . Before each analysis, the required amount of biological sample was ground to a fine
104 powder in liquid nitrogen with a Mixer Mill MM 400 (Retsch, Haan, Germany).

106 Identification of fungal pathogens

107 The fungal isolates were identified according to the methods of Crous and Gams¹³ (Pch), Essakhi *et al.*¹⁴ (Pal),
108 Fischer¹⁵ (Fom), Phillips¹⁶ (Botryosphaeriaceae), Crous *et al.*¹⁷ (Botryosphaeriaceae) and Carter¹⁸ (*E. lata*).

110 Protein extraction

111 Total protein fraction of woody samples was isolated using a phenol-based procedure according to Spagnolo *et al.*¹⁰. The powdered tissue was placed in microtubes ($0.30 \pm 0.01\text{g}$ of powder per 2.0 mL microtubes) and then
112 resuspended in 1.0 mL of cold acetone. After vortexing thoroughly for 30 s, the tubes were centrifuged at 10000g
113 for 5 min at 4°C . The resultant pellet was washed once more with cold acetone. The pellet was sequentially rinsed
114 with cold 80% acetone 3 times or until the supernatant was colorless, then resuspended in 1.0 mL of cold 20%
115 (w/v) trichloroacetic acid (TCA)/ H_2O . The suspension was sonicated in a water bath at 4°C for 10 min. After
116 centrifugation, the pellets were sequentially washed twice with 20% (w/v) TCA/ H_2O and twice with 80% (v/v)
117 acetone. This pellet was air-dried and the dry powder was resuspended in 0.7 mL dense sodium dodecyl sulphate
118 (SDS) buffer [30% (w/v) sucrose, 2% (w/v) SDS, 0.1 M tris(hydroxymethyl) aminomethane (Tris) -HCl pH 8.0,
119 5% (v/v) 2-mercaptoethanol]. Then 0.7 mL of a 90% phenol solution (Sigma-Aldrich, St. Louis, MO, USA) was
120 added, and the resulting mixture was vortexed for 30 s. The phenol phase, recovered by centrifugation at 10000g
121 for 5 min at 4°C , was separated in two aliquots. One of 0.7 mL was transferred to a 15 mL Falcon tubes while an
122 aliquot of 0.1 mL was placed in a 1.5 mL microtubes. Further steps were followed in parallel. After addition of 5
123 volumes of cold 0.1 M ammonium acetate in methanol, proteins were precipitated from the phenol phase over-
124 night at -20°C . The precipitated proteins were recovered by centrifugation, washed twice with cold 0.1 M
125 ammonium acetate in methanol and twice with 80% (v/v) acetone. The final pellet was air-dried and stored at -80°C .
126 The pellet retrieved from the aliquot of 0.1 mL was dissolved in 100 mL of 8 M urea for protein quantification
127 using the Pierce 660 nm Protein Assay Kit (Thermo Fisher scientific, Waltham, MA, USA) and bovine serum
128 albumin (BSA) as standard. After quantification, protein samples were solubilised in a sample buffer consisting of
129 7 M urea, 2 M thiourea, 4% (w/v) 3-[(3-cholamidopropyl)-dimethylammonio]-1-propanesulfonate (CHAPS), 0.5%
130 (v/v) immobilized pH gradient (IPG) buffer 3-10, 60 mM 1,4-dithiothreitol (DTT), and traces of bromophenol blue.

132 Two-Dimensional Electrophoresis (2-DE)

133 For preparative 2-DE analysis, samples containing approximately $40\ \mu\text{g}$ of total protein fraction were diluted in
134 a mixture containing sample buffer and 10% (v/v) glycerol to a final volume of $125\ \mu\text{L}$. Immobilized pH gradient
135 (IPG) gel strips (ReadyStrip IPG, pH 4–7, 7cm, Bio-Rad, Hercules, CA, USA) were actively rehydrated over-night
136 at 20°C with the mixture. Isoelectric focusing (IEF) was conducted at 20°C in an IPGphor unit (Amersham
137 Pharmacia, Sweden) as follows: a linear increase from 50 to 4000 V to give a total of 10000 V/h. Focused proteins
138 were reduced and subsequently alkylated according to Görg *et al.*¹⁹ IPG strips were then placed on the top of
139 vertical slabs of polyacrylamide (12% T and 2.6% C) and sealed by a layer of 1% (w/v) low melting point agarose,
140 0.15 M Bis-Tris/ 0.1 M HCl, and 0.2% (w/v) SDS. Electrophoretic migration along the second dimension was
141 performed using a Mini-Protean 3 Cell (Bio-Rad) under a voltage of 30 V for 20 min, followed by 150 V for 1.5 h.
142 After completion of SDS-polyacrylamide gel electrophoresis (PAGE), gels were stained with colloidal Coomassie
143

144 Brilliant Blue using the PageBlue™ Protein Staining Solution (Fermentas, USA) following the manufacturer's
145 instructions.
146

147 **Image Analysis**

148 Digitized images at 36.6 μm resolution were obtained using the GS-800 scanner and Quantity One 4.6.2
149 software (Bio-Rad, Hercules, CA). Computerized 2D gel analysis, including spot detection and quantification, was
150 performed using the PDQuest Basic 8.0.1 software (Bio-Rad, Hercules, CA, USA). The relative molecular mass
151 was calibrated with internal protein markers (Precision Plus Protein Standards, Bio-Rad, Hercules, CA) after co-
152 migration during the 2nd dimension. Quantification of detected protein spots was performed calculating the relative
153 optical density×area (relative OD×area) in the gels. Normalization was set up according to the total spot density.
154 Three biological repetitions per group were considered to detect qualitative and quantitative differences in protein
155 expression among the six groups. The mean relative OD×area±SD ($n=3$) values of each group were finally used to
156 estimate relative expression level (relative OD×area %) of each protein spot among the groups. Differences among
157 the means were evaluated by the Dunn's Multiple Comparison Test after that the null hypothesis (equal means) was
158 rejected in the Kruskal-Wallis test, assuming a significance of $p \leq 0.05$. The relative expression ratio to C1 in the
159 other groups was also estimated. Values higher than 1.5 or lower than 0.5 were discussed.
160

161 **Protein identification by mass spectrometry**

162 Protein spots of interest were excised manually and submitted to in-gel digestion. Reduction, alkylation and
163 tryptic in-gel digestion were performed as previously described.¹⁰ Tryptic digests were analyzed by C18 reversed
164 phase nanoHPLC on a nanoHPLC-Chip/MS system (Agilent Technologies, Palo Alto, USA) coupled to an ion trap
165 amaZon (Bruker Daltonics, Bremen, Germany) mass spectrometer. The complete system was fully controlled by
166 Hystar 3.2 (Bruker Daltonics). For tandem MS experiments, the system was operated in the data-dependant mode
167 with 6 MS/MS scans.

168 Mass data collected during nanoLC-MS/MS were processed, converted into “.mgf” files with DataAnalysis 4.0
169 (Bruker Daltonics) and interpreted using the MASCOT 2.3.02 algorithm (Matrix Science, London, UK) and Open
170 Mass Spectrometry Search Algorithm (OMSSA). Searches were performed without any molecular weight or
171 isoelectric point restrictions against an in-house generated protein database composed of protein sequences of *Vitis*,
172 human keratins and trypsin, downloaded from the National Center for Biotechnology Information nonredundant
173 database (NCBI nr, June 19, 2012) concatenated with reversed copies of all sequences (total 138416 entries).
174 Database searching was carried out using the following parameters: 2 missed cleavages; a parent and fragment
175 mass tolerance of ± 0.25 Da; carbamidomethylation, N-terminal acetylation, oxidized methionine as variable
176 modification.

177 Mascot and OMSSA results were loaded into the Scaffold 3.6.5 software (Proteome Software Inc., Portland,
178 USA). To minimize false positive identifications, results were subjected to very stringent Mascot and OMSSA
179 filtering criteria as follows: 1) for the identification of proteins, all peptides were validated with both algorithms
180 (Mascot and OMSSA); 2) for proteins identified with two peptides or more, OMSSA -Log(E-Value) scores were
181 greater than 7, Mascot ion minus identity scores (the 95% Mascot significance threshold) greater than -5 and each
182 ion scores were higher than 30; 3) in the case of single peptide hits, OMSSA -Log(E-Value) scores were higher
183 than 8, Mascot ion minus identity scores were greater than 10 and unique peptide ion scores greater than 30. The
184 target-decoy database search allowed us to control and estimate the false positive identification rate of the study.²⁰
185 Thus, the final catalogue of proteins corresponded to an estimated false positive rate below 1%.

186 A list of all identified proteins with Mascot and OMSSA is provided in Table S1 and the spectra of single
187 peptide hits in Table S2 (Supporting Information). The MS data have been submitted to the PRIDE proteomics
188 identification database²¹ (www.ebi.ac.uk/pride) under accession numbers # (27098).
189

190 **Spot clustering and functional classification of identified proteins**

191 Identified protein spots were classified according to the relative expression values (relative OD×area%) in each
192 group by performing hierarchical tree clustering with the NCSS 8 software by NCSS Statistical Analysis &
193 Graphics (<http://www.ncss.com/>). Clustering parameters were set up as follows: group average (unweighted pair-
194 group) linkage type, Euclidean distance method, average absolute deviation scaling method, cluster cutoff at 0.01.

195 A functional classification of the identified proteins was performed by using GenomeNet Database Resources
196 (<http://www.genome.jp/kegg>) or according to their role described in the literature.
197

198 **RNA Extraction**

199 Total RNA was isolated from 3×50 mg of woody tissues powder using the plant RNA purification reagent
200 (Invitrogen, Cergy Pontoise, France). The RNA pellet was resuspended in 20 µl of RNase free water, then treated
201 with RQ1 DNase enzyme (Promega) and quantified by measuring the absorbance at 260 nm according to the
202 manufacturer's instructions (Biowave DNA Spectrophotometer, Biochrom WPA).
203

204 **Real-time RT-PCR analysis of gene expression**

205 In total, 150 ng of total RNA were reverse-transcribed using the Verso SYBR 2-step QRT ROX enzyme
206 (ABgene, Surrey, UK) according to the manufacturer's protocol. PCR conditions were as described by Bézier *et*
207 *al.*²². Expression of 11 targeted genes selected from proteomic results was tracked by quantitative Reverse-
208 Transcripts Polymerase Chain Reaction (qRT-PCR), including the α -chain elongation factor 1 gene (*EF1- α*), which
209 was used as the internal standard to normalize the starting template of cDNA. This analysis was performed using
210 the following gene-specific primers presented in Table S3 (Supporting Information). The cDNA sequence was
211 obtained on NCBI Blast protein website (<http://blast.ncbi.nlm.nih.gov/>). We use the accession number (gi) of the
212 targeted protein (*Vitis vinifera*) to search and obtain the corresponding cDNA sequence. Primers were designed at
213 60°C T_m to amplify fragments from 110 to 220 bp using Primer express 2.0 software (Applied Biosystems). The
214 cDNA sequences are tested under NCBI Blast nucleotide to check their theoretical specific character to the targeted
215 gene and not to a gene family. The amplification specificity of each qRT-PCR was confirmed by the presence of a
216 single peak in the melt curve analysis, and no primer dimers were detected using agarose gel electrophoresis.
217 Reactions were carried out in a real-time PCR detector Chromo 4 apparatus (Bio-Rad) using the following thermal
218 profile: 15 s at 95°C (denaturation) and 1 min at 60°C (annealing/extension) for 40 cycles. The efficiency of the
219 primer sets was estimated by performing real-time PCR on several dilutions. The results were normalized with the
220 *EF1- α* gene and expressed relative to the control corresponding to asymptomatic- or BST-wood collected from
221 control plants. PCR reactions were performed in duplicate for three biological replicates, for each of the 6 groups
222 of samples. The results presented correspond to the means (\pm standard deviation) of the independent experiments.
223
224

225 **Results**

226 **Isolation of fungal species associated with grapevine trunk diseases**

227 The major causal agents of esca proper (Pch, Pal and Fom) were isolated from discolored woody tissues of the
228 three groups of plants analyzed (E and A affected, and control plants). Fom was directly linked to white decay.
229 Other fungi associated with grapevine trunk diseases, such as Botryosphaeriaceae species¹² and *Eutypa lata*²³, were
230 also isolated from woody tissues as shown in the Table S4.
231

232 In contrast, no fungi were isolated from non-discolored wood of the trunk and branches of control or diseased
233 plants (Table S4, Supporting Information).
234

235 **Differences in protein yield and number of spots detected among the groups of samples**

236 Protein yield was related to the type of wood samples, asymptomatic (C1, A1, E1) or BST (C2, A2, E2), but not
237 to the absence/presence of external symptoms. In fact, values of 0.31 \pm 0.03, 0.33 \pm 0.04 and 0.27 \pm 0.06 μ g mg⁻¹ of
238 fresh weight were observed for C1, A1 and E1, respectively, whereas values of 0.11 \pm 0.07, 0.12 \pm 0.08 and
239 0.10 \pm 0.06 μ g mg⁻¹ of fresh weight were obtained in the case of C2, A2 and E2, respectively. This lower protein
240 yield recorded in the BST wood samples could be the result of an altered physiology and/or the possible presence
241 of phenolic compounds interfering with the protein extraction.²⁴ On the other hand, the average number of protein
242 spots resolved on 2-DE gels was not significantly different among the six groups in the pH range used in this study
243 (4–7). Indeed, means ranged between 148 \pm 8 (A1) and 152 \pm 6 (E2). This might indicate that the proteome variations
244 occurring in the BST wood were mainly quantitative, as observed for the 30 differentially expressed protein spots
245 selected for identification (Figure 2, Supporting Information). The coefficients of variation associated with the

246 number of spots detected in the biological replicates were 3.74%, 5.02%, 5.11%, 4.00%, 3.70% and 4.01% for C1,
247 C2, A1, A2, E1, and E2, respectively.

248

249 **Identification of differentially expressed protein spots by nanoLC-MS/MS**

250 Protein spots detected in at least 2 biological repetitions of a given group were analyzed among all the groups.
251 Accordingly, a total of 148 protein spots were matched and compared by image analysis. Among the differentially
252 expressed spots for at least one group, 30 spots were subjected to in gel trypsin digestion followed by
253 nanoLC-MS/MS analysis. In most cases, more than one protein was identified in the same spot (Table S1,
254 Supporting Information), resulting in a large number of unique proteins to be analyzed (153). The 153 identified
255 proteins were classified into nineteen functional groups by using the GenomeNet Database Resources website
256 (<http://www.genome.jp/kegg>) and other reports in the literature (Figure 1). Among these identified proteins, 12%
257 were classified as being involved in the stress response, 10% in carbohydrate metabolism and in cellular processes,
258 8% in translation, 7% in transcription, 6% in protein destination and 5% in secondary metabolism (Figure 1). A
259 selection based on their biological importance was performed and, a total of 53 proteins were chosen (Table 1). As
260 shown in Table 1, experimental and theoretical molecular weights (Mw) correlated more closely than related
261 isoelectric points (pI). Indeed, Pearson's correlation coefficient ($p \leq 0.05$) was 0.920 and 0.402 for the Mw and pI
262 comparisons, respectively. The strongest discrepancies concerning the Mw were recorded in spots s3409, s6201,
263 s6203 and s7319, where ratio values as high as two were found. In the case of the pI comparison, it should be noted
264 that for some identified proteins the theoretical pI was found to be higher than 7 (spots s4004, s5422, s6001, s6201,
265 s6203, s7319 and s8102). Such differences, as well as the presence of the same protein in several spots, may be due
266 to the presence of protein variants and/or post-translational modifications, truncation or degradation products.¹⁰ The
267 discrepancies between experimental and theoretical Mw values could also be due to the fact that some protein
268 sequences contained a signal peptide in the database, which effectively increases the theoretical molecular mass
269 since the whole sequences are considered for the calculation.

270

271 **Differentially expressed proteins among the groups of samples**

272 Comparison of the relative expression values of the spots from the six groups showed that C2 had a similar
273 protein expression to the other groups the largest number of times (75), whereas the lowest frequency was observed
274 for both E2 (54) and E1 (52) (Figure 2A). When all the possible pairs were considered separately, the most similar
275 protein expression was recorded for the pairs C1-A1 and A2-E2 (18 spots), whereas the least similar was observed
276 for E1-E2 and C1-E2 (7 spots) (Figure 2B). Although to a lesser extent than for the A2-E2 pair, a weak similar
277 expression to A2 and E2 was also observed when the C2 group was considered (16 and 13 spots, respectively)
278 (Figure 2B). However, taken together C2, A2 and E2 had a similar expression in only 8 spots (1202-1208-2211-
279 2302-3409-4102-7102-8102) (Figure 2B). Therefore, protein expression in these groups did not seem to be only
280 linked to the presence of the same symptom (BST) but to other factors. In this regard, it is important to note that the
281 percentage of GLSD agent isolation was similar for C2, A2 and E2 (50–70% of the positive samples), but different
282 frequencies of positive samples (where at least one fungal colony was detected) of 86%, 50% and 41% were
283 observed for E2, A2 and C2, respectively. Large differences in protein expression were apparent among the groups
284 from asymptomatic wood (C1, A1 and E1). In fact, except for the C1-A1 pair, the number of spots with similar
285 protein expression became lower when the E1 group was considered (Figure 2B). When the three groups were
286 considered together, they showed similar expression values only in five spots (3409-6201-6203-6420-7102).
287 Protein expression was therefore not strictly linked to the type of wood samples. Indeed, proteome variations were
288 induced in the asymptomatic wood of apoplectic (A1) and esca proper affected (E1) plants. Moreover, differences
289 in protein expression between asymptomatic and BST wood did not show the same trend among the groups of
290 plants (C1-C2, A1-A2, E1-E2), being more marked in A and E affected plants. Actually, the most similar protein
291 expression was registered for the C1-C2 pair (17), followed by A1-A2 (12) and E1-E2 (7) (Figure 2B).

292

293 **Proteomic changes in asymptomatic and BST wood of trunk**

294 To obtain a comprehensive overview of the dynamics of protein abundance associated with foliar symptom
295 emergence, hierarchical tree clustering was performed and protein abundance patterns were grouped together
296 without a priori knowledge of the biological reasons for the existence of these groups (Figure 3). Moreover,
297 proteins were described and discussed according to their putative biological functions (Figure 1).

298 The hierarchical tree clustering analysis of the six spots (s2212, s2303, s3409, s6201, s6420, s7319) of
299 carbohydrate metabolism (including glycolysis metabolism) indicated that the distribution of these spots was
300 different between asymptomatic and BST wood, regardless of the disease form. Malate dehydrogenase (s6201) and
301 transaldolase isoform 1 (s2302) were always grouped in the same cluster in BST wood (C2, A2 and E2) but not in
302 asymptomatic wood (C1, A1 and E1). During foliar symptom emergence, malate dehydrogenase (s6201),
303 transaldolase (s2302), acid phosphatase 1-like (s2212), hexokinase 2 (s3409), isocitrate dehydrogenase (s6420)
304 were all down-regulated in asymptomatic wood of diseased plants, irrespective of the disease form (A or E).
305 Conversely, a glyoxylate reductase isoform (s7319) and transaldolase (s2302) were over-expressed in A2 and E2
306 (Table 1, Figure 4). In parallel to the proteomic analysis, we followed the *transaldolase* gene expression, which did
307 not correlate with the protein level (Figure 5). No predictions of protein levels present in the organ could be
308 formulated with transcript analysis.

309 Plant growth was especially perturbed. A GEM-like protein 5 (s1208) implicated in tissue growth was
310 identified. Expression levels in A1 and E1 were lower than those observed in C1. Additionally, an endo-(1,3)-(1,4)-
311 β -D-glucanase (s2211) associated with modulation of cell growth showed weak expression in wood of diseased
312 plants A1 and E1 (Table 1, Figure 4). Dihydrofolate reductase (DHFR, s5316) involved in cell division was down-
313 regulated during symptom emergence. Alteration of cell wall biogenesis was also observed and concerned the
314 production of the cell wall sugar D-apiose. The polymyxin resistance protein ARNa (s5422) has 97% homology to
315 an UDP-apiose/UDP-xylose synthase. The over-regulation of this enzyme identified in both A2 and E2 may reflect
316 a stress response of the plant to increase cell wall integrity (Figure 4).

317 Nitrogen metabolism was also altered with the appearance of foliar symptoms; an ornithine aminotransferase
318 (s6420) was down-expressed in A1, A2 and E1. A similar trend was observed for the nitrogen regulatory protein II
319 (PII, s6001), a nuclear-encoded plastid protein that regulates ornithine and arginine biosynthesis, for which lowest
320 levels were observed in A1 and E1.

321 The distribution of the ten spots containing proteins involved in protein synthesis and degradation (s3409,
322 s4004, s4216, s4221, s5422, s6201, s6203, s7210, s7319, s8102) (Table 1, Figure 3) was different in control plants
323 depending on the sample state, C1 and C2 (Figure 3). Expression of nine of these spots (s4004, s4216, s4221,
324 s5422, s6201, s6203, s7210, s7319, s8102) was altered in the trunk of grapevines affected by apoplexy (Figure 3).
325 Instead, the relative abundance of seven spots (s4004, s4216, s4221, s5422, s7210, s7319, s8102) was affected
326 during GLSD appearance (Figure 3). The impact of foliar symptom emergence on proteins variations varied
327 depending on the type of sample. For instance, proteasome subunit beta (s7210) was down-regulated in
328 asymptomatic wood (A1, E1) and up-regulated in BST wood (A2, E2).

329 Two small heat shock proteins (smHSPs) were identified: a smHSP chloroplastic (smHSPCP, s3105-s3107-
330 s3202-s3203) and a 17.9 kDa class II HSP (HSP17.9; s4004-s6001-s6101) (Table 1). Hierarchical tree clustering
331 also indicated that, in these cases, the distribution of the corresponding spots depended on the sample state,
332 asymptomatic or BST wood (Figure 3). These spots were mainly grouped in the clusters (1, 2 or 3) with the highest
333 abundance in C1, E1 and A1. They were also grouped in clusters (2, 3, 4 or 5) with a lower abundance in wood
334 with the presence of fungi. The two smHSPs (s3105-s3107-s3202-s3203 and s4004-s6001-s6101) showed lower
335 abundance in A1 and E1 in comparison to C1. Results of transcript analysis for *smHSPCP* corroborated the results
336 of protein expression (Figure 5). Hierarchical tree clustering indicated that the distribution of the spots (s4221,
337 s5215, s6002, s6201, s7210), containing proteins involved in antioxidant system, was independent of the sample
338 state in control plants (Figure 3). In contrast, the abundance of these spots was affected during foliar symptom
339 emergence depending on the proteins and disease form (Figures 2 and 3). The GST dehydroascorbate reductase
340 protein (GST DHAR, s6201) showed a similar trend of expression in asymptomatic and BST wood, with a decrease
341 of abundance in diseased plants (Figure 4). The two isoforms of glutathion S-transferase (GST) F9 (s4221-s5215
342 and s7102) and the superoxide dismutase chloroplastic (SODCP) (s4102) were down-expressed in A1 and E1,
343 whereas GSTF9s (s4221-s5215 and s7210) and SOD isoform II (s6002) were over-expressed in A2 and E2 (Figure
344 3). No significant down- or up-regulation was observed for the related *GSTF9* or *SODCP* genes (Figure 5).

345 Regarding the defense response, a hypothetical protein homolog to the “pathogenesis-related” protein (PRP)
346 PRP-17 (s4221) was identified (Table 1). In the control plants, the relative abundance in C1 seemed to be 2-fold
347 lower than in C2. The increase in abundance was higher in E1 and E2 than in A1 and A2 (Figure 4). The transcript
348 analysis suggested weak up-regulation of *PRP-17* expression only in E1 and no correlation with the protein
349 abundance in BST samples (Figure 5). Endochitinase (s1202) and thaumatin-like protein (TLP, s1105) showed

350 decreased abundance in A1 and E1 and increased abundance in A2 and E2. *TLP* expression was significantly
351 induced in A2 (80-fold) and E2 (26-fold) in comparison to C2 (Figure 5) and correlated positively with the protein
352 abundance in BST wood of diseased grapevine. Spots containing PRP10 (s4102-s6101-s8102) showed a decrease
353 in relative abundance in A1 and E1. Protein and transcript expressions of PRP-10 were not significantly different
354 between BST samples.

355 Secondary metabolism-related proteins identified in this study were involved in the phenylpropanoids pathway.
356 Three sub-groups of proteins that changed in abundance with foliar symptom emergence were distinguished: the
357 first comprised a single protein with increased abundance in both asymptomatic and BST woods (s4216, chalcone
358 flavonone isomerase), the second also included a single protein with increased abundance in BST wood alone
359 (s7319, isoflavone reductase homolog), while the third one included two proteins with decreased abundance in both
360 asymptomatic and BST wood (s4212-s5316, isoflavone reductase-like (IFRL) protein 4; s6420, leucoanthocyanidin
361 dioxygenase) (Figure 4). Results of the supplementary analysis of transcripts highlighted the up-regulation of *CHI*
362 in asymptomatic wood and the down-regulation of *IFRL4* (Figure 5). S-adenosyl-L methionine (SAM) synthase
363 plays an important role in the production of SAM leading to polyphenols. Hierarchical tree clustering indicated that
364 the distribution of the two forms of SAM (s6420 and s5422) was varied depending on the sample observed; the two
365 spots were grouped in the same cluster for E1 and C2 but not for C1, A1, A2 and E2 (Figure 3). The results
366 revealed the lowest level for SAMs isoform 2 (s6420) in A1, E1 and A2 and the highest abundance in A2 and E2
367 for SAMs isoform 4 (s5422) (Figure 4). The exact role of the six stress-induced proteins in plant defense is unclear.
368 These six proteins showed decreased abundance in asymptomatic wood of diseased plants. They were identified as
369 an uncharacterized protein (s1007) belonging to the PLAT family (polycystin-1 lipoyxygenase alpha toxin domain),
370 a stem specific protein TSJT1 (SSP, s4221-s5215-s6201), a glycine RNA binding protein (GR-RBP, s6002), a salt
371 stress root protein RS1 isoform 3 (RS1; s2211); a 14-3-3 protein 1-like (14-3-3, s1202-s2212) and a major latex
372 protein (MLP) like protein 28 (s7102-s7319) (Table 1). No correlation was observed between the transcript
373 analysis and the proteins expression of both 14-3-3 and SSP (Figures 4 and 5). Nevertheless, four of them were
374 accumulated in E2 and A2: 14-3-3 (s1202-s2212), GR-RBP (s6002), RS1 (s2211), and SSP (s4221-s5215-s6201)
375 (Figure 4). Moreover, transcript analysis of *SSP* and *14-3-3* indicated significant up-regulation in E2 compared to
376 A2 and C2 (Figure 5). These results revealed that protein and transcript expression during foliar symptom
377 emergence varies according to the type of sample (asymptomatic or BST wood).

378

379 Discussion

380

381 Primary metabolism altered in the trunk of apoplectic and esca proper affected grapevines

382 Recognition of pathogens results in significant reprogramming of the plant cell to activate and deploy defense
383 responses to halt the pathogen growth. If primary metabolism is reconfigured to support the increased demands for
384 the defense response, a lower protein abundance for enzymes involved in carbon metabolism is likely to be
385 observed, particularly in asymptomatic wood. These proteins, which are involved in metabolic pathways like
386 gluconeogenesis, citrate and pyruvate metabolism, were expressed in control plants. The observed accumulation of
387 glyoxylate reductase (s7319) in the presence of fungal agents (BST wood) may support its putative role in the
388 stress response.²⁵ In addition, carbohydrate increase is believed to be a metabolic signal that induces both the
389 expression of defense-related genes and the repression of photosynthesis.²⁶⁻²⁸ For example, the accumulation of
390 soluble hexoses can regulate programmed cell death in plants²⁹, suggesting that sugar sensing mediates a direct link
391 between carbohydrate metabolism and the defense response. As a consequence of the repression of primary
392 metabolism, a negative effect on cell growth/division was observed, e.g., diminution of the relative abundance of
393 DHFR (s5316) and endo-(1,3)- β -glucanase (s2211). DHFR decrease may stop the growth of woody tissues; Wu *et*
394 *al.*³⁰ reported that inhibition of cell growth in *Arabidopsis* was correlated to inhibition of DHFR activity. Our
395 isoform of endo-(1,3)- β -glucanase was recently linked to modulation of cell growth in grapevine berries.³¹ All
396 these data suggest inhibition of cell growth during foliar symptom emergence.

397 Alteration of nitrogen metabolism and especially the proline pathway was also observed. PII (s6001), a nuclear
398 protein that regulates a key enzyme of the ornithine and arginine biosynthesis³² and the ornithine aminotransferase
399 (OAT, s6420) involved in the ultimate formation of proline to ornithine, showed a decrease in the relative
400 abundance in diseased grapevine. Many studies have implicated proline metabolism in the plant-pathogen
401 interaction^{33,34} and disease resistance.³⁵ Our study suggests that decreased expression of these proteins could be

402 related to symptom expression. However, certain amino acids, e.g., proline, asparagine, alanine, are known to
403 stimulate hyphae growth of fungal pathogen.³⁶ The inhibition of these pathways by the plant itself may limit
404 colonization of the grapevine by the causal agents.

405 Proteins involved in protein synthesis and degradation also play an important role in the plant-pathogen
406 interaction. Among the proteins that responded to GLSD and/or apoplexy symptom emergence, various 40SRP
407 (s8102) or 60SRP (s4004, s5422, s7319, s8102) proteins and subunits of the proteasome complex known to be
408 altered in response to grapevine-pathogen interactions were identified. After *Xylella fastidiosa* inoculation, 40SRP-
409 S25 expression was strongly enhanced in the grapevine susceptible-genotype in comparison to the resistant-one.³⁷
410 The over-regulation of ribosomal proteins indicates a positive role in aggravating pathogenesis by altering the
411 translation machinery.³⁸ In the case of two subunits of the ubiquitin/proteasome complex, α -5 and β -6 were slightly
412 up-regulated in grapevine in response to *Phytoplasma* infection.³⁹ Vincent *et al.*⁴⁰ correlated apparent activation of
413 the proteolysis machinery to reduced ability of the plant to cope physiologically with water or salt stresses. The
414 activation of proteolysis machinery observed in BST wood of both apoplexy and GLSD plants may increase
415 disease development. Nevertheless, proteasome activity has been also linked with activation of defense reactions.⁴¹
416 In support of this, putative defense proteins were identified at higher abundance in BST than in asymptomatic
417 wood.

418 **The majority of proteins identified in the trunk of grapevine affected by esca proper and apoplexy were** 419 **related to cell rescue and defense**

420 Remarkably, the majority of proteins identified in this study were related to cell rescue and defense. Among
421 them, small heat shock proteins (smHSPs, s3105-s3107-s3202-s3203 and s4004-s6001-s6101) play an important
422 chaperone role in maintaining cellular functions when plants are subject to a variety of stress. Because of their
423 well-known role as chaperones, HSPs were classified in the protein destination category. However, many studies
424 have suggested they are involved in heat stress initially and various biotic or abiotic stresses later.⁴² In grapevine,
425 Yang *et al.*³⁷ reported that smHSPs were differentially expressed in Pierce's Disease (PD)-resistant and PD-
426 susceptible genotypes and suggested that these proteins might be implicated in resistance. Consistent with our data,
427 a decrease in abundance of smHSPs in asymptomatic wood was observed and could be related to the emergence of
428 foliar symptoms. A similar down-regulated expression of one smHSP chloroplastic gene was observed in the leaves
429 of a susceptible cultivar (Riesling) during *Plasmopara viticola* infection.⁴³ These results suggest that smHSPs could
430 be systemically induced in grape and may play a role in resistance. Since an important key regulator of smHSPs is
431 H₂O₂, the synthesis of smHSPs can be considered as an adaptative mechanism in which cellular protection against
432 oxidative stress is essential.^{44,45}

433 In this study, the expression of five proteins involved in the antioxidant system was also altered: two of them
434 showed a decrease in abundance (SODCP, s4102 and GST DHAR, s6201), one an increase (SODII, s6002) and the
435 other two a decrease/increase (two isoforms of GSTF9, s4221-s5215 and s7102) in asymptomatic/BST wood of
436 diseased plants, respectively. Reduced expression of SOD and GST was observed in asymptomatic green stem of
437 apoplectic plants.¹⁰ Moreover, other recent studies have reported repression of *SOD* in leaves of apoplectic or esca
438 proper affected plant.^{5,8} Therefore, the antioxidant system associated with cell plant tolerance seems to be perturbed
439 in symptomatic plants. These physiological perturbations were observed in various parts of the grapevine:
440 leaves^{5,8,9}, stems¹⁰ and woody tissues (this study). Recent studies have described a significant increase in the level
441 of transcripts encoding GSTs in herbaceous organs of compatible grapevine affected by *Erysiphe necator*⁴⁶, *Eutypa*
442 *lata*⁴⁷, *P. viticola*⁴³, esca proper and apoplexy.¹⁰ These results suggest that the GST system is not related to the
443 plant's tolerance against pathogenic fungi. Moreover, comparison between a resistant- and susceptible-grapevine
444 cultivar showed that both GST and catalase (another antioxidant enzyme) are not related to the innate resistance
445 capabilities of the resistant cultivar against fungi and is probably instead related to environmental conditions.⁴⁸ As
446 indicated by Valtaud *et al.*⁹, glutathione (GSH) metabolism provides new markers for the presence of esca in
447 grapevine, which are not correlated to a hypothetical marker for tolerance of a cultivar. To prevent oxidative
448 damage, GSTs perform diverse catalytic as well as non-catalytic roles in the detoxification of xenobiotics, such as
449 toxins.⁴⁹ The toxin eutypin, produced by another grapevine trunk disease agent, *Eutypa lata*, is metabolically active
450 in the aldehyde form.⁵⁰ The major wood-infesting fungi Pch, Pal and Fom have been shown to produce diverse
451 toxins.⁵¹ Toxins might undergo similar processing leading to their detoxification in BST wood. In spite of the over-
452 regulation of antioxidant enzymes in BST wood observed here, the grapevines were affected by either apoplexy- or
453

454 GLSD- symptoms. Several studies have reported that GSH is important in the plant's pathogen response regulation
455 and that NPR1 and TGA1 are key redox-controlled regulators that modulate the systemic acquired resistance and
456 PRP expression.⁵²⁻⁵⁴

457 Regarding the defense response, a PRP-10 protein (s4102-s6101-s8102), an endochitinase (s1202), a TLP
458 (s1105) and a homolog to a PRP-17 (s4221) protein were identified and changed in abundance according to the
459 disease form expression and type of wood. PRP-10s are a large group of proteins, constitutively expressed⁵⁵ or
460 induced in grapevine by pathogen infection.^{37,43,47,56} In the present study, the lower abundance of PRP-10 may
461 indicate that it plays a defensive role in the grapevine response to esca fungi infection. This is supported by the
462 increase of *PRP-10* expression noted in grapevine cell cultures in response to elicitors produced by Pch.⁵⁷ The class
463 IV endochitinase, known to be induced in grapevine by pathogen infection⁴⁷, takes part in plant defense through the
464 hydrolysis of chitin fragments of fungal cell walls. Interestingly, its abundance changed by decreasing in
465 asymptomatic wood and increasing in BST wood of diseased vines. Similar trends of protein expression were
466 observed for TLP. These results were surprising because TLPs are normally expressed at low levels in healthy
467 plants and rapidly accumulate to high levels in response to biotic stress.^{46,47} The thaumatin-like proteins are defined
468 as a PRP and, along with osmotin-like proteins, form the PR-5 family.⁵⁸ Members of this family are known to have
469 antifungal⁵⁹ and anti-oomycete properties.⁶⁰ Considering both the expression levels of endochitinase and TLP, it is
470 obvious that their change in the trunk (decrease in asymptomatic wood or increase in BST wood) has no preventive
471 effects on the onset of foliar symptoms. Finally, a hypothetical protein homolog to PRP-17 was up-regulated in the
472 trunk of diseased vines, but this induction does not avoid foliar symptom appearance. Similar induction has already
473 been observed during grapevine-trunk diseases agent interaction. Indeed, transcript analysis of a PRP-17 gene
474 (called *NtPR27-like*) has suggested it is induced in grapevine leaves in response to *E. lata* infection.⁴⁷

475 Many plant defense responses involve the production of secondary metabolites.⁶¹ S-adenosyl-L methionine
476 (SAM) synthase (s) participates in the production of SAM, leading to phenolic compounds involved in defense
477 responses.⁶² Thus, transcripts of SAMs have been shown to be up-regulated in the *E. necator*- and *P. viticola*-
478 resistant cultivar (Regent), suggesting a role for SAMs in the intrinsic resistance capability of this cultivar.⁴⁸ Our
479 study identified two SAMs (s5422 and s6420) with opposite evolution in trunk samples of vines affected by the
480 foliar symptom emergence. Spagnolo *et al.*¹⁰ showed that the lowest level of SAMs occurred in symptomatic green
481 stems of A and E affected plants. Functions of phenylpropanoid compounds in plant defense range from preformed
482 or inducible physical and chemical barriers against infection to signal molecules involved in local and systemic
483 signaling for defense gene induction.⁶³ With regard to the phenylpropanoid pathway, a chalcone flavonone
484 isomerase (CHI) (s4216) was identified, which was more abundant in the trunk of diseased vines and especially in
485 BST wood. Since CHI appears to be involved in flux control into the flavonoid and isoflavonoid pathways⁶³, our
486 results suggest that these pathways are positively altered in response to foliar symptom emergence. In addition, two
487 isoforms of isoflavone reductase (IFR, involved in biosynthesis of isoflavonoid phytoalexins, s7319 and s4212-
488 s5316) and one leucoanthocyanidin dioxygenase (LDOX, involved in the anthocyanins biosynthesis, s6420) were
489 identified and showed a decrease in relative abundance in the trunk of diseased vines. Hence, alteration of the
490 isoflavonoid/anthocyanin pathways in the asymptomatic wood might be related to the presence of foliar symptoms
491 and/or to their possible imminent development. Similarly, Camps *et al.*⁴⁷ associated the induction of *LDOX* genes
492 expression to the lack of leaf symptoms on grapevine affected by *E. lata*. Production of phenylpropanoid
493 compounds was stimulated in BST wood of either apoplectic or esca proper affected vines. This induction may be
494 targeted especially on the flavonoid branch.

495 Among the six stress-inducing proteins with unclear role, three have already been described in various plant-
496 microbe interactions. The major latex protein (MLP), known to be associated with pathogen defense responses⁶⁴,
497 was down-regulated in grapevine leaves to phytoplasma infection.⁵⁶ An increase of GR-RBP proteins has been
498 reported in pea during *Peronospora viciae* infection, providing evidence of its role in the response of plants to
499 pathogen.⁶⁵ Finally, many studies have reported that 14-3-3 transcripts are accumulated in susceptible- and
500 resistant-plants affected by pathogenic fungi.⁶⁶ It may be speculated that these proteins play an active part in the
501 defense response by making the host cell more resistant to infection and penetration.

502 503 **Different responses in asymptomatic and BST wood independent of disease form**

504 This study also indicated that the wood protein quantity was more affected by the nature of the sample than the
505 diseased form. A comparison of protein expression in asymptomatic and BST woods (C1/C2, A1/A2, E1/E2)

506 revealed three interesting categories/trends related to the presence/absence of wood and foliar symptoms. In the
507 first category, we observed an increase of protein expression in BST wood with respect to asymptomatic wood
508 (s1105, s1202, s1208, s2211, s2212, s2302, s4221, s4216, s5215, s6002, s7319). The main proteins identified in
509 these spots belonged to the detoxification system (GST, SOD), primary metabolism (GEM-like, transaldolase,
510 endo- β -D-glucanase), stress-induced proteins (14-3-3, RS1, SSP), phenylpropanoid metabolism (CHI, IFR-L4) and
511 PR proteins (TLP). The significant induction of the detoxification system in symptomatic plants confirms its role as
512 a stress marker of GLSD infection⁹ and supports the toxin hypothesis in explaining at least the GLSD symptoms
513 development.⁸ The strong protein expression from the phenylpropanoid pathway in BST wood is in accordance
514 with the observed accumulation of phenolic compounds.⁶⁷ In contrast, the second category was characterized by
515 higher protein expression in asymptomatic wood than in BST wood (s1007, s3105, s3107, s3202, s3203, s4004,
516 s6203, s6420) (Figure 4, Table 1). Most of these proteins were smHSPs, which suggests that these proteins play an
517 important defensive role in the grapevine wood before or at the beginning of extensive fungi colonization. These
518 smHSPs may act as molecular chaperones, resulting in the maintenance of cellular conditions suitable for inducible
519 plant defense responses.⁶⁸ Finally, the third category was characterized by proteins showing higher expression
520 levels in asymptomatic or BST wood of control plant than in apoplectic and esca proper affected plants (s4212,
521 s5316, s6201). The major proteins included in this category belonged to the antioxidant system (DHAR2),
522 proteasome activity, stress-induced proteins (SSP) and isoflavonoid pathway (IFR-L4). This finding suggests that
523 these proteins might have an important role in preventing or limiting foliar symptom emergence.
524

525 **Conclusion**

526 Different quantitative protein expressions related to the nature of samples, BST wood (high rate of GLSD
527 and/or esca agent inoculum) or asymptomatic wood (low or absent inoculum) were more pronounced than those
528 related to the expressed foliar symptoms, either apoplexy or GLSD. In asymptomatic wood, alteration of protein
529 accumulation was mainly characterized by down-regulation of proteins involved in various biological functions,
530 such as cell growth or division (GEM-like, transaldolase, endo-(1,3)- β -glucanase) and plant defense responses
531 (smHSP, TLP, PRP-17, IFR-L4). In BST wood, characterized by the extensive presence of grapevine trunk disease
532 agents, over-regulation of proteins involved in plant defense (14-3-3), phenylpropanoid pathway (CHI) and PR
533 proteins (TLP, PRP-17) was shown. It could therefore be speculated that a diverse effect occurs in terms of protein
534 expression in asymptomatic- or BST-wood of plants affected by esca proper or apoplexy during symptoms
535 expression. Finally, remarkable alterations of the anti-oxidant system were observed in both types of wood during
536 foliar symptom emergence, thus supporting the hypothesis that fungal toxins from woody tissues are involved in
537 foliar symptom development.^{3,69}
538

539 **Acknowledgements**

540 This research was financed by the national CPER (Contract Project État-Région) program - task E1 altesca and
541 CASDAR (Compte d'Affectation Spéciale au Développement Agricole et Rural). Proteomic studies were
542 supported by the CNRS, the "Agence National de la Recherche" (ANR) and the "Region Alsace". The company
543 Moët & Chandon is thanked for making available the vineyard used as the experimental plot in this study.
544

545 **Supporting information**

546 Supporting Information Available: This material is available free of charge via the Internet at
547 <http://pubs.acs.org>.
548

549 **References**

- 550 (1) Fussler, L.; Kobes, N.; Bertrand, F.; Mauray, M.; Grosman, J.; Svary, S. A characterisation of grapevine
551 trunk diseases in France from data generated by the National Grapevine Wood Survey. *Phytopathology*
552 **2008**, *98*, 571–579.
- 553 (2) Bertsch, C.; Ramirez-Suero, M.; Magnin-Robert, M.; Larignon, P.; Chong, J.; Abou-Mansour, E.;
554 Spagnolo, A.; Clément, C.; Fontaine, F. Trunk diseases of grapevine: complex and still poorly understood
555 syndromes described as eutypa dieback, esca and black dead arm. *Plant Pathol.* **2012** Doi: 10.1111/j.1365-
556 3059.2012.02674.x
- 557 (3) Surico, G.; Mugnai, L.; Marchi, G. The esca disease complex. In *Integrated management of*

- diseases caused by fungi, phytoplasma and bacteria; Ciancio, A.; Mukerji, K. G., Eds.; Springer: Heidelberg, 2008; pp119–136.
- (4) Surico, G. Towards a redefinition of the diseases within the esca complex of grapevine. *Phytopathol. Mediterr.* **2009**, *48*, 5–10.
- (5) Letousey, P.; Baillieul, F.; Perrot, G.; Rabenoelina, C.; Boulay, M.; Vaillant-Gaveau, N.; Clément, C.; Fontaine, F. Early events prior to visual symptoms in apoplectic form of esca disease. *Phytopathology* **2010**, *100*, 424–431.
- (6) Larignon, P.; Fontaine, F.; Farine, S.; Clément, C.; Bertsch, C. Esca et Black Dead Arm: deux acteurs majeurs des maladies du bois chez la vigne. *CR Acad SCI III-Vie* **2009**, *333*, 765–783.
- (7) Petit, A. N.; Vaillant, N.; Boulay, M.; Clément, C.; Fontaine, F. Alteration of photosynthesis in grapevines affected by esca. *Phytopathology* **2006**, *96*, 1060–1066.
- (8) Magnin-Robert, M.; Letousey, P.; Spagnolo, A.; Rabenoelina, F.; Jacquens, L.; Mercier, L.; Clément, C.; Fontaine, F. Leaf stripe form of esca induces alteration of photosynthesis and defence reactions in presymptomatic leaves. *Func. Plant Biol.* **2011**, *38*, 856–866.
- (9) Valtaud, C.; Foyer, C. H.; Fleurat-Lessard, P.; Bourbouloux, A. Systemic effects on leaf glutathione metabolism and defence protein expression caused by esca infection in grapevines. *Func. Plant Biol.* **2009**, *36*, 260–279.
- (10) Spagnolo, A.; Magnin-Robert, M.; Alayi, T. D.; Cilindre, C.; Mercier, L.; Schaeffer-Reiss, C.; Van Dorsselaer, A.; Clément, C.; Fontaine, F. Physiological changes in green stems of *Vitis vinifera* L. cv. Chardonnay in response to esca proper and apoplexy revealed by proteomic and transcriptomic analyses. *J. Proteome Res.* **2012**, *11*, 461–475.
- (11) Mugnai, L.; Graniti, A.; Surico, G. Esca (black measles) and brown wood-streaking: two old and elusive diseases in grapevines. *Plant Dis.* **1999**, *83*, 404–418.
- (12) Spagnolo, A.; Marchi, G.; Peduto, F.; Phillips, A. J. L.; Surico, G. Detection of Botryosphaeriaceae species within grapevine woody tissues, with particular emphasis on the *Neofusicoccum parvum*/*N. ribis* complex. *Eur. J. Plant Pathol.* **2011**, *129*, 485–500.
- (13) Crous, P. W.; Gams, W. *Phaeomoniella chlamydospora* gen. et comb. Nov., a causal organism of Petri grapevine decline and esca. *Phytopathol. Mediterr.* **2000**, *39*, 112–118.
- (14) Essakhi, S.; Mugnai, L.; Crous, P. W.; Groenewald, J. Z.; Surico, G. Molecular and phenotypic characterization of novel *Phaeoacremonium* species isolated from esca diseased grapevines. *Persoonia* **2008**, *21*, 119–134.
- (15) Fischer, M. A. A new wood-decaying basidiomycete species associated with esca of grapevine: *Fomitiporia mediterranea* (Hymenochaetales). *Mycol. Prog.* **2002**, *1*, 315–324.
- (16) Phillips, A. J. L. *Botryosphaeria* species associated with diseases of grapevines in Portugal. *Phytopathol. Mediterr.* **2002**, *41*, 3–18.
- (17) Crous, P. W.; Slippers, B.; Wingfield, M. J.; Rheeder, J.; Marasas, W. F. O.; Phillips, A. J. L.; Alves, A.; Burgess, T.; Barber, P.; Groenewald, J. Z. Phylogenetic lineages in the Botryosphaeriaceae. *Stud. Mycol.* **2006**, *55*, 235–253.
- (18) Carter, M. V. Wood and root diseases caused by fungi. In *Compendium of grape diseases*, 3rd ed.; Pearson, R. C.; Goheen, A. C., Eds; APS Press: St-Paul, MN, 1994; pp 32–34.
- (19) 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*, 122–124.
- (20) Elias, J. E.; Gygi, S. P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* **2007**, *4*, 207–214.
- (21) Vizcaíno, J. A.; Côté, R.; Reisinger, F.; Barsnes, H.; Foster, J. M.; Rameseder, J.; Hermjakob, H.; Martens, L. The Proteomics Identifications database: 2010 update. *Nucleic Acids Res.* **2010**, *38*, D736–D742.
- (22) Bézier, A.; Lambert, B.; Baillieul, F. Study of defence-related gene expression in grapevine leaves and berries infected with *Botrytis cinerea*. *Eur. J. Plant Pathol* **2002**, *108*, 111–120.
- (23) Kuntzmann, P.; Villaume, S.; Larignon, P.; Bertsch, C. Esca, BDA and Eutypiosis: foliar symptoms, trunk lesions and fungi observed in diseased vinestocks in two vineyards in Alsace. *Vitis* **2010**, *49*, 71–76.

- 609 (24) Charmont, S.; Jamet, E.; Pont-Lezica, R.; Canut, H. Proteomic analysis of secreted proteins from
610 *Arabidopsis thaliana* seedlings: improved recovery following removal of phenolic compounds.
611 *Phytochemistry* **2005**, *66*, 453–461.
- 612 (25) Allan, W. L.; Clark, S. M.; Hoover, G. J.; Shelp, B. J. Role of plant glyoxylate reductases during stress: a
613 hypothesis. *Biochem. J.* **2009**, *423*, 15–22.
- 614 (26) Berger, S.; Papadopoulos, M.; Schreiber, U.; Kaiser, W.; Roitsch, T. Complex regulation of gene
615 expression, photosynthesis and sugar levels by pathogen infection in tomato. *Physiol. Plantarum* **2004**,
616 *122*, 419–428.
- 617 (27) Ehness, R.; Ecker, M.; Godt, D. E.; Roitsch, T. Glucose and stress independently regulate source and sink
618 metabolism and defense mechanisms via signal transduction pathways involving protein phosphorylation.
619 *Plant Cell* **1997**, *9*, 1825–1841.
- 620 (28) Herbers, K.; Meuwly, P.; Frommer, W. B.; Métraux, J. P.; Sonnewald, C. Systemic acquired resistance
621 mediated by the ectopic expression of invertase: possible hexose sensing in the secretory pathway. *Plant*
622 *Cell* **1996**, *8*, 793–803.
- 623 (29) Kim, M.; Lim, J. H.; Ahn, C. S.; Park, K.; Kim, G. T.; Kim, W. T.; Pai, H. S. Mitochondria-associated
624 hexokinases play a role in the control of programmed cell death in *Nicotiana benthamiana*. *Plant Cell*
625 **2006**, *18*, 2341–2355.
- 626 (30) Wu, K.; Atkinson, I. J.; Cossins, E. A.; King, J. Methotrexate resistance in *Datura innoxia* (uptake
627 metabolism of methotrexate in wil-type and resistant cell lines). *Plant Physiol.* **1993**, *101*, 477–483.
- 628 (31) Schlosser, J.; Olsson, N.; Weis, M.; Reid, K.; Peng, F.; Lund, S.; Bowen, P. Cellular expansion and gene
629 expression in the developing grape (*Vitis vinifera* L.). *Protoplasma* **2008**, *232*, 255–265.
- 630 (32) Ferrario-Méry, S.; Besin, E.; Picho, O.; Meyer, C.; Hodges, M. The regulatory PII protein controls arginine
631 biosynthesis in *Arabidopsis*. *FEBS Letters* **2006**, *580*, 2015–2020.
- 632 (33) Deuschle, K.; Funck, D.; Forlani, G.; Stransky, H.; Biehl, A.; Leister, D.; Vand der Graff, E.; Kunze, R.;
633 Frommer, W. B. The role of [Delta]1-pyrrolidone-5-carboxylate dehydrogenase in proline degradation.
634 *Plant Cell* **2004**, *16*, 3413–3425.
- 635 (34) Mitchell, H.; Ayliffe, M.; Rashid, K.; Pryor, A. A rust-inducible gene from flax (*fls1*) is involved in proline
636 catabolism. *Planta* **2006**, *23*, 213–222.
- 637 (35) Raj Shetty, S. N.; Shetty, H. S. Proline: an inducer of resistance against pearl millet downy mildew disease
638 caused by *Sclerospora graminicola*. *Phytoparasitica* **2004**, *32*, 523–527.
- 639 (36) Bélanger, R. R.; Manion, P. D.; Griffin, D. H. Amino acid content of water-stressed plantlets of *Populus*
640 *tremuloides* clones in relation to clonal susceptibility to *Hypoxyylon mammatum* *in vitro*. *Can. J. Bot.* **1990**,
641 *68*, 26–29.
- 642 (37) Yang, L.; Lin, H.; Takahashi, Y.; Chen, F.; Walker, M. A.; Civerolo, E. I. Proteomic analysis of grapevine
643 stem in response to *Xylella fastidiosa* inoculation. *Physiol. Mol. Plant Pathol.* **2011**, *75*, 90–99.
- 644 (38) Golkari, S.; Gilbert, J.; Prashar, S.; Procnier, J. D. Microarray analysis of *Fusarium graminearum*-induced
645 wheat genes: identification of organ-specific and differentially expressed genes. *Plant Biotechnol. J.* **2007**,
646 *5*, 38–49.
- 647 (39) Margaria, P.; Palmano, S. Response of the *Vitis vinifera* L. cv. 'Nebbiolo' proteome to Flavescence dorée
648 phytoplasma infection. *Proteomics* **2011**, *11*, 212–214.
- 649 (40) Vincent, D.; Ergül, A.; Bohlman, C.; Tattersall, E. A. R.; Tillett, R. L.; Wheatley, M. D.; Woolsey, R.;
650 Quilici, D. R.; Joets, J.; Schlauch, K.; Schooley, D. A.; Cushman, J. C.; Cramer, G. R. Proteomic analysis
651 reveals differences between *Vitis vinifera* L. cv. Chardonnay and cv. Cabernet Sauvignon and their
652 responses to water deficit and salinity. *J. Exp. Bot.* **2007**, *58*, 1873–1892.
- 653 (41) Suty, L.; Lequeu, J.; Lançon, A.; Etienne, P.; Petitot, A. S.; Blein, J. P. Preferential induction of 20S
654 proteasome subunits during elicitation of plant defence reactions: towards the characterization of « plant
655 defence proteasomes ». *Int. J. Biochem. Cell B* **2003**, *35*, 637–650.
- 656 (42) Al-Whaibi, M. H. Plant heat-shock proteins: a mini review. *J. King. Saud. Univ.* **2011**, *23*, 139–150.
- 657 (43) Polesani, M.; Desario, F.; Ferrarini, A.; Zamboni, A.; Pezzoti, M.; Kortekamp, A.; Polverari, A. cDNA-
658 AFLP analysis of plant and pathogen genes expressed in grapevine with *Plasmopara viticola*. *BMC*
659 *Genomics* **2008**, *9*, 142–155.

- 660 (44) Banzet, N.; Richard, C.; Desveaux, Y.; Kazmaizer, M.; Gagnon, J.; Triantaphylidès, C. Accumulation of
661 small heat shock proteins, including mitochondrial HSP22, induced by oxidative stress and adaptative
662 response in tomato cells. *Plant J.* **1998**, *13*, 519–527.
- 663 (45) Vanderauwera, S.; Zimmermann, P.; Rombauts, S.; Vandenabeele, S.; Langebartels, C.; Grisse, W.;
664 Inzé, D.; Van Breusegem, F. Genome-wide analysis of hydrogen peroxide regulated gene expression in
665 *Arabidopsis* reveals a high light-induced transcriptional cluster involved in anthocyanin biosynthesis. *Plant*
666 *Physiol.* **2005**, *139*, 806–821.
- 667 (46) Fung, R. W. M.; Gonzalo, M.; Fekete, C.; Kovacs, L. G.; He, Y.; Marsch, E.; McIntyre, L. M.;
668 Schachtman, D. P.; Qiu, W. Powdery mildew induces defence-oriented reprogramming of the
669 transcriptome in a susceptible but not in a resistant grapevine. *Plant Physiol.* **2008**, *146*, 236–249.
- 670 (47) Camps, C.; Kappel, C.; Lecomte, P.; Léon, C.; Gomès, E.; Coutos-Thévenot, P.; Delrot, S. A
671 transcriptomic study of grapevine (*Vitis vinifera* cv. Cabernet-Sauvignon) interaction with the vascular
672 ascomycete fungus *Eutypa lata*. *J. Exp. Bot.* **2010**, *61*, 1719–1737.
- 673 (48) Figueiredo, A.; Fortes, A. M.; Ferreira, S.; Sebastiana, M.; Choi, Y. H.; Sousa, L.; Acioli-Santos, B.;
674 Pessoa, F.; Verpoorte, R.; Pais, M. S. Transcriptional and metabolic profiling of grape (*Vitis vinifera* L.)
675 leaves unravel possible innate resistance against pathogenic fungi. *J. Exp. Bot.* **2008**, *59*, 3371–3381.
- 676 (49) Frova, C. The plant glutathione transferase gene family: genomic structure, functions, expression and
677 evolution. *Physiol. Plantarum* **2003**, *119*, 469–479.
- 678 (50) Guillen, P.; Guis, M.; Martinez-Reina, G.; Colrat, S.; Dalmeyrac, S.; Bouzayen, M.; Roustan, J.-P.; Fallot,
679 J.; Pech, J.-C.; Latché, A. A novel NADPH-dependent aldehyde reductase gene from *Vigna radiata* confers
680 resistance to the grapevine fungal toxin eutypine. *Plant J.* **1998**, *16*, 335–343.
- 681 (51) Andolfi, A.; Mugnai, L.; Luque, J.; Surico, G.; Cimmino, A.; Evidente, A. Phytotoxins produced by fungi
682 associated with grapevine trunk diseases. *Toxins* **2011**, *3*, 1569–1605.
- 683 (52) Després, C.; Chubak, C.; Rochon, A.; Clark, R.; Bethune, T.; Desveaux, D.; Fobert, P. R. The *Arabidopsis*
684 NPR1 disease resistance protein is a novel cofactor that confers redox regulation of DNA binding activity
685 to the basic domain/leucine zipper transcription factor TGA1. *Plant Cell* **2003**, *15*, 2181–2191.
- 686 (53) Mou, Z.; Fan, W.; Dong, X. Inducers of plant systemic acquired resistance regulate NPR1 function through
687 redox changes. *Cell* **2003**, *27*, 935–944.
- 688 (54) Lindemayr, C.; Sell, S.; Müller, B.; Durner, J. Redox regulation of the NPR1-TGA1 system of *Arabidopsis*
689 *thaliana* by nitric oxide. *Plant Cell* **2010**, *22*, 2894–2907.
- 690 (55) Wu, F.; Yan, M.; Li, Y.; Chang, S.; Song, X.; Zhou, Z.; Gong, W. cDNA cloning, expression, and
691 mutagenesis of a PR-10 protein SPE-16 from the seeds of *Pachyrrhizus erosus*. *Biochem. Biophys. Res.*
692 *Commun.* **2003**, *312*, 761–766.
- 693 (56) Margaria, P.; Abbà, S.; Palmani, S. Novel aspects of grapevine response to phytoplasma infection
694 investigated by a proteomic and phospho-proteomic approach with data integration into functional
695 networks. *BMC Genomics* **2013**, *14*, 38–52.
- 696 (57) Lima, M. R. M.; Ferreres, F.; Dias, A. C. P. Response of *Vitis vinifera* cell cultures to *Phaeoemoniella*
697 *chlamydospora*: changes in phenolic production, oxidative state and expression of defence-related genes.
698 *Eur. J. Plant Pathol.* **2012**, *132*, 133–146.
- 699 (58) Afroz, A.; Ali, G. M.; Mir, A.; Komatsu, S. Application of proteomics to investigate stress-induced
700 proteins for improvement in crop protection. *Plant Cell Rep.* **2011**, *30*, 745–763.
- 701 (59) Monteiro, S.; Barakat, M.; Piçarra-Pereira, M. A.; Teixeira, A. R.; Ferreira, R. B. Osmotin and thaumatin
702 from grape: a putative general defense mechanism against pathogenic fungi. *Phytopathology* **2003**, *93*,
703 1505–1512.
- 704 (60) Colditz, F.; Niehaus, K.; Krajinski, F. Silencing of PR-10-like proteins in *Medicago truncatula* results in an
705 antagonistic induction of other PR protein and in an increased tolerance upon infection with the oomycete
706 *Aphanomyces euteiches*. *Planta* **2007**, *226*, 57–71.
- 707 (61) La Camera, S.; Gouzerth, G.; Dhondt, S.; Hoffmann, L.; Fritig, B.; Legrand, M.; Heitz, T. Metabolic
708 reprogramming in plant innate immunity: the contributions of phenylpropanoid and oxylipin pathways.
709 *Immunol. Rev.* **2004**, *198*, 267–284.
- 710 (62) Roje, S. S-Adenosyl-L-methionine: beyond the universal methyl group donor. *Phytochemistry* **2006**, *67*,
711 1686–1698.

- 712 (63) Dixon, R. A.; Achnine, L.; Kota, P.; Liu, C. J.; Reddy, M. S. S.; Wang, L. The phenylpropanoid pathway
713 and plant defence-a genomics perspective. *Mol. Plant Pathol.* **2002**, *3*, 371–390.
- 714 (64) Osmark, P.; Boyle, B.; Brisson, N. Sequential and structural homology between intracellular pathogenesis-
715 related proteins and a group of latex proteins. *Plant Mol. Biol.* **1998**, *38*, 1243–1246.
- 716 (65) Amey, R. C.; Schleicher, T.; Slinn, J.; Lewis, M.; McDonald, H.; Neill, S. J.; Spencer-Phillips, P. T. N.
717 Proteomic analysis of a compatible interaction between *Pisum sativum* (pea) and the downy mildew
718 pathogen *Peronospora viciae*. *Eur. J. Plant Pathol.* **2008**, *122*, 41–55.
- 719 (66) Roberts, M. R.; Salinas, J.; Collinge, D. B. 14-3-3 proteins and the response to abiotic and biotic stress.
720 *Plant Mol. Biol.* **2002**, *50*, 1031–1039.
- 721 (67) Del Rio, J. A.; Gonzalez, A.; Fuster, M. D.; Botia, J. M.; Gomez, P.; Frias, V.; Ortuño, A. Tylose formation
722 and changes in phenolic compounds of grape roots infected with *Phaeoconiella chlamyospora* and
723 *Phaeoacremonium* species. *Phytopathol. Mediterr.* **2001**, *40*, 394–399.
- 724 (68) Maimbo, M.; Ohnishi, K.; Hikichi, Y.; Yoshioka, H.; Kiba, A. Induction of a small heat shock protein and
725 its functional roles in *Nicotiana* plants in the defence response against *Ralstonia solanacearum*. *Plant*
726 *Physiol.* **2007**, *145*, 1588–1599.
- 727 (69) Andolfi, A.; Cimmino, A.; Evidente, A.; Iannaccone, M.; Capparelli, R.; Mugnai, L.; Surico, G. A new
728 flow cytometry technique to identify *Phaeoconiella chlamyospora* exopolysaccharides and study
729 mechanisms of esca grapevine foliar symptoms. *Plant Dis.* **2009**, *90*, 1490–1503.
- 730

731

732
733
734

Table 1. Identified proteins differentially expressed in the black streaked or asymptomatic trunk wood of control, apoplectic and esca proper affected plants.

Spot ^a	Mw/pI exp. ^b	Mw/pI theor. ^c	Matched Protein ^d	Number of unique peptides ^e	Total number of peptides ^f	Coverage % ^g	Accession No. ^h
1007	11.00/4.40	18.98/4.84	PREDICTED: uncharacterized protein LOC100854560 [<i>Vitis vinifera</i>]	2	2	11	gi 359496362
1105	19.50/4.85	24.03/4.94	putative thaumatin-like protein [<i>Vitis</i> hybrid cultivar]	7	7	39	gi 7406716
1202	24.00/4.80	28.76/4.80	PREDICTED: 14-3-3-like protein [<i>Vitis vinifera</i>]	4	4	20	gi 359492927
1202	24.00/4.80	27.51/5.38	class IV endochitinase [<i>Vitis vinifera</i>]	1	1	6	gi 2306813
1208	29.50/4.18	31.54/6.59	PREDICTED: GEM-like protein 5 [<i>Vitis vinifera</i>]	1	1	4	gi 225468805
2211	22.50/5.12	20.90/4.83	PREDICTED: salt stress root protein RS1 isoform 3 [<i>Vitis vinifera</i>]	2	2	16	gi 225424944
2211	22.50/5.12	26.08/5.22	PREDICTED: endo-1,3;1,4-beta-D-glucanase [<i>Vitis vinifera</i>]	2	2	11	gi 225436938
2212	23.00/4.90	29.31/5.05	PREDICTED: acid phosphatase 1-like [<i>Vitis vinifera</i>]	3	3	17	gi 225445051
2212	23.00/4.90	24.03/4.80	PREDICTED: 14-3-3 protein 1-like [<i>Vitis vinifera</i>]	12	12	56	gi 359492927
2302	40.25/5.00	48.08/5.51	PREDICTED: transaldolase isoform 1 [<i>Vitis vinifera</i>]	10	10	30	gi 225425280
3105	18.75/5.30	25.01/6.35	PREDICTED: small heat shock protein, chloroplastic [<i>Vitis vinifera</i>]	6	6	24	gi 225455238
3107	18.12/5.45	25.01/6.35	PREDICTED: small heat shock protein, chloroplastic [<i>Vitis vinifera</i>]	5	5	24	gi 225455238
3202	17.50/4.90	25.01/6.35	PREDICTED: small heat shock protein, chloroplastic [<i>Vitis vinifera</i>]	10	10	45	gi 225455238
3203	17.19/5.05	25.01/6.35	PREDICTED: small heat shock protein, chloroplastic [<i>Vitis vinifera</i>]	8	8	43	gi 225455238
3409	45.12/5.40	49.65/4.95	PREDICTED: tubulin alpha-1 chain [<i>Vitis vinifera</i>]	15	15	47	gi 225435758
3409	45.12/5.40	52.45/5.51	PREDICTED: hexokinase-2, chloroplastic [<i>Vitis vinifera</i>]	15	15	40	gi 225457987
3409	45.12/5.40	54.45/5.71	PREDICTED: mitochondrial-processing peptidase subunit alpha [<i>Vitis vinifera</i>]	8	8	22	gi 225445041
4004	12.45/5.60	17.34/5.94	PREDICTED: 17.9 kDa class II heat shock protein [<i>Vitis vinifera</i>]	3	5	44	gi 225429604
4004	12.45/5.60	15.62/10.58	PREDICTED: 60S ribosomal protein L27-3 isoform 1 [<i>Vitis vinifera</i>]	1	1	8	gi 225423704
4102	12.60/5.36	21.71/5.87	PREDICTED: superoxide dismutase [Cu-Zn], chloroplastic [<i>Vitis vinifera</i>]	4	4	30	gi 225436450
4102	12.60/5.36	17.09/5.96	pathogenesis-related protein 10 [<i>Vitis</i> hybrid cultivar]	2	2	18	gi 163914213
4212	31.75/5.78	33.82/5.50	TPA: isoflavone reductase-like protein 4 [<i>Vitis vinifera</i>]	10	15	45	gi 76559892
4216	22.55/5.59	25.06/5.26	Chalcone--flavonone isomerase 2 [<i>Vitis vinifera</i>]	7	7	39	gi 158514257
4216	22.55/5.59	24.51/5.48	PREDICTED: proteasome subunit alpha type-2-B [<i>Vitis vinifera</i>]	2	2	14	gi 225423722
4221	20.70/5.78	24.89/5.24	PREDICTED: glutathione S-transferase F9 [<i>Vitis vinifera</i>]	10	10	54	gi 225446791
4221	20.70/5.78	25.23/5.67	PREDICTED: stem-specific protein TSJT1 [<i>Vitis vinifera</i>]	7	7	48	gi 225432548
4221	20.70/5.78	25.31/5.85	hypothetical protein VITISV_038846 [<i>Vitis vinifera</i>]	6	6	35	gi 147784683
4221	20.70/5.78	24.90/5.31	PREDICTED: proteasome subunit beta type-6 [<i>Vitis vinifera</i>]	1	2	12	gi 225429850
5215	21.25/6.06	25.23/5.67	PREDICTED: stem-specific protein TSJT1 [<i>Vitis vinifera</i>]	7	7	47	gi 225432548
5215	21.25/6.06	24.89/5.24	PREDICTED: glutathione S-transferase F9 [<i>Vitis vinifera</i>]	10	10	51	gi 225446791
5316	31.75/5.97	33.82/5.50	TPA: isoflavone reductase-like protein 4 [<i>Vitis vinifera</i>]	10	14	45	gi 76559892
5316	31.75/5.97	29.22/5.74	PREDICTED: dihydrofolate reductase [<i>Vitis vinifera</i>]	3	3	14	gi 225440230

735

Spot ^a	Mw/pI exp. ^b	Mw/pI theor. ^c	Matched Protein ^d	Number of unique peptides ^e	Total number of peptides ^f	Coverage % ^g	Accession No. ^h
5422	41.87/6.11	39.17/5.39	PREDICTED: peroxidase 12-like [<i>Vitis vinifera</i>]	10	10	36	gi 359493149
5422	41.87/6.11	43.56/5.63	PREDICTED: bifunctional polymyxin resistance protein ArnA [<i>Vitis vinifera</i>]	3	3	11	gi 225436371
5422	41.87/6.11	44.75/10.42	PREDICTED: 60S ribosomal protein L4 [<i>Vitis vinifera</i>]	6	6	21	gi 225439761
5422	41.87/6.11	42.91/5.49	S-adenosylmethionine synthase 4	2	2	6	gi 223635287
6001	11.87/6.08	17.34/5.94	PREDICTED: 17.9 kDa class II heat shock protein [<i>Vitis vinifera</i>]	5	7	60	gi 225429604
6001	11.87/6.08	21.82/9.20	PREDICTED: nitrogen regulatory protein P-II homolog [<i>Vitis vinifera</i>]	4	4	24	gi 225437667
6002	10.65/6.10	16.31/6.32	PREDICTED: glycine-rich RNA-binding protein GRP1A-like [<i>Vitis vinifera</i>]	7	7	57	gi 359475330
6002	10.65/6.10	15.26/5.49	PREDICTED: superoxide dismutase [Cu-Zn] isoform 2 [<i>Vitis vinifera</i>]	3	3	30	gi 225451120
6101	13.12/6.00	17.34/5.94	PREDICTED: 17.9 kDa class II heat shock protein [<i>Vitis vinifera</i>]	1	2	20	gi 225429604
6101	13.12/6.00	17.09/5.96	pathogenesis-related protein 10 [<i>Vitis</i> hybrid cultivar]	3	3	25	gi 163914213
6201	22.70/6.18	24.61/6.88	PREDICTED: proteasome subunit beta type-1 [<i>Vitis vinifera</i>]	3	3	28	gi 225453909
6201	22.70/6.18	25.24/5.67	PREDICTED: stem-specific protein TSJT1 [<i>Vitis vinifera</i>]	4	4	27	gi 225432548
6201	22.70/6.18	23.67/5.79	PREDICTED: glutathione S-transferase DHAR2 [<i>Vitis vinifera</i>]	4	4	29	gi 225448831
6201	22.70/6.18	43.73/8.09	PREDICTED: malate dehydrogenase, chloroplastic-like [<i>Vitis vinifera</i>]	2	2	8	gi 225452831
6203	19.37/6.16	21.72/5.81	PREDICTED: flavoprotein wrbA-like [<i>Vitis vinifera</i>]	5	5	31	gi 359496220
6203	19.37/6.16	17.09/5.96	pathogenesis-related protein 10 [<i>Vitis</i> hybrid cultivar]	2	2	18	gi 163914213
6203	19.37/6.16	31.52/7.71	PREDICTED: ATP-dependent Clp protease proteolytic subunit 5, chloroplastic isoform 1 [<i>Vitis vinifera</i>]	4	4	12	gi 225437581
6420	42.69/6.22	43.11/5.65	S-adenosylmethionine synthase 2 [<i>Vitis vinifera</i>]	10	10	43	gi 223635284
6420	42.69/6.22	51.31/6.61	PREDICTED: probable ornithine aminotransferase [<i>Vitis vinifera</i>]	12	12	36	gi 225443914
6420	42.69/6.22	46.43/5.80	PREDICTED: eukaryotic initiation factor 4A-3-like [<i>Vitis vinifera</i>]	9	9	25	gi 225464928
6420	42.69/6.22	46.27/5.99	PREDICTED: isocitrate dehydrogenase [NADP] [<i>Vitis vinifera</i>]	2	2	7	gi 225466253
6420	42.69/6.22	40.17/5.54	leucoanthocyanidin dioxygenase [<i>Vitis amurensis</i>]	2	2	8	gi 224038270
7102	13.15/6.39	17.18/5.62	PREDICTED: MLP-like protein 28 [<i>Vitis vinifera</i>]	6	6	49	gi 225424272
7210	20.60/6.57	24.61/6.88	PREDICTED: proteasome subunit beta type-1 [<i>Vitis vinifera</i>]	5	5	38	gi 225453909
7210	20.60/6.57	24.94/5.78	PREDICTED: glutathione S-transferase F9 [<i>Vitis vinifera</i>]	8	9	42	gi 225446793
7319	31.37/6.32	33.87/5.76	TPA: isoflavone reductase-like protein 5 [<i>Vitis vinifera</i>]	14	14	54	gi 76559894
7319	31.37/6.32	17.18/5.62	PREDICTED: MLP-like protein 28 [<i>Vitis vinifera</i>]	2	2	22	gi 225424272
7319	31.37/6.32	34.28/5.49	PREDICTED: glyoxylate reductase isoform 1 [<i>Vitis vinifera</i>]	2	2	12	gi 359482924
7319	31.37/6.32	35.04/6.05	PREDICTED: isoflavone reductase homolog [<i>Vitis vinifera</i>]	1	1	4	gi 359491809
7319	31.37/6.32	34.46/9.00	PREDICTED: 60S ribosomal protein L5-like isoform 2 [<i>Vitis vinifera</i>]	2	2	6	gi 225434146
7319	31.37/6.32	30.02/9.31	PREDICTED: 60S ribosomal protein L5-like [<i>Vitis vinifera</i>]	1	2	11	gi 359495418
8102	14.25/6.81	17.09/5.96	pathogenesis-related protein 10 [<i>Vitis</i> hybrid cultivar]	7	7	50	gi 163914213
8102	14.25/6.81	17.08/10.44	PREDICTED: 40S ribosomal protein S15-like isoform 1 [<i>Vitis vinifera</i>]	2	2	24	gi 225437253

737
738
739

Table 1. Continued

Spot ^a	Mw/pI exp. ^b	Mw/pI theor. ^c	Matched Protein ^d	Number of unique peptides ^e	Total number of peptides ^f	Coverage % ^g	Accession No. ^h
8102	14.25/6.81	17.15/6.61	PREDICTED: transcription factor BTF3 [<i>Vitis vinifera</i>]	2	2	24	gi 225457458
8102	14.25/6.81	17.81/8.81	PREDICTED: 60S ribosomal protein L12 [<i>Vitis vinifera</i>]	4	4	32	gi 225449052
8102	14.25/6.81	20.86/9.97	PREDICTED: 60S ribosomal protein L11-1-like [<i>Vitis vinifera</i>]	1	1	5	gi 225443760

740

741

^a spot code as reported in Figures 3, 4 and S2.

742

^b deduced spot molecular mass (kDa) and isoelectric point as observed on the gels.

743

^c theoretical molecular mass (kDa) and isoelectric point computed using the compute Mw/pI tool in ExPASy Bioinformatics Resource Portal.

744

^d protein identified via the MASCOT and OMSSA search engines against in house made database from NCBI database .

745

^e number of unique peptides matching the protein sequence.

746

^f total number of peptides matching the protein sequence.

747

^g percentage of the protein sequence covered by the matching peptides.

748

^h accession No. of the matched protein as reported in the NCBI database.

749

750

751

752

753

754

755

756

757

758

759

760

761

762

763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807

Figure captions

Figure 1. Functional classification of identified proteins from the trunk of fieldgrown *V. vinifera* affected or not by the grapevine trunk disease. The proteins were classified into 15 functional groups by using GenomeNet Database Resources website (<http://www.genome.jp/kegg>) and reports in the literature.

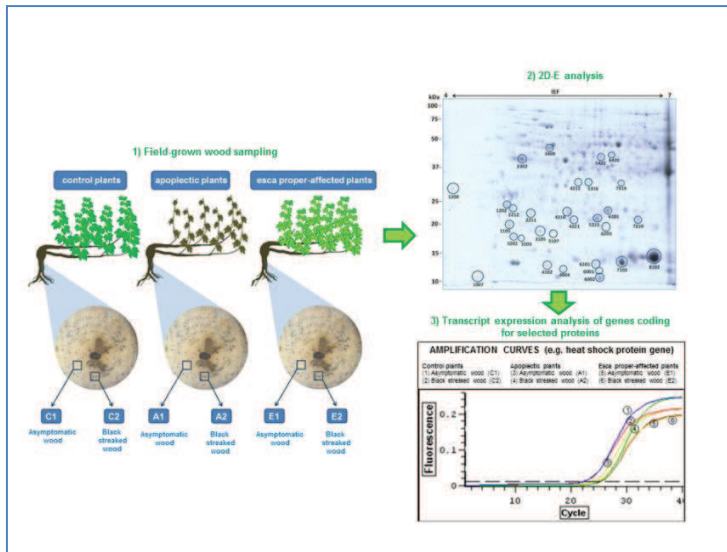
Figure 2. Pairwise comparison. (A) Pairwise comparison for all the possible pairs of group samples in one spot performed for all the spots (150 total comparisons per group). The number of times where the relative expression value of each group was considered to be similar to another (ratio comprised between 0.5 and 1.5) is reported. Asymptomatic (C1, A1, E1) and black streaked (C2, A2, E2) trunk wood of control (C1, C2), apoplectic (A1, A2) and esca proper affected (E1, E2) plants. (B) Pairwise comparison involving samples from asymptomatic (C1, A1, E1) and black streaked (C2, A2, E2) trunk wood performed for all the spots (30 total comparisons per group). The number of times (spots) where their relative expression was considered to be similar (ratio comprised between 0.5 and 1.5) is reported.

Figure 3. Hierarchical tree clustering based on relative expression values of the 30 differentially expressed protein spots identified in the groups of asymptomatic (C1, A1, E1) and black streaked wood (C2, A2, E2) from control (C1, C2), apoplectic (A1, A2) and esca proper affected plants (E1, E2). The relative value of optical density×area (relative OD×area%) of the total value observed in the gel images was considered as measure of expression. Relative mean values are from three relative values obtained by dividing the value of each spot to the total density of the corresponding gel. Hierarchical tree clustering was performed using the NCSS 8 software by NCSS Statistical Analysis & Graphics (<http://www.ncss.com/>). Clustering parameters were set up as follows: group average (unweighted pair-group) linkage type, Euclidean distance method, average absolute deviation scaling method, cluster cutoff at 0.01. Numbered parenthesis indicates clusters from highest values (cluster 1) to the smallest (cluster 5). Column a: spot number; column b: relative expression value.

Figure 4. Quantification of differentially expressed proteins extracted from asymptomatic (C1, A1, E1) and black streaked (C2, A2, E2) trunk wood of control (C1, C2), apoplectic (A1, A2) and esca proper affected (E1, E2) vines. Vertical axis represents the relative value of optical density×area (relative OD×area%) of the total value observed in the gel images as a measure of expression. Histograms indicate the mean quantitative values (\pm standard deviation SD) of three independent experiments. Relative mean values (\pm SD) are from three relative values obtained by dividing the value of each spot to the total density of the corresponding gel. Differences among the means were evaluated by the Dunn's Multiple Comparison Test after that the null hypothesis (equal means) was rejected in the Kruskal-Wallis test, assuming a significance of $p \leq 0.05$. One asterisk indicates a statistically significant difference with the control C1. Two or three asterisks indicate a statistically significant difference between the two classes where they were added. The relative expression ratio to C1 in the other groups is reported.

Figure 5. Expression levels of eleven selected genes in asymptomatic (C1, A1 and E1) and black streaked trunk wood (C2, A2 and E2) of control (C), apoplectic (A) and esca proper affected (E) 26 year-old standing vines cultivar Chardonnay by quantitative reverse-transcription polymerase chain reaction. Results correspond to means (\pm standard deviation) of three independent experiences (C1, C2, A1, A2, E1 and E2). Gene expression was considered as significantly up- or down-regulated to the $1 \times$ controls, when changes in relative expression were $> 2 \times$ or $< 0.5 \times$, respectively.

808
809
810
811
812
813
814
815
816
817
818
819
820
821
822



Total proteins and RNA were extracted from asymptomatic (C1, A1, E1) and black streaked (C2, A2, E2) trunk wood of control (C1, C2), apoplectic (A1, A2) and esca proper affected (E1, E2) vines. After 2D-E analyses, 153 differentially expressed proteins were identified by nanoLC-MS/MS. Abundance in the different samples and function of 53 of them were discussed. The mRNA expression of 11 of them was also evaluated by qRT-PCR.

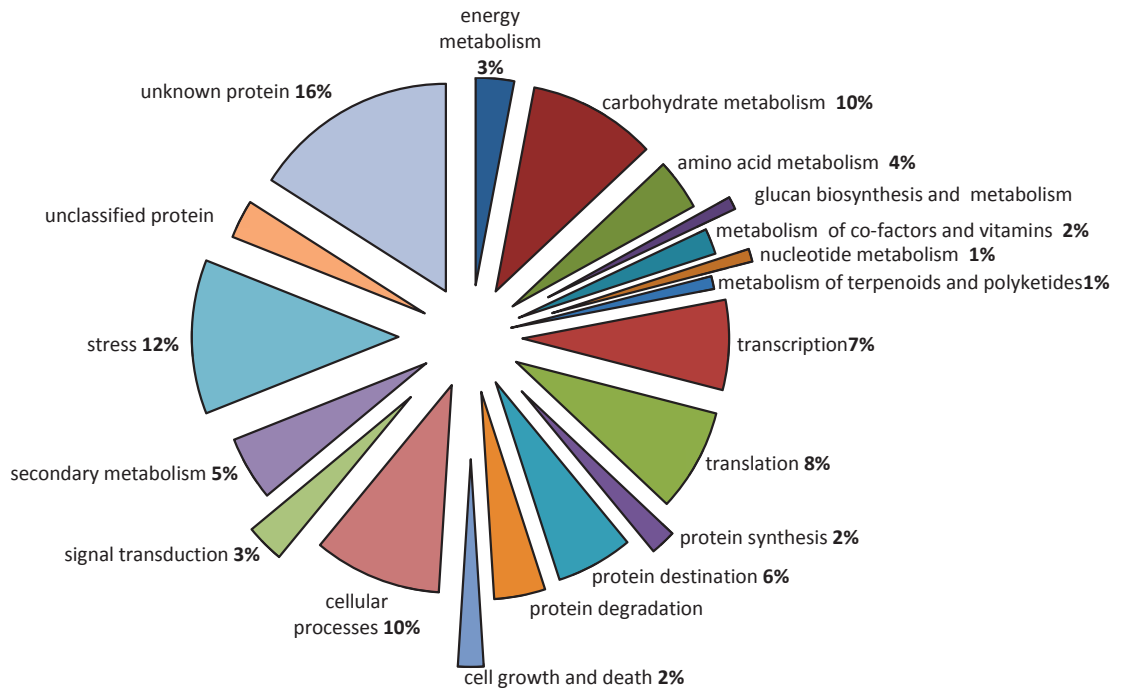


Figure 1.

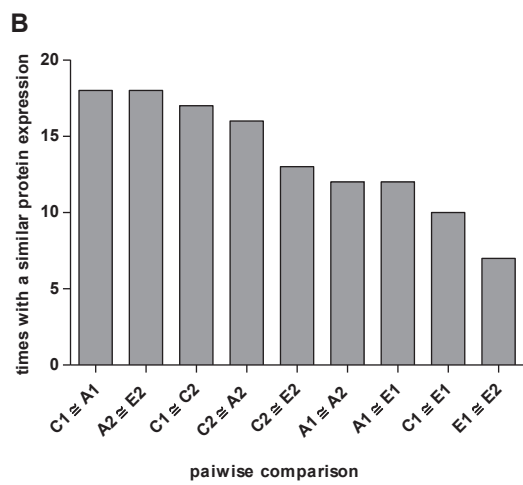
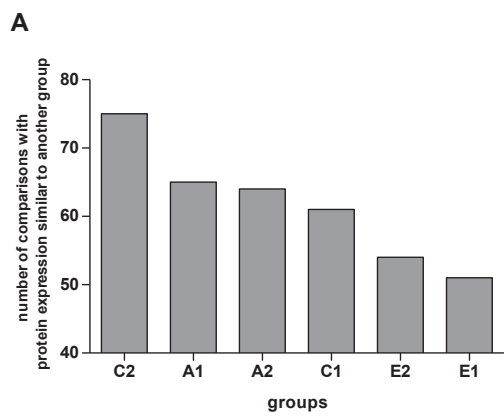


Figure 2.

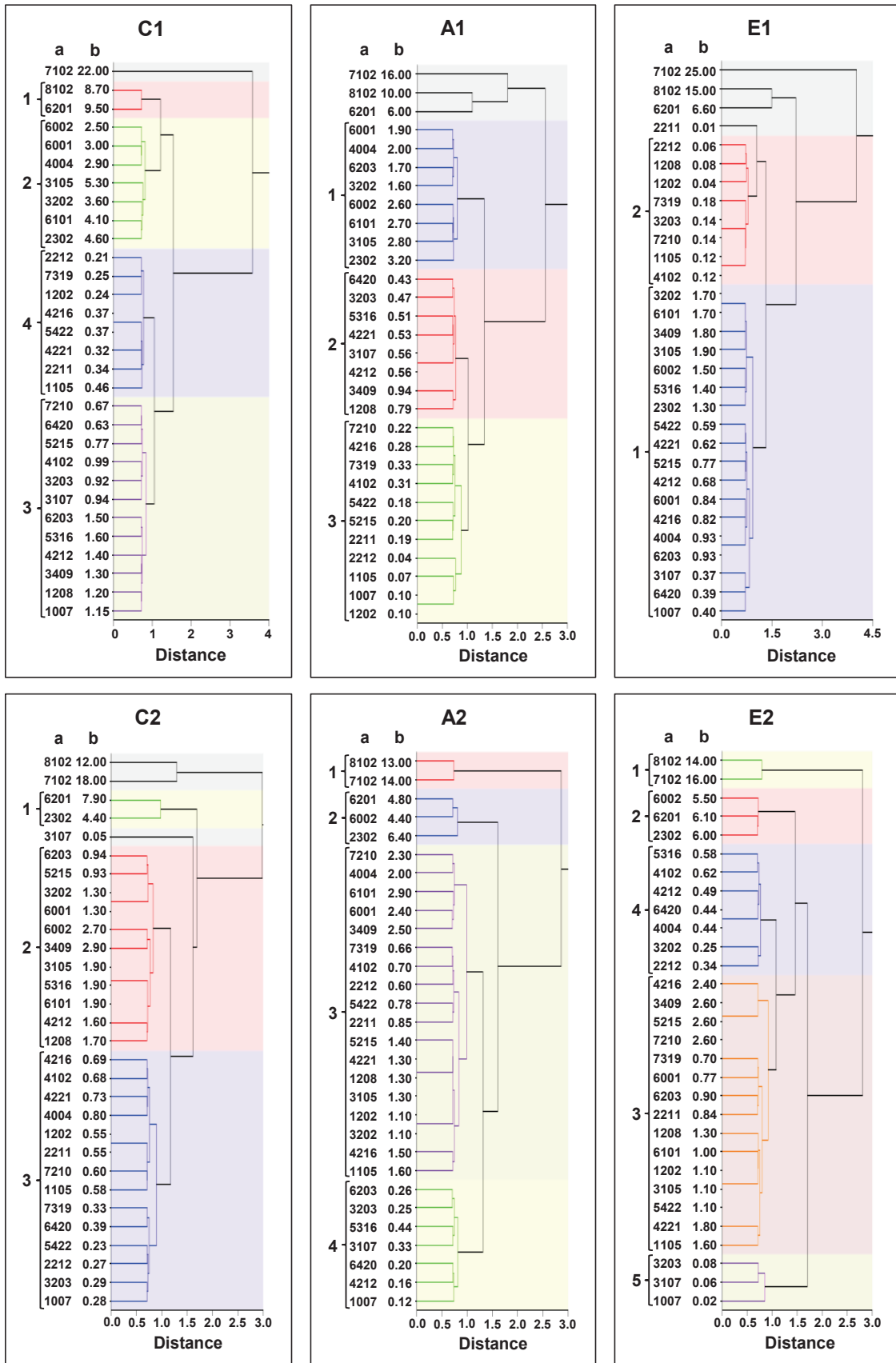


Figure 3.



Figure 4.

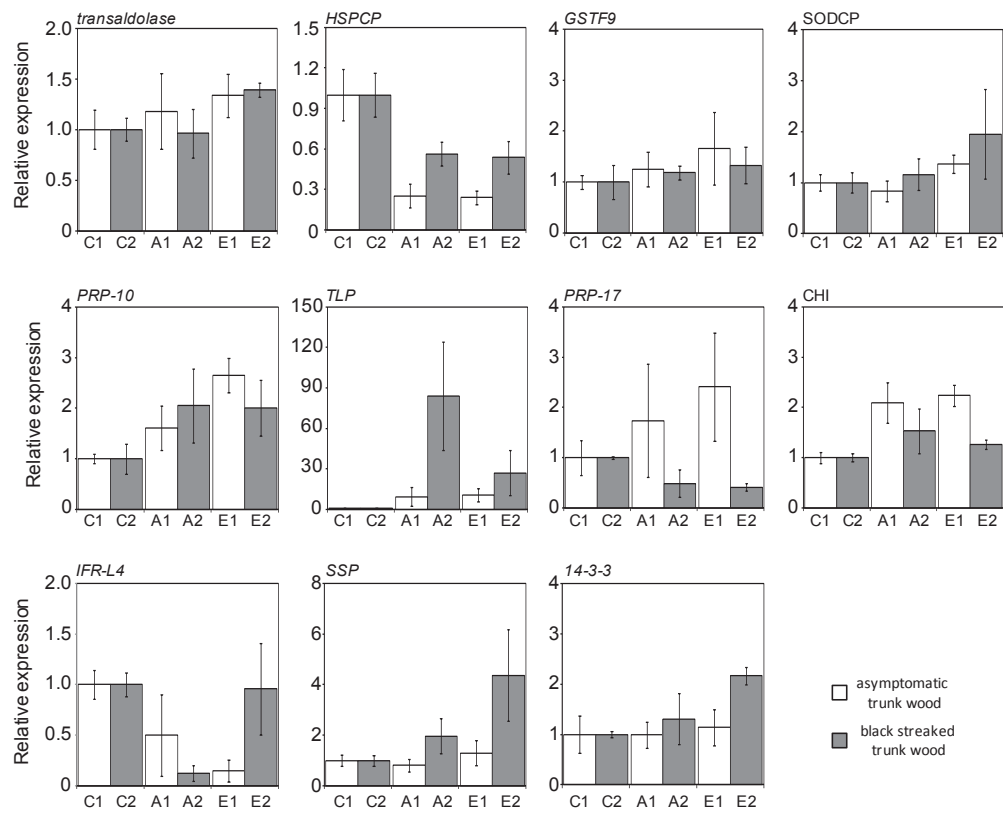
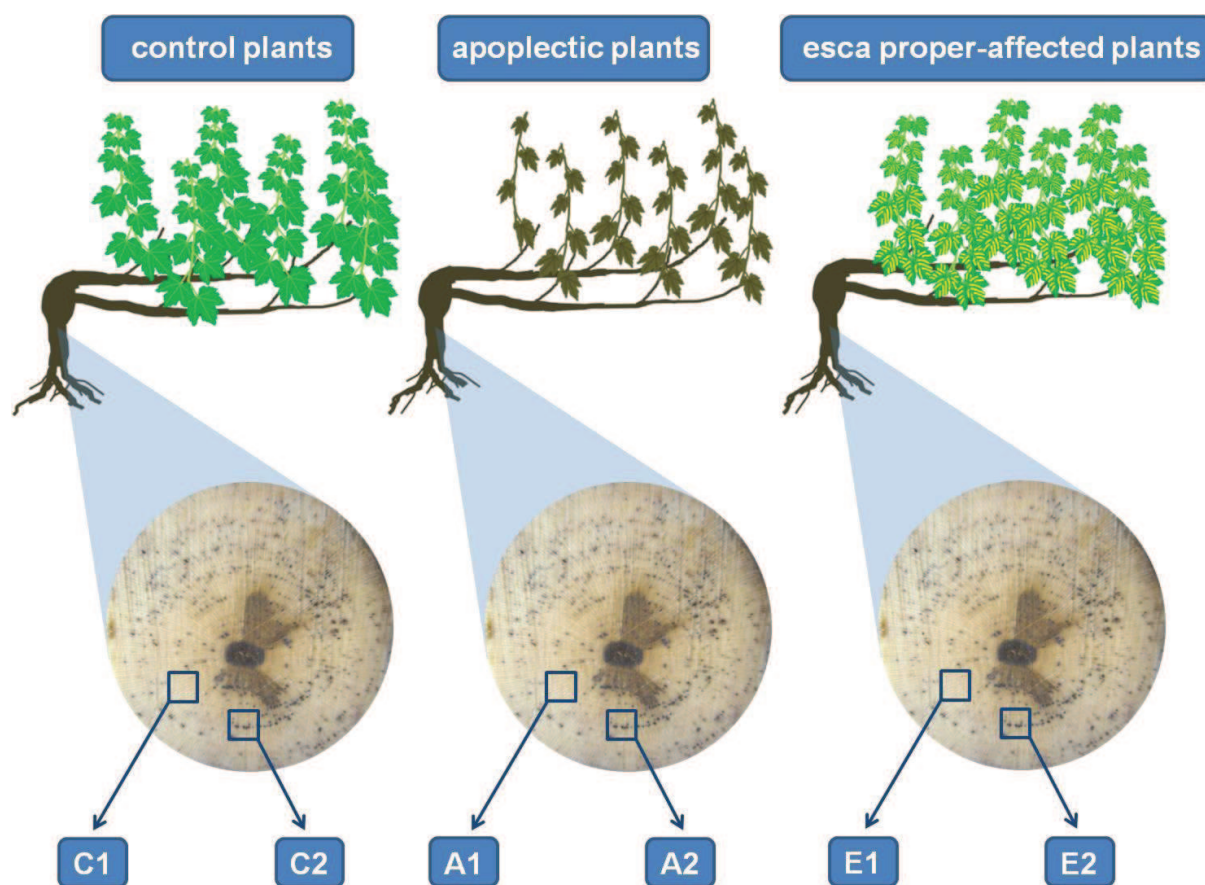
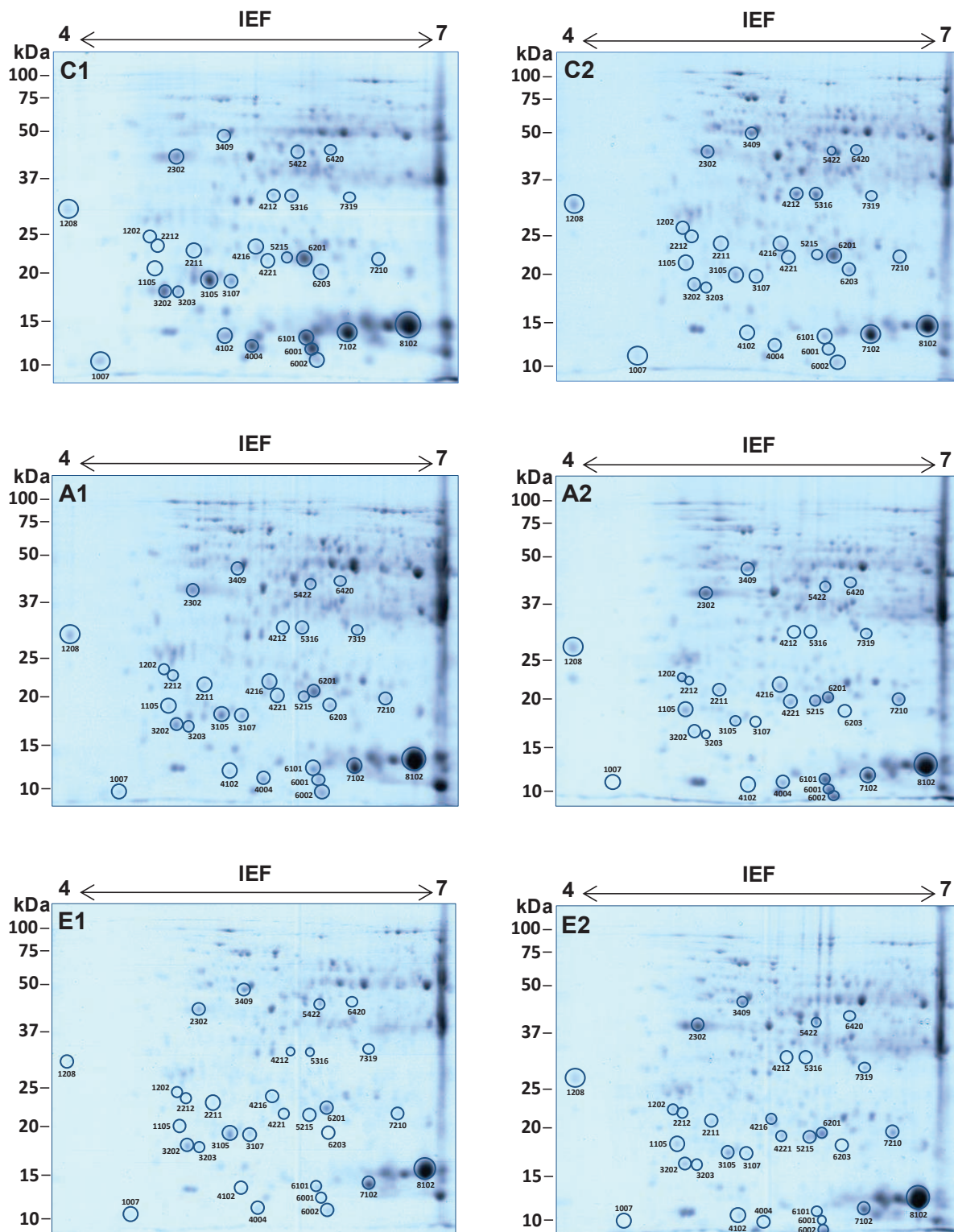


Figure 5.



Supplementary Fig. 1. Sampling strategy. Samples of asymptomatic (C1, A1, E1) and black streaked (C2, A2, E2) trunk wood were recovered from control (C1, C2) and apoplectic (A1, A2) or esca proper affected (E1, E2) plants. Control plants had not shown either apoplexy or GLSD symptoms since 2001.



Supplementary Fig. 2. Map of the 30 identified protein spots quantitatively differentially expressed among the groups of wood samples: asymptomatic (C1, A1, E1) and black streaked (C2, A2, E2) trunk wood of control (C1, C2), apoplectic (A1, A2) or esca proper affected (E1, E2) plants. Isoelectric focusing (IEF) was performed on precast dry polyacrylamide 7 cm length gels ReadyStrip IPG (pH 4–7). The relative molecular mass (kDa) was calibrated with standard protein markers (Prestained SDS-PAGE Standards, Bio-Rad, USA) after co-second dimensional electrophoresis. Only spots detected in at least one replica-gel of each sample were chosen for identification (indicated with a circle in the figure).

Supplementary Table 3. Primers of genes analyzed by real-time reverse-transcription polymerase chain reaction.

Gene	Primer sequences	Genbank or TC TIGR accession number	PCR primer efficiency (%)
<i>EF1-α</i> (elongation factor 1- α , housekeeping gene)	5'-GAACTGGGTGCTTGATAGGC-3' 5'-AACCAAAATATCCGGAGTAAAAGA-3'	GU585871	100.5
<i>Transaldolase</i>	5'-GAATCCTGCCTACCTGACACA-3' 5'-CATGGTCAACAAATGCTTGGA-3'	XM_00271478	90.1
<i>HSPCP</i> (heat shock protein, chloroplastic)	5'-TCGGTGGAGGATGACTTGCT-3' 5'-CGTGTGCTGTACGAGCTGAAG-3'	XM_002272382	103.3
<i>GSTF9</i> (glutathione S-transferase, phi class)	5'-GGAGCAGTGGCTAGAGGTTGA-3' 5'-CATCTGCAGGAAACCCCATCT-3'	EU181422	90.9
<i>SODCP</i> (superoxide dismutase, chloroplastic)	5'-GGGAAACATAGTTGCCAATGC-3' 5'-TCATGAACCACCAAGGCTCTT-3'	XM_002274385	96.3
<i>PRP-10</i> (protein PR 10)	5'-TGGTGAAGGCCACAAATTCA-3' 5'-GCGCCCTTGGTGTGGTACT-3'	XM_002274206	96.4
<i>TLP</i> (thaumatin-like protein)	5'-GCTACGGTTCACCCCAAA-3' 5'-CCATTGATGTCAGCGGAACA-3'	AJ237999	89.9
<i>PRP-17</i> (protein PR 17)	5'-TGCCGATTATGTGAGGTTGAAG-3' 5'-CCAGAAGCTCCACGAAGAAG-3'	XM_002269434	99.1
<i>CHI</i> (chalcone isomerase)	5'-GCAGAAGCCAAAGCCATTGA-3' 5'-GCCGATGATGGACTCCAGTAC-3'	XM_002282072	93.2
<i>IFRL4</i> (isoflavone reductase)	5'-GGATCGTGTAAATGCGGTTGA-3' 5'-GCCTGGCTGGACCAATGTAG-3'	BN000709	96.0
<i>SSP</i> (stem specific protein TSJT1)	5'-GGGAAGGTGCCTCTGTCTTG-3' 5'-TTGAAGGTTGCACCCCATATC-3'	XM_002280622	98.9
<i>14-3-3</i> (proteins 14-3-3)	5'-TGATACCTTCGGCTTCCACAA-3' 5'-GCGCCAAGCCCAACCT-3'	FJ824802	96.4

Supplementary Table 4. Percentage of isolation of esca proper causal agents (Pch, Pal and Fom) and other fungi associated to grapevine trunk diseases (as Botryosphaeriaceae species and *Eutypa lata*) from the discoloured or non discoloured trunk or branch wood of apoplectic, esca proper-affected and control plants. Percentage values refer to the number of samples (portion of trunk or branch wood) which these fungal pathogens were isolated from.

	Apoplectic plants		Esca proper-affected plants		Control plants	
	discolored	non discolored	discolored	non discolored	discolored	non discolored
trunk wood	a (63.2)	a (0.0)	a (53.8)	a (0.0)	a (66.7)	a (0.0)
	b (17.5)	b (0.0)	b (15.4)	b (0.0)	b (25.0)	b (0.0)
	c (5.3)	c (0.0)	c (23.3)	c (0.0)	c (8.3)	c (0.0)
	d (10.5)	d (0.0)	d (15.4)	d (0.0)	d (8.3)	d (0.0)
	e (0.0)	e (0.0)	e (0.0)	e (0.0)	e (25.0)	e (0.0)
	f (6.5)	f (0.0)	f (5.0)	f (0.0)	f (8.3)	f (0.0)
branch wood	a (55.5)	a (0.0)	a (60.0)	a (0.0)	a (80.0)	a (0.0)
	b (0.0)	b (0.0)	b (0.0)	b (0.0)	b (0.0)	b (0.0)
	c (0.0)	c (0.0)	c (30.0)	c (0.0)	c (0.0)	c (0.0)
	d (0.0)	d (0.0)	d (0.0)	d (0.0)	d (0.0)	d (0.0)
	e (11.1)	e (0.0)	e (20.0)	e (0.0)	e (0.0)	e (0.0)
	f (22.2)	f (0.0)	f (20.0)	f (0.0)	f (20.0)	f (0.0)

^a *Phaeoconiella chlamydospora* (Pch)

^b *Phaeoacremonium aleophilum* (Pal) and/or other *Phaeoacremonium* spp.

^c Botryosphaeriaceae spp.

^d *Fomitiporia mediterranea* (Fom) and/or other *Fomitiporia* spp.

^e *Eutypa lata*

^f Other fungi not associated to grapevine trunk diseases

Conclusion et perspectives

Dans une **première partie**, notre étude a porté sur les **tiges herbacées** de vignes de Chardonnay, puis dans une **seconde partie**, nous avons étudié les **trons de vignes** de Chardonnay, où siègent les champignons responsables de l'esca. Les banques de données protéiques des agents pathogènes impliqués dans cette maladie de bois de vigne étant peu fournies, nous avons utilisé une approche basée sur le séquençage *de novo* (PepNovo) et l'identification par homologies de séquences (MS-BLAST). Cette approche de recherche par homologie de séquence bien plus difficile à mettre en place par rapport à la recherche conventionnelle a conduit à une quantité très importante de données qui sont au-delà des capacités des outils de gestion de données en protéomique comme scaffold. Ce qui a conduit à l'adaptation des outils EXCEL pour le traitement de ces données de protéomique.

La combinaison de stratégies d'analyses protéomiques basées sur les gels 2D et l'identification des protéines par combinaison de la recherche conventionnelle et séquençage *de novo* PepNovo/MS-BLAST, a permis d'identifier les protéines différentiellement exprimées dans les trons de vigne infectés par l'esca. Notre stratégie d'identification des protéines a pu être appliquée avec succès à plusieurs variétés de vignes infectées par l'esca, et est en cours d'application pour d'autres études d'interactions plantes-pathogènes. La recherche par séquençage *de novo* PepNovo/MS-BLAST a permis de confirmer jusqu'à 83% des protéines identifiées par la recherche conventionnelle. La recherche conventionnelle basée sur les banques protéiques, a permis d'identifier que 6% des protéines de champignons sur l'ensemble des résultats puisque les génomes de la plupart des champignons impliqués dans la maladie de l'esca ne sont pas séquencés.

Même si les mêmes agents pathogènes sont responsables de l'esca sur les différentes variétés de vignes étudiées, il apparaît que ces variétés de vignes, ont des réponses différentes face à l'attaque de l'esca, puisque seulement 11 sur près de 70 protéines différentiellement exprimées sont communes entre les différentes variétés de vigne atteintes d'esca. En effet, les plants de vignes ont exprimé un grand nombre de protéines isoformes différents d'une variété à l'autre comme les HSP, tubulin, ribosomal protein, ATP synthase subunit, glutathione S-transferase etc. Quelques protéines de vigne identifiées par analyse protéomique, ont été suivies par PCR. Les résultats de quantification de PCR ont été dans certains cas en désaccord avec la quantification sur gel 2D. Ces résultats montrent une fois de plus que l'expression d'un gène n'est pas strictement corrélée à la quantité de protéine réellement exprimée dans un tissu. Néanmoins, la comparaison de la quantification par PCR ou par gel 2D entre les différences symptômes de l'esca a permis de montrer de remarquables différences quantitatives entre les symptômes de l'esca. Grâce aux fonctions des protéines identifiées par BLAST et l'annotation fonctionnelle par Gene Ontology, les protéines de vignes identifiées ont pu être classées en trois groupes, (i) métabolisme et énergie, (ii) tolérance aux stress, et (iii) la réponse de défense. Cependant, la plupart des protéines fongiques identifiées n'ont pas de fonctions biologiques définies et donc n'ont pas pu être classées.

La recherche d'homologies de séquence par *de novo* PepNovo/MS-BLAST, génère une quantité très importante de données (notamment les peptides des variants protéiques, les peptides spécifiques à des protéines non séquencés). Cependant la gestion des données correspondant à des centaines de protéines et des milliers de peptides comme nous l'avons vu est très fastidieuse à cause du manque d'outils de protéomique capable dégrader et traiter plus facilement les données de recherche par homologies de séquences. C'est pourquoi cette approche est très rarement utilisée dans le cadre des études protéomiques malgré les nombreux avantages qu'elle offre. Il est donc indispensable de développer des outils de gestion de donnée de recherche par homologies de séquences.

Plusieurs peptides séquencés restent incomplets à cause de certaines parties de leurs spectres MS/MS qui sont peu informatifs ou correspondent à des peptides modifiés par des PTMs. Il n'existe pas à ce jour d'outils de séquençage *de novo* ou de recherche conventionnelle à même de prendre en compte toutes les PTMs. Ces recherches prendraient non seulement des temps de calcul très longs, mais aussi entraîneraient l'indentification de nombreux faux positifs puisque de nombreuses PTMs sont isobariques. De plus, la fragmentation en spectrométrie de masse des peptides portant quelques PTMs est complexe et encore trop mal connue pour être modélisée et intégrée dans les algorithmes de séquençage *de novo*. Le séquençage *de novo* se limite en général uniquement à la recherche d'écarts de masse correspondant aux acides aminés sur un spectre MS/MS. Beaucoup de progrès restent donc à faire pour pouvoir interpréter tous les peptides modifiés. Cependant, les outils de séquençage *de novo* existant sont de véritables outils indépendants de vérification et de validation des résultats de recherche conventionnelle, et de recherche des peptides candidats aux protéines des organismes dont les génomes ne sont pas encore séquencés.

Cette étude a contribué à mettre en évidence les voies métaboliques activées lors de l'infection de l'esca chez les différentes variétés de vigne étudiées. De plus cette étude a démontré que les mêmes voies métaboliques sont activées au cours des deux symptômes de l'esca. Cette a permis également d'identifier les protéines homologues de champignons impliquées dans cette maladie de l'esca et donc les possibles voies métaboliques de champignons activés au cours de la maladie de l'esca.

Références bibliographiques du chapitre 2

1. Surico, G. Introduction - The Grapevine and Wine Production through the Ages. *Phytopathol. Mediterr.* **2006**, 39, (1), 03-10.
2. Surico, G. Towards a redefinition of the diseases within the esca complex of grapevine. *Phytopathol. Mediterr.* **2009**, 48, (1), 05-10.
3. Surico, G. Towards Commonly Agreed Answers to Some Basic Questions on Esca. *Phytopathol. Mediterr.* **2001**, 40, (3), S487–S490.
4. Calzarano, F.; Seghetti, L.; Del Carlo, M.; Cichelli, A. Effect of Esca on the Quality of Berries, Musts and Wines. *Phytopathol. Mediterr.* **2004**, 43, (1), 125-135.
5. Crous, P. W.; Gams, W. «*Phaeomoniella chlamydospora*» Gen. et Comb. Nov., a Causal Organism of Petri Grapevine Decline and Esca. *Phytopathol. Mediterr.* **2000**, 39, (1), 112-118.
6. Mostert, L.; Halleen, F.; Fourie, P.; Crous, P. W. A review of *Phaeoacremonium* Species Involved in Petri Disease and Esca of Grapevines. *Phytopathol. Mediterr.* **2006**, 45, S12-S29.
7. Di Marco, S.; Cesari, A.; Calzarano, F.; Mazzullo, A. The Control of Esca : Status and Perspectives. *Phytopathol. Mediterr.* **2000**, 39, (1), 232-240.
8. Calzarano, F.; Amalfitano, C.; Seghetti, L.; Agostino, V. Foliar treatment of esca-proper affected vines with nutrients and bioactivators. *Phytopathol. Mediterr.* **2007**, 46, (2), 207–217.
9. Waite, H.; May, P. The Effects of Hot Water Treatment, Hydration and Order of Nursery Operations on Cuttings of *Vitis vinifera* Cultivars. *Phytopathol. Mediterr.* **2005**, 44, (2), 144–152.
10. Petit, A. N.; Vaillant, N.; Boulay, M.; Clement, C.; Fontaine, F. Alteration of photosynthesis in grapevines affected by esca. *Phytopathol. Mediterr.* **2006**, 96, (10), 1060-6.
11. Lima, M. R.; Felgueiras, M. L.; Graca, G.; Rodrigues, J. E.; Barros, A.; Gil, A. M.; Dias, A. C. NMR metabolomics of esca disease-affected *Vitis vinifera* cv. Alvarinho leaves. *J Exp Bot* **2010**, 61, (14), 4033-42.
12. Tabacchi, R.; Fkyera, C.; Dubin, G.-M., Phytotoxins from Fungi of Esca of Grapevine. *Phytopathol. Mediterr.* **2006**, 39, (1), 156-161.
13. Gonzalez-Fernandez, R.; Jorin-Novo, J. V. Contribution of proteomics to the study of plant pathogenic fungi. *Journal of proteome research* **2012**, 11, (1), 3-16.
14. Wielsch, N.; Thomas, H.; Surendranath, V.; Waridel, P.; Frank, A.; Pevzner, P.; Shevchenko, A. Rapid validation of protein identifications with the borderline statistical confidence via de novo sequencing and MS BLAST searches. *Journal of proteome research* **2006**, 5, (9), 2448-56.
15. Spagnolo, A.; Magnin-Robert, M.; Alayi, T. D.; Cilindre, C.; Mercier, L.; Schaeffer-Reiss, C.; Van Dorsselaer, A.; Clement, C.; Fontaine, F. Physiological changes in green stems of *Vitis vinifera* L. cv. Chardonnay in response to esca proper and apoplexy revealed by proteomic and transcriptomic analyses. *J Proteome Res* **2012**, 11, (1), 461-75.
16. Hall, B. K. Homology, homoplasy, novelty, and behavior. *Developmental psychobiology* **2012**, 1-9.
17. Koonin, E. V. Orthologs, paralogs, and evolutionary genomics. *Annual review of genetics* **2005**, 39, 309-38.
18. Vizcaino, J. A.; Cote, R.; Reisinger, F.; Barsnes, H.; Foster, J. M.; Rameseder, J.; Hermjakob, H.; Martens, L. The Proteomics Identifications database: 2010 update. *Nucleic acids research* **2010**, 38, (Database issue), D736-42.
19. Magnin-Robert, M.; Letousey, P.; Spagnolo, A.; Rabenoelina, F.; Jacquens, L.; Mercier, L.; Clément, C.; Fontaine, F. Leaf stripe form of esca induces alteration of photosynthesis and defence reactions in presymptomatic leaves. *Functional Plant Biology* **2011**, 38, (11), 856-866.
20. Elias, J. E.; Haas, W.; Faherty, B. K.; Gygi, S. P. Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. *Nature methods* **2005**, 2, (9), 667-75.

21. Balgley, B. M.; Laudeman, T.; Yang, L.; Song, T.; Lee, C. S. Comparative evaluation of tandem MS search algorithms using a target-decoy search strategy. *Molecular & cellular proteomics : MCP* **2007**, *6*, (9), 1599-608.
22. Frank, A.; Pevzner, P. PepNovo: de novo peptide sequencing via probabilistic network modeling. *Anal Chem* **2005**, *77*, (4), 964-73.
23. Pevtsov, S.; Fedulova, I.; Mirzaei, H.; Buck, C.; Zhang, X. Performance evaluation of existing de novo sequencing algorithms. *Journal of proteome research* **2006**, *5*, (11), 3018-28.
24. Bringans, S.; Kendrick, T. S.; Lui, J.; Lipscombe, R. A comparative study of the accuracy of several de novo sequencing software packages for datasets derived by matrix-assisted laser desorption/ionisation and electrospray. *Rapid communications in mass spectrometry* **2008**, *22*, (21), 3450-4.
25. Wielsch, N.; Thomas, H.; Surendranath, V.; Waridel, P.; Frank, A.; Pevzner, P.; Shevchenko, A. Rapid validation of protein identifications with the borderline statistical confidence via de novo sequencing and MS BLAST searches. *J Proteome Res* **2006**, *5*, (9), 2448-56.

Chapitre 3 :

Etude et identification de nouveaux variants d'hémoglobine humaine

Cette étude a été réalisée en collaboration avec le Pr. E. Bissé de la Faculté de Médecine "Universitätsklinikum Freiburg" Allemagne.

Lors de cette étude, deux variants d'hémoglobine ont été détectés chez deux patients au cours d'un examen hématologique. Et pour mettre en évidence les acides aminés mutés sur chacun de ces variants d'hémoglobine, nous avons utilisé une stratégie analytique basée sur les outils de l'analyse protéomique (LC et MS). Ainsi deux nouveaux variants de la chaîne de globine alpha et delta d'hémoglobine humaine ont pu être caractérisés.

I. Contexte de cette étude

1. L'hémoglobine et les hémoglobinopathies

L'hémoglobine est une métalloprotéine du fer donc la principale fonction est le transport du dioxygène chez les mammifères. L'hémoglobine est le constituant majeur des globules rouges (environ 97% du contenu des globules rouges secs)¹. L'hémoglobine est constituée de quatre sous-unités (chaînes de globine): quatre chaînes polypeptidiques identiques deux à deux (Figure 1).

Chez l'homme, il existe différents types d'hémoglobines qui se distinguent par l'assemblage de leurs sous-unités et leur ordre d'apparition au cours des différents stades du développement.

L'hémoglobine humaine chez l'adulte est composée de : 97% d'hémoglobine HbA ($\alpha_2\beta_2$), de 2-3% de HbA2 ($\alpha_2\delta_2$) et de moins de 1% d'hémoglobine fœtale HbF ($\alpha_2\gamma_2$) (Figure 1).

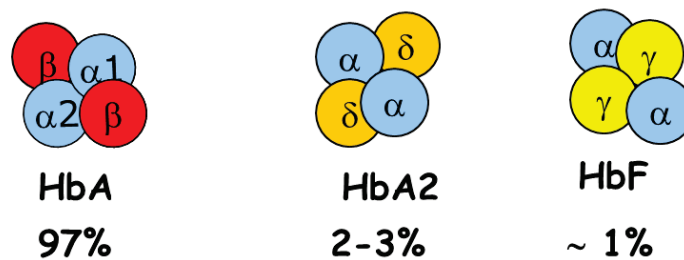


Figure 1. Composition de l'hémoglobine du sang d'un homme adulte.

A chacune de ces globines est associé un groupement prosthétique, l'hème, qui contient en son cœur un ion Fe^{2+} capable de fixer l' O_2 et le CO_2 dissous dans le sang (Figure 2).

Pendant le stade embryonnaire, deux chaînes de globine ζ (zeta) apparentées à la chaîne α , peuvent se lier à d'autres chaînes, comme par exemple les globines ϵ (epsilon) (Hb Gower 1 : ($\zeta_2\epsilon_2$)) ou γ (Hb Portland : ($\zeta_2\gamma_2$)). Ces formes d'hémoglobine disparaissent progressivement au cours de la grossesse pour laisser place à l'hémoglobine HbA, HbA2 et HbF (Figure 3).

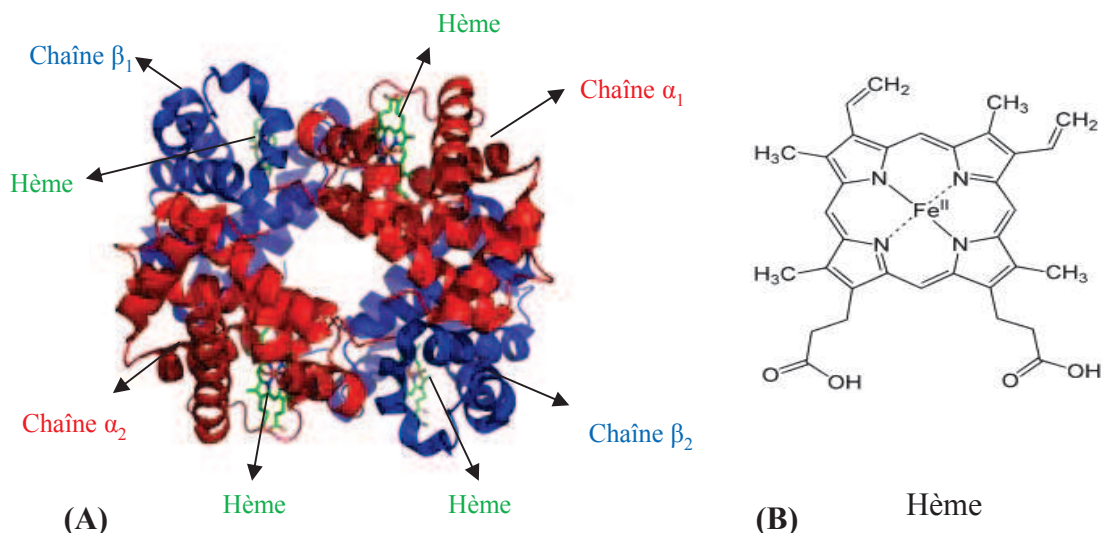


Figure 2. (A) structure 3D de l'hémoglobine HbA0 humaine avec ces 2 chaînes de globines α et deux chaînes de globine β dans lesquelles logent 4 hèmes et (B) l'hème de groupe b avec son ion Fe^{2+} central.

Les sites de synthèse des hémoglobines diffèrent également au cours du développement. Pendant la phase fœtale, les chaînes de globines sont principalement synthétisées dans le foie, tandis qu'après la naissance, elles sont synthétisées dans les précurseurs des globules rouges, dans la moelle osseuse.

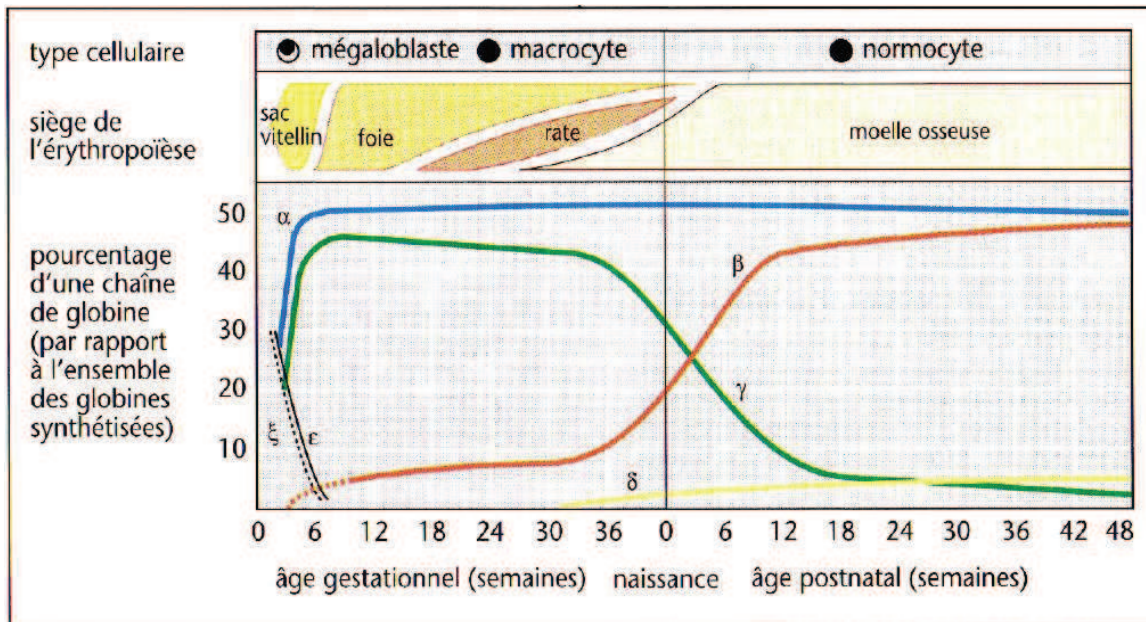


Figure 3. Variation de la concentration des chaînes de globine humaine au cours du développement ².

Chez l'homme, les gènes des chaînes apparentées à la chaîne β (famille β : β, γ, δ, ε) sont situés à proximité les uns des autres formant un cluster de gènes sur le chromosome 11, tandis que les gènes des chaînes apparentées à la α (famille α : α, ζ) se situent sur le chromosome 16 (Figure 4).

Les gènes de globines alpha sont codés par 2 gènes distincts alpha 1 (HBA1) et alpha 2 (HBA2) situés côte à côte sur le chromosome 16. Il y a également deux gènes qui codent pour les deux chaînes de globine gamma 1 et gamma 2 situés sur le chromosome 11 (Figure 4).

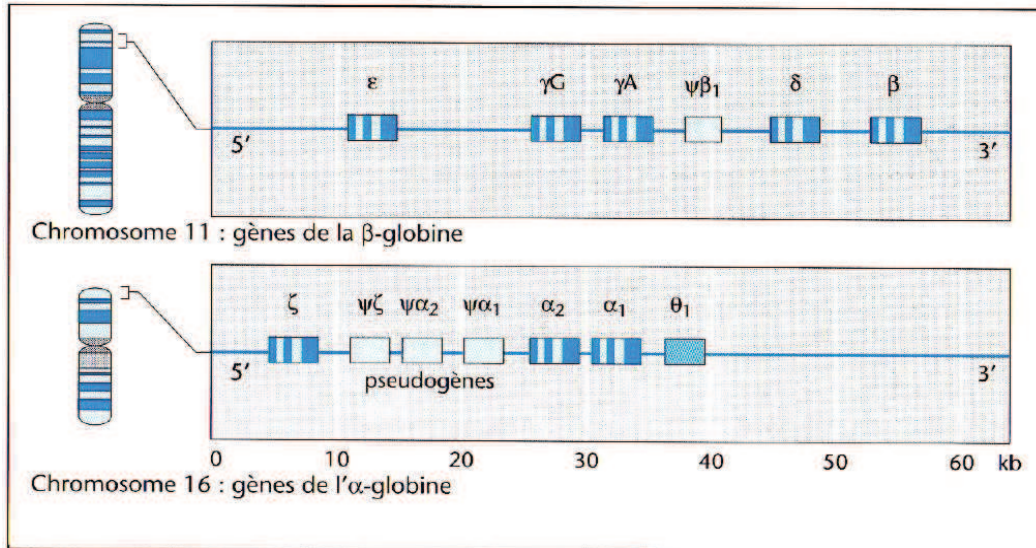


Figure 4. Les gènes de l'hémoglobine humaine et leurs positions sur les chromosomes 11 et 16 ³.

Les gènes des chaînes de globines peuvent subir des mutations qui conduisent à de nouveaux variants d'hémoglobine. Ces mutations peuvent être dues soit à: (1) une délétion de codons (ex. Hb Bruxelles ⁴, Hb Gun Hill ⁵); (2) une substitution d'un codon par un autre (ex. Hb S ⁶); (3) une insertion de codon (ex. Hb Cranston ⁷).

Les mutations les plus fréquentes, sont les mutations ponctuelles, n'affectant qu'un seul codon qui conduisent à des centaines de variants d'hémoglobines (Tableau 1) qui sont impliqués dans de nombreuses pathologies nommées hémoglobinopathies.

Les hémoglobinopathies peuvent être réparties en deux catégories suivant qu'elles soient quantitatives (thalassémies) ou qualitatives (liés aux variants d'hémoglobine). Les thalassémies sont des défauts ou

insuffisances de synthèse des chaînes de globine qui peuvent toucher les chaînes de globine alpha (alpha thalassémie) ou la beta (beta thalassémie).

Tableau 1. Bilan des variants de globines répertoriés dans la base de référence des mutations d'hémoglobine humaine, Hbvar (5 novembre 2012) : (<http://globin.bx.psu.edu/hbvar>).

	Nombres d'entrée
Nombre d'entrées dans la banque	1542
Nombre variants d'hémoglobine	1156
Nombre de variants des chaînes α	699
Nombre de variants des chaînes β	811
Nombre de variants des chaînes δ	98
Nombre de variants des chaînes γ	124
Nombre de variants dus à une insertion de gènes	63
Nombre de variants dus à une fusion de gènes	9
Nombre de variants dus à une délétion de gènes	187
Nombre de variants dus à une substitution de gènes	1259

Les conséquences fonctionnelles et cliniques de ces mutations sont différentes. La mutation d'un acide aminé par un autre peut entraîner des changements des propriétés physicochimiques des chaînes de globines, qui peuvent, soit conduire à des anomalies hématologiques avec des conséquences plus ou moins graves (par exemple, la perte d'affinité avec l'oxygène), soit être sans incidence clinique dans le cas des mutations dites silencieuses. L'hémoglobinopathie la plus répandue avec une pathologie sévère est la drépanocytose ⁸, qui résulte d'une substitution sur la chaîne beta de l'acide glutamique en position 6 par une valine.

On estime aujourd'hui à environ 5%, la population mondiale porteuse des gènes caractéristiques des hémoglobinopathies (www.who.int) et plus de 300000 enfants porteurs de formes graves d'hémoglobinopathies. Mais ce pourcentage peut atteindre parfois jusqu'à 45% dans certaines régions du globe comme par exemple à l'est de l'Afrique ⁹.

Face à l'incidence majeure des hémoglobinopathies sur la santé publique, de vastes programmes nationaux ont été mis en place dans les centres de services hématologiques pour le diagnostic précoce des hémoglobinopathies afin de permettre une meilleure prise en charge des patients ¹⁰.

2. Problématique et objectifs de l'étude

Les troubles de synthèse des chaînes de globine qui conduisent à des variants d'hémoglobine, sont très fréquents et constituent un problème majeur de santé publique. Les diagnostics de ces variants sont donc souvent nécessaires pour: (1) confirmer les diagnostics prévisionnels des hémoglobinopathies tels que la drépanocytose ; (2) expliquer les anomalies hématologiques tels que les anémies et microcytoses ; (3) identifier les anomalies dans les phases pré-symptomatiques lors des diagnostics prénatales ; (4) identifier les fœtus ou les individus présentant des risques d'hémoglobinopathies et offrir aux patients des choix éclairés ; (5) fournir des conseils génétiques aux futurs mariés et parents.

Les méthodes de détection des variants de globines dans les laboratoires d'hématologie sont basées sur la chromatographie échangeuse d'ions, l'HPLC en phase inverse (RP-HPLC) et IEF (focalisation isoélectrique). On note également l'utilisation de plus en plus croissante d'autres méthodes pour l'identification des anomalies des gènes de globines dont les méthodes d'analyse d'ADN, la spectrométrie de masse et les techniques basées sur l'immunologie ¹¹⁻¹³.

Mais très souvent les résultats d'analyse de ces techniques peuvent présenter des biais lorsqu'il s'agit de nouveaux variants d'hémoglobine. En effet, la méthode de dosage d'un hémolysat, par chromatographie échangeuse d'ions, repose sur l'analyse des hémoglobines natives sous forme de tétramères. Pour un hémolysat provenant d'un patient qui présente une mutation mineure, le temps de rétention de l'hémoglobine mutée ne sera pas en général suffisamment affecté pour que la mutation soit mise en évidence par ce dosage. Quant à l'analyse d'ADN, dans le cas de nouveaux variants, les mutations peuvent provoquer le changement de

l'amorce au niveau de l'ADN. Les modifications post-traductionnelles peuvent provoquer la conversion d'un acide aminé en un autre (ex. Hb Bristol¹⁴). A ces difficultés, s'ajoute la question de savoir lequel des gènes doit être séquencé. Puisque par exemple dans le cas d'un variant de la chaîne alpha, deux gènes différents (HBA1 et HBA2) codent pour les chaînes de globines identique alpha 1 et alpha 2.

Pour ces multiples raisons, l'identification de nouveaux variants d'hémoglobine passe par plusieurs techniques d'analyse des chaînes de globine par HPLC et spectrométrie de masse en identifiant précisément la mutation. C'est dans ce contexte que se situe ce chapitre de thèse.

L'objectif de ce travail de thèse était d'étudier deux échantillons provenant de deux patients, pour lesquels, les analyses hématologiques ne permettaient pas d'émettre un diagnostic clinique, à l'aide d'outils d'analyse protéomique, dans le but d'identifier et de caractériser les mutations des hémoglobines de ces deux patients.

Pour des raisons de confidentialité, seule l'identification d'un des deux variants d'hémoglobine étudiés est décrite en détail dans ce chapitre de thèse.

II. Identification de nouveaux variant

1. Description des échantillons

Les hémolysats de deux patients (nous appellerons dans la suite de ce travail ces échantillons Nal et Bre) lors des examens hématologiques ont présenté des profils de chromatographie échangeuses d'ions différents de l'hémolysat normal. En effet, des pics supplémentaires ne correspondant à aucune forme d'hémoglobine connue, ont été observés.

Pour identifier la nature de ces hémoglobines élués dans les pics supplémentaires suspectés, plusieurs collectes de fractions de ces pics ont été réalisées afin d'enrichir les hémoglobines anormales de ces patients.

Les fractions d'hémoglobines anormales collectées ont été analysées par LC-MS et nanoLC-MS/MS pour l'identification et la caractérisation de ces globines.

2. Stratégie d'identification de nouveaux variants

La stratégie d'analyse LC-MS et nanoLC-MS/MS pour l'identification et la caractérisation des nouveaux variants de chaînes de globine est représentée sur la Figure 5.

Les protocoles utilisés pour l'identification et la caractérisation de chaînes de globines mutée sont détaillés dans la partie expérimentale générale.

Les chaînes de globines sont dans un premier temps séparées par chromatographie sur phase inverse, et les masses de protéines séparées sont mesurées avec un spectromètre de masse de type Q-TOF qui permet d'obtenir une résolution et une précision de masse est nécessaire pour mesurer de différences de masse de protéines de 15 kDa. Les différences de masse entre masse mesurées et masse théoriques de chaînes de globines sont calculées pour poser une première hypothèse sur les possibles mutations qui peuvent induire cette différence de masse (Annexe 6, Tableau A 2).

Dans un second temps, les protéines ne correspondant à aucune forme de globine connues sont collectées, soumises à une étape de réduction/alkylation et digérer avec la trypsine ou d'autres enzyme et analysées par nanoL-MS/MS, afin d'identifier la chaîne de globine, et l'acide aminé muté par interprétation de spectre MS/MS du peptide muté.

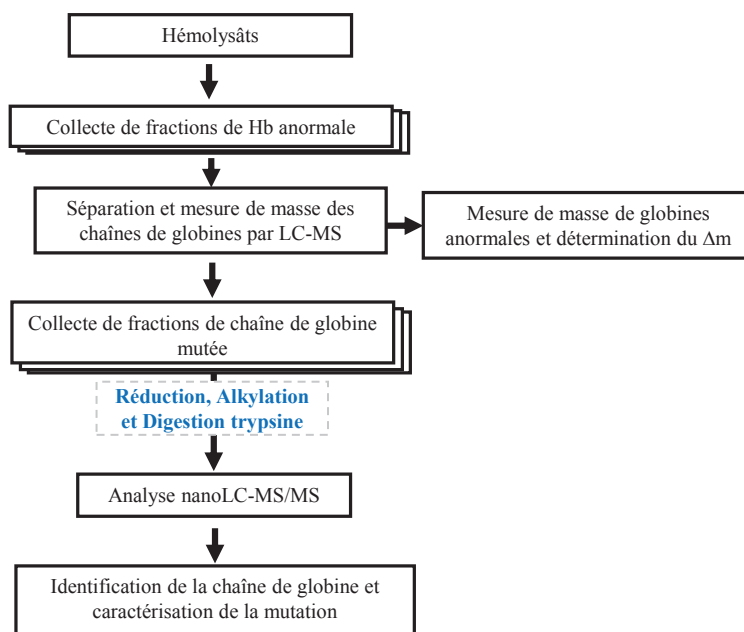


Figure 5. Stratégie globale d'analyse LC-MS et LC-MS/MS adoptée pour l'identification et la caractérisation de nouveaux variants de globine.

3. Principaux résultats

Pour des raisons de confidentialité, seul l'identification d'un des deux variants d'hémoglobine (échantillon Bre) étudiés sera décrit en détail dans cette partie des résultats.

a. LC-MS des chaînes de globines intactes È

L'hémolysât (échantillon Bre) a été analysé par LC-MS (Figure 6) pour séparer et mesurer la masse de chaque globine (Tableau 2). Le chromatogramme (A) correspond à de l'hémoglobine HbA0 ($\alpha_2\beta_2$) purifiée servant de témoin. Les pics (A1) et (A2) correspondent respectivement à des chaînes de globines normales beta et alpha.

La comparaison du chromatogramme (B) avec le chromatogramme de référence (A) et la mesure de masse des différents pics montre que le pic (B5) est le pic de la globine alpha normale. Les mesures de masse de protéines éluées ont permis d'attribuer le pic (B4) à la chaîne de globine delta normale (Tableau 2). Le pic (B3) correspond à une protéine dont la masse mesurée est de $15938,4 \pm 0,2$ Da, soit avec une différence de masse de +14 Da par rapport à la chaîne de globine delta normale. Le pic (B2) correspond à une protéine de $14391,3 \pm 0,3$ Da, soit une différence de masse supérieure à 130 Da par rapport à toutes les chaînes de globine, et ne peut donc pas correspondre à une mutation ponctuelle d'un acide aminé d'une chaîne de globine (Annexe 6, Tableau A 2). Cette protéine ne correspond probablement pas à une chaîne de globine. Le pic (B1) correspond à une contamination d'une molécule de $m/z = 700$ (1+) qui pourrait provenir de la purification.

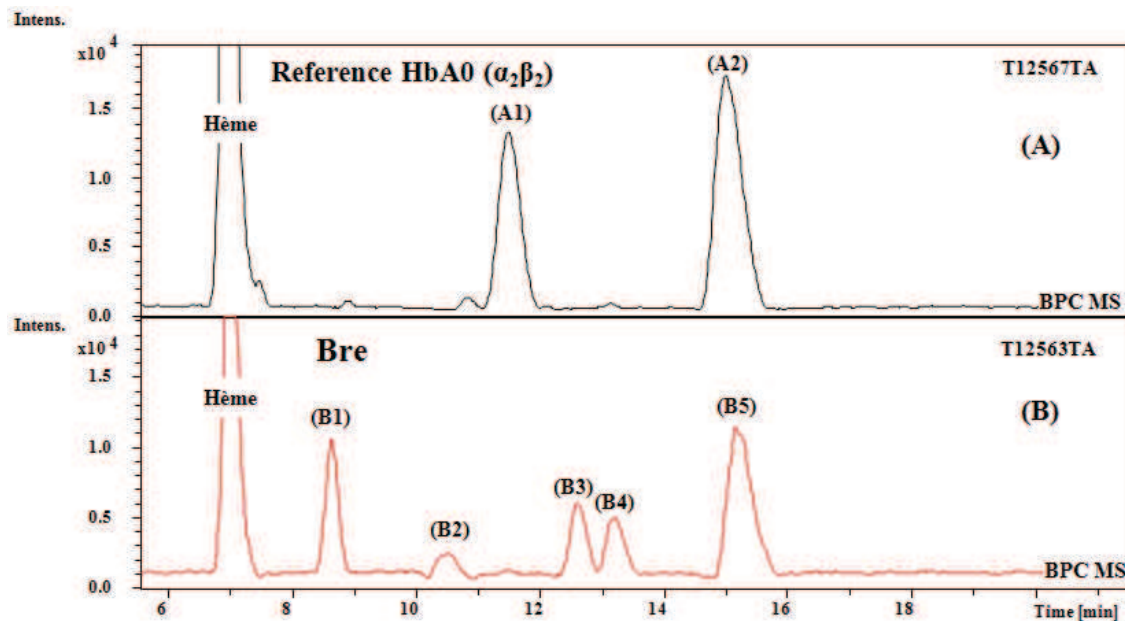


Figure 6. Chromatogramme (BPC, LC-MS) des chaînes de globines: (A) HbA ($\alpha_2\beta_2$) servant de référence et (B) échantillon Bre.

Tableau 2. Liste de masses moyennes expérimentales et théoriques (Mm_{exp} et Mm_{theo} normale), temps de rétention (Tr) correspondant à chaque pic chromatographique de échantillon Bre et la référence HbA0 ($\alpha_2\beta_2$). Les masses moyennes théoriques ont été calculées par Comput pI/Mw (www.expasy.org) avec la séquence de chaque protéine sans la méthionine initiale (Annexe 6, Tableau A 1).

Echantillon	Pic	Tr	Mm_{exp} (Da)	Mm_{theo} normale (Da)	Pic collecté	Δm (Da)
(B) Référence HbA ($\alpha_2\beta_2$)	Hème	7 min	$616,2 \pm 0,002$	616,17	-	-
	(A1) Chaîne β	11,5 min	$15867,6 \pm 0,3$	15867,22	-	-
	(A2) Chaîne α	15 min	$15126,7 \pm 0,3$	15126,36	-	-
(C) Bre	Hème	7 min	$616,2 \pm 0,001$	616,17	-	-
	(B1) contaminant	8,6 min	$700,2 \pm 0,005$	-	-	-
	(B2)	10,5 min	$14391,3 \pm 0,3$	-	(B2)	-
	(B3) Chaîne δ^*	12,6 min	$15938,4 \pm 0,2$	-	(B3)	$\Delta m (\delta^*-\delta) = +14$
	(B4) Chaîne δ	13,2 min	$15924,5 \pm 0,4$	15924,29	-	-
	(B5) Chaîne α	15,1 min	$15126,5 \pm 0,3$	15126,36	-	-

Les différences de masse entre la protéine inconnue (*) et les différentes chaînes de globine normales ont été ensuite calculées pour identifier les possibles mutations pouvant entraîner ces différences de masse (Tableau 3).

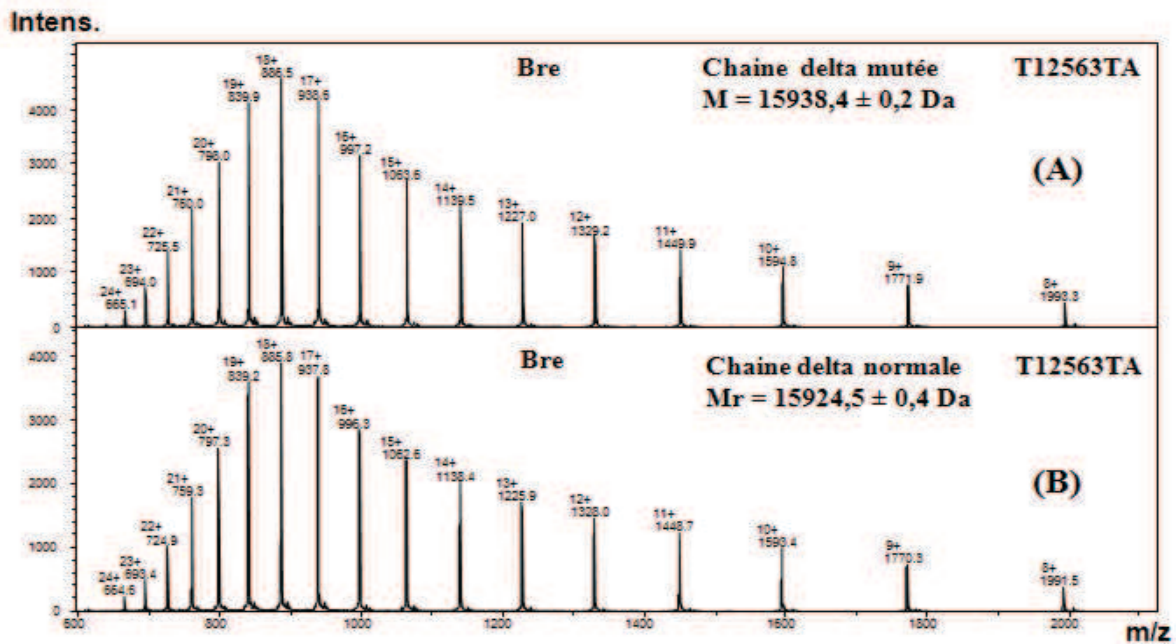


Figure 7. Spectre de masse de la chaîne de globine delta normale de $15924,5 \pm 0,4$ Da (B) et delta mutée de $15938,4 \pm 0,2$ (A) de l'échantillon Nal, respectivement dans les pics B3 et B4 de la figure 6.

Tableau 3. Différences de masse entre les chaînes de globines mutées et les mutations possibles des d'acides aminés correspondant. Les possibles mutations ont été extraites du tableau de différence des masses induites par le changement d'un acide aminé de la banque de référence de variants d'hémoglobines (HbVar database) (Annexe 6, Tableau A 2).

Echantillon	Globine mutées	M _{exp} (Da)	M _{theo normale} (Da)	ΔM (Da)	Possibles mutations
(C) Bre	(⊖) Globine δ [*]	15938,4	15924,29	+14,1 ± 0,2	Gly -> Ala Val -> Leu Val -> Ile Asp -> Glu Ser -> Thr Asn -> Lys

Les deux pics (B2) et (B3) non identifiés qui nous intéressent donc, ont été ensuite collectés pour l'analyse nanoLC-MS/MS afin d'identifier les différentes protéines éluées et de caractériser le ou les acides aminés mutés.

b. Identification et caractérisation des variants d'hémoglobine

Le digeste tryptique des fractions des pics collectés pour l'échantillon Bre ont été analysés par nanoLC-MS/MS (Tableau 4).

Tableau 4. Résultats d'identification des digestes des pics (B2), (B3) après recherche dans une banque SWISSPROT.

Echantillon	Pic	M _{exp} (Da)	M _{theo normale} (Da)	Protéines identifiées	Couverture de séquence
(C) Bre	(B2)	14391,3	28870,31	Carbonic anhydrase 1	55%
	(B3)	15938,4	15924,29	Globine δ [*]	83%

La chaîne de globine delta a été identifiée dans le pic (B3) (Figure 6) avec des couvertures de séquences de 83 % (Figure 9). Cette chaîne de globine correspond donc à un variant de la chaîne de globine delta d'hémoglobine humaine car la masse de la protéine entière expérimentale et théorique sont différentes. Ce résultat est donc en accord avec les analyses LC-MS des protéines entières.

Hemoglobin|subunit delta (P02042): couverture de séquence 82%

(δT5)

MVHLTPEEKT AVNALWGKVN VDAVGGEALG RLLVYPWTQ RFFESFGDLS SPDAVMGNPK
(δT6+δT7)

VKAHGK KVLG AFSGLAHLN NLKGTFSQLS ELHCDKLHVD PENFRLG NV LCVLARNFG

KEFTPQMCAA YQKVVAGVAN ALAHKYH

Figure 8. Représentation des séquences peptidiques tryptiques identifiées (en rouge) sur les séquences de la chaîne de globine delta identifiée après l'analyse nanoLC-MS/MS dans l'échantillon Bre. Les peptides non identifiés sont indiqués δT5 et δT6 (en noire).

L'analyse par LC-MS/MS n'a pas permis d'identifier certains peptides (ex : RFFESFGDLSPDAVMGNPK (δT5) et VKAHGK (δT6+ δT7) Figure 8). Ces peptides n'ont pas été sélectionnés pour la fragmentation ou ont été mal fragmentés. Pour vérifier leur présence dans les différents digestes, une extraction des courants d'ions de ces peptides non séquencés a été effectuée. Mais aucun de ces peptides n'a été identifié. Un de ces peptides pourrait donc correspondre au peptide dont un des acides aminés est muté.

Dans les banques de données protéiques, les séquences de globines sont données avec une méthionine N-terminale. Mais *in vivo*, les chaînes de globine humaine sont clivées de cette méthionine et exprimées dans les érythrocytes humains sous forme de globine mature sans la méthionine N-terminale¹⁵. C'est pourquoi cette méthionine n'a pas été identifiée sur les deux chaînes d'hémoglobine.

La carbonic anhydrase 1 (CA1, EC 4.2.1.1, P00915), qui est une métalloprotéine cytoplasmique a été identifiée dans les pics (A1) et (C2). Chez les animaux, l'un des rôles des protéines CA, est l'inter-conversion du dioxyde de carbone en bicarbonate pour maintenir l'équilibre acido-basique dans le sang et les tissus afin de

faciliter le transport du dioxyde de carbone ¹⁶. La présence de cette protéine dans l'échantillon Bre est due à une contamination introduite lors de l'enrichissement des hémoglobines mutées.

Pour localiser les pics d'éluion du peptide muté dans le chromatogramme nanoLC-MS/MS, les courants d'ions des peptides tryptiques qui n'ont pas été identifiés pour chaque chaîne de globine additionnés d'une différence de masse +14 Da ont été extraits. Les spectres MS/MS du peptide muté a été annoté pour identifier et localiser l'acide aminé muté.

Identification et caractérisation de la mutation de la globine delta mutée

Les extraits des courants d'ions des peptides de la chaîne δ , δT (peptide théoriques de la digestion tryptique *in silico* de la chaîne de globine delta (Annexe 6, Tableau A 3) +14 Da montrent que seul le peptide FFESFGDLSSPDAVMGNPK +14 Da ($\delta T5+14$ Da) a donné un pic sur le chromatogramme du digeste tryptique de la fraction de la chaîne de globine delta mutée de l'échantillon Bre (Figure 9).

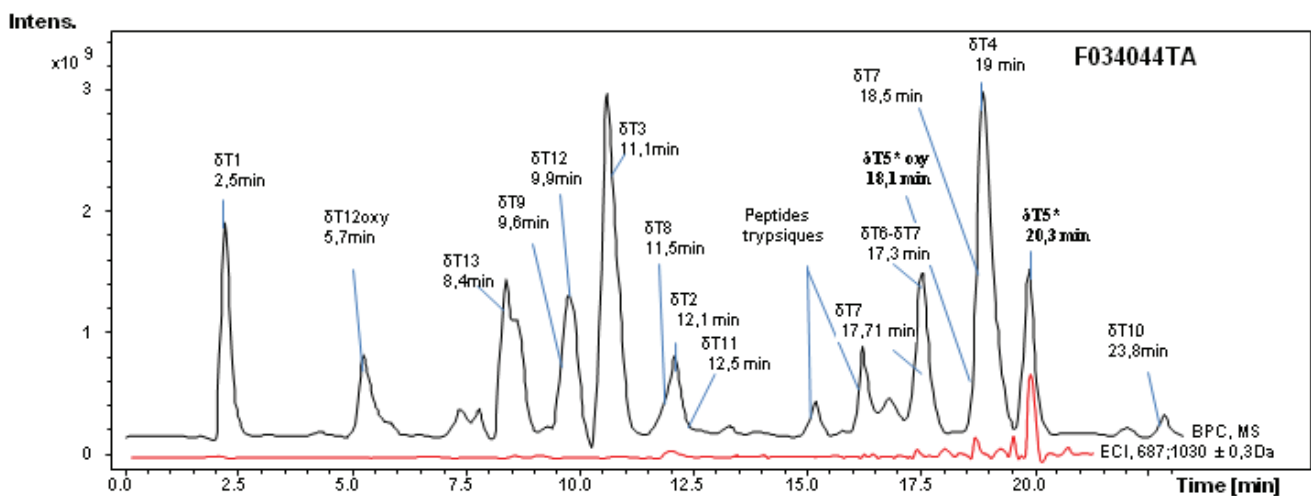


Figure 9. Chromatogramme (BPC, nanoLC-MS/MS) du digeste tryptique de la chaîne de globine delta mutée de l'échantillon Bre (δT peptides tryptiques de la chaîne de globine delta, oxy peptide avec (M) oxydé, * peptide muté). En rouge le courant d'ion du peptide muté $\delta T5^*$ = FFESFGDLSSPDAVMGNPK +14 Da di- et tri-chargé ($m/z = 1030$ ($z = 2+$) et 687 ($z = 3+$)).

Ce peptide détecté sous forme di et tri-chargé ($m/z = 1030$ ($2+$) et $m/z = 687$ ($3+$)) a été sélectionné et fragmenté en MS/MS. Les spectres de ce peptide muté ($\delta T5^*$, $\delta T5+14$ Da = FFESFGDLSSPDAVMGNPK +14 Da) ont donc été annotés manuellement pour identifier et localiser l'acide aminé muté (Figure 11 et 12).

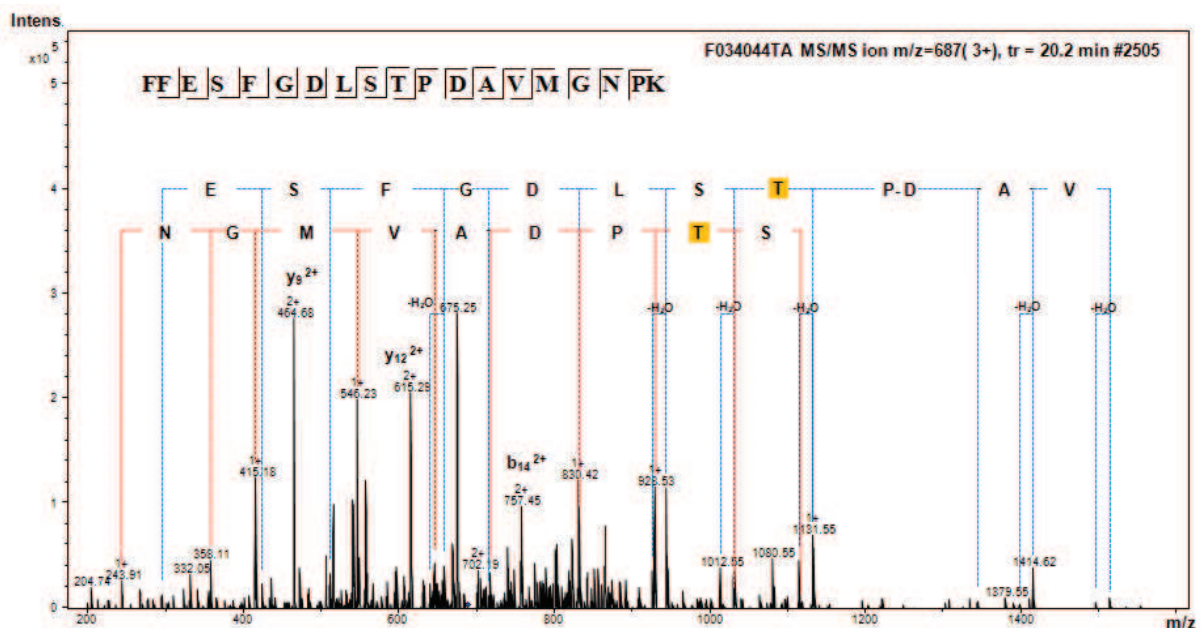


Figure 11. Spectre MS/MS annoté du peptide $\delta T5^*$ muté correspondant au $m/z = 687$ ($3+$) de la chaîne de globine delta mutée de l'échantillon Bre. La (T) remplace la (S)⁵⁰ dans le peptide muté. Les ions y sont indiqués en rouge et les ions b sont en bleu.

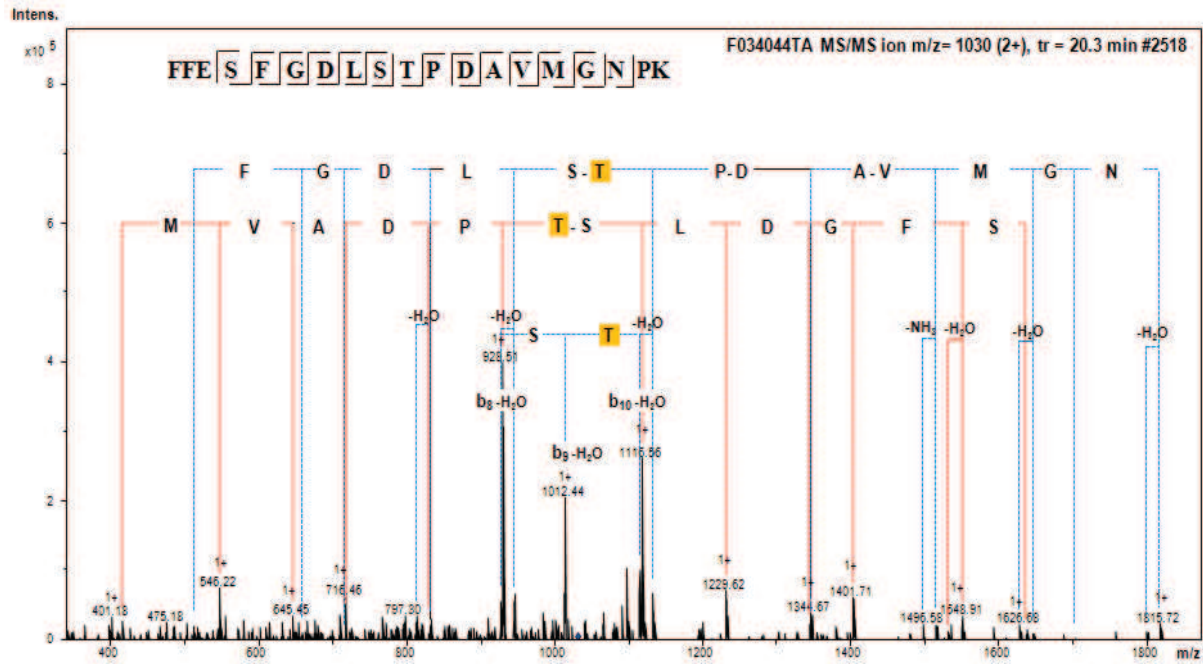


Figure 12. Spectre MS/MS annoté du peptide $\delta T5^*$ muté correspondant au $m/z = 1030$ ($2+$) de la chaîne de globine delta de l'échantillon Bre. La (T) remplace la (S)⁵⁰ dans le peptide muté. Les ions y sont indiqués en rouge et les ions b sont en bleu.

Par ailleurs, le peptide muté a également été identifié avec une oxydation sur la méthionine FFESFGDLST⁵⁰PDAVM^{oxy}GNPK ($\delta T5^*$ oxy sur la Figure 10).

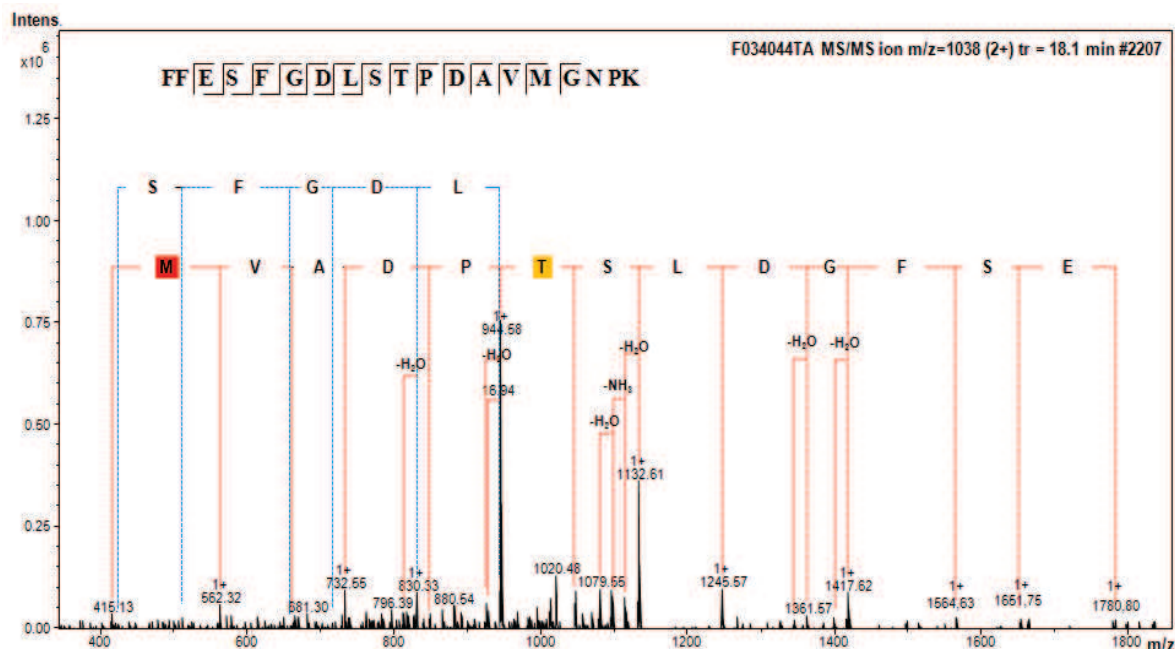


Figure 13. Spectre MS/MS annoté du peptide $\delta T5^*$ de la chaîne de globine delta de l'échantillon Bre avec la méthionine (M) oxydée ($\delta T5^*$ oxy) correspondant à $m/z = 1038$ ($2+$). Les ions y sont indiqués en rouge et les ions b sont en bleu.

Conclusion

Les spectres de fragmentation du peptide muté de la chaîne de globine delta ($\delta T5^*$ oxy et $\delta T5^*$) de l'échantillon Bre a permis de localiser la mutation en position 50, une mutation de la sérine en thréonine (Ser \rightarrow Thr). La différence de masse entre $\delta T5^*$ (2058 Da) et $\delta T5$ (2044 Da) ou entre $\delta T5^*$ oxy (2074 Da) et $\delta T5$ oxy (2060 Da) est de +14 Da.

Cette différence de masse est la même que celle observée lors de la mesure de masse de chaîne de globine δ^* muté et la chaîne de globine δ normale de l'échantillon Bre. Ces deux résultats de mesure de masse de la protéine entière et celle obtenue par séquençage du peptide muté de l'échantillon Bre sont concordant.

Ce variant de globine δ n'a jamais été décrit à notre connaissance. Le séquençage du gène de globine delta de l'échantillon Bre est en cours pour confirmer les résultats obtenus par spectrométrie de masse.

Conclusion et perspectives

Grâce à notre stratégie basée sur la purification des chaînes de globines par HPLC suivie de l'analyse par spectrométrie de masse, nous avons pu caractériser deux nouveaux variants d'hémoglobine humaine (chaînes alpha et delta), pour lesquels les analyses hématologiques ne permettaient pas d'émettre de diagnostic clinique.

L'analyse par spectrométrie de masse a permis d'orienter les analyses de l'ADN de la globine alpha de l'échantillon Na1. Le séquençage des gènes alpha a confirmé la mutation détectée par spectrométrie de masse.

La mutation de la chaîne de globine delta de l'échantillon Bre décrite dans ce chapitre de thèse est une mutation de la sérine en position 50 en thréonine. Les analyses des gènes delta de l'échantillon Bre sont en cours de réalisation pour confirmer la mutation identifiée par spectrométrie de masse.

Des tests d'affinités à l'oxygène sont en cours de réalisation afin de voir l'impact de chaque mutation sur l'affinité avec l'oxygène.

Ces 2 nouveaux cas de variants seront publiés et intégrés dans la banque de référence de variants d'hémoglobine (HbVar database).

Annexe 6: Tableaux

Tableau A 1: Masses théoriques moyennes (Mm theo) des différentes globines humaines. Les masses moyennes théoriques ont été calculées par Comput pI/Mw (www.expasy.org) avec la séquence de chaque protéine extraites de SWISSPROT sans la méthionine initiale.

Famille de globine	Chaîne Globines	de	Mm theo	N° swissprot	Chromosome
alpha	alpha		15126.36	P69905	11
	zeta		15505.83	P02008	
beta	beta		15867.22	P68871	16
	delta		15924.29	P02042	
	epsilon		16071.63	P02100	
	gamma 1		16009.28	P69891	
	gamma 2		15995.25	P69892	

Tableau A 2: Différence de masse induite par le changement d'un acide aminé extraite de la banque de référence de variants d'hémoglobine (HbVar database). Exemple une mutation entrainant une augmentation de +10 Da une Ser → Pro mais une différence de - 10 Da est une Pro → Ser. ($\Delta m = (M_{\text{variant}} - M_{\text{théorique}})$).

	0	1	2	3	4	5	6	7	8	9
0	Leu->Ile Gln->Ly	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			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	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
20										Gly->Trp

Tableau A 3 : liste des peptides de digestion trypsique générée *in silico* par (PeptideMass, expasy.org) de la globine delta humaine, les masses théoriques correspondant à ces peptides et les différents m/z selon les états de charge. (MC, missed cleavages)

Globine delta humaine							
	10	20	30	40	50	60	
	MVLSPADKTN	VKAAWGKVG	HAGEYGAEAL	ERMFLSFPTT	KTYFPHFDLS	HGSAQVKGHG	
	70	80	90	100	110	120	
	KKVADALTN	VAHVDDMPN	LSALSDLHAH	KLRVDPVNFK	LLSHCLLVLT	AAHLPAEFTP	
	130	140					
	AVHASLDKFL	ASVSTVLTSK	YR				
Nom du peptide	MC	Position	Peptide muté	Mtheo	MH ⁺	MH ⁺²	MH ⁺³
δT1	0	2-9	VHLTPEEK	951,5	952,5	476,8	318,2
δT2	0	10-18	TAVNALWGK	958,5	959,5	480,3	320,5
δT3	0	19-31	VNVDAVGGEALGR	1255,7	1256,7	628,8	419,6
δT4	0	32-41	LLVVYPWTQR	1273,7	1274,7	637,9	425,6
δT5	0	42-60	FFESFGDLSSPDAVMGNPK	2043,9	2044,9	1023,0	682,3
δT5	0	42-60	FFESFGDLSSPDAVMGNPK	2059,9	2060,9	1031,0	691,6
δT5*	1	42-60	FFESFGDLSIPDAVMGNPK	2057,9	2058,9	1030,0	687,0
δT5*oxy	2	42-60	FFESFGDLSIPDAVMGNPK	2073,9	2074,9	1038,0	692,3
δT6	0	61-62	VK	245,2	246,2	123,6	82,7
δT7	0	63-66	AHGK	411,2	412,2	206,6	138,1
δT9	0	68-83	VLGAFSDGLAHLDNLK	1668,9	1669,9	835,4	557,3
δT10	0	84-96	GTFSQSELHCDK	1463,7	1464,7	732,8	488,9
δT11	0	97-105	LHVDPENFR	1125,6	1126,6	563,8	376,2
δT12	0	106-117	LLGNVLCVRLAR	1268,8	1269,8	635,4	423,9
δT13	0	118-121	NFGK	464,2	465,2	233,1	155,7
δT14	0	122-133	EFTPMQAAYQK	1440,7	1441,7	721,3	481,2
δT15	0	134-145	VVAGVANALAHK	1148,7	1149,7	575,3	383,9
δT16	0	146-147	YH	318,1	319,1	160,1	107,0
δT1 + δT2	1	2-18	VHLTPEEKTAVNALWGK	1892,0	1893,0	947,0	631,7
δT2 + δT3	1	10-31	TAVNALWGKVNVDVAVGGEAL GR	2196,2	2197,2	1099,1	733,1
δT3 + δT4	1	19-41	VNVDAVGGEALGRLLVVYPW TQR	2511,4	2512,4	1256,7	838,1
δT4 + δT5	1	32-60	LLVVYPWTQRFFESFGDLSS PDAVMGNPK	3299,6	3300,6	1650,8	1100,9
δT5 + δT6	1	42-62	FFESFGDLSSPDAVMGNPKV	2271,1	2272,1	1136,5	758,0
δT6 + δT7	1	61-66	VKAHGK	638,4	639,4	320,2	213,8
δT7 + δT8	1	63-67	AHGKK	539,3	540,3	270,7	180,8
δT8 + δT9	1	67-83	KVLGAFSDGLAHLDNLK	1797,0	1798,0	899,5	600,0
δT9 + δT10	1	68-96	VLGAFSDGLAHLDNLKGTFS QSELHCDK	3114,6	3115,6	1558,3	1039,2
δT10 + δT11	1	84-105	GTFSQSELHCDKLVHVDPEN FR	2571,2	2572,2	1286,6	858,1
δT11 + δT12	1	97-117	LHVDPENFRLLGNVLCVRLA	2376,3	2377,3	1189,2	793,1
δT12 + δT13	1	106-121	LLGNVLCVRLARNFGK	1715,0	1716,0	858,5	572,7
δT13 + δT14	1	118-133	NFGKEFTPMQAAYQK	1886,9	1887,9	944,5	630,0
δT14 + δT15	1	122-145	EFTPMQAAYQKVVAGVANA LAHK	2571,3	2572,3	1286,7	858,1
δT15 + δT16	1	134-147	VVAGVANALAHKYH	1448,8	1449,8	725,4	483,9

Références bibliographiques du chapitre 3

1. Weed, R. I.; Reed, C. F.; Berg, G. Is hemoglobin an essential structural component of human erythrocyte membranes? *The Journal of clinical investigation* **1963**, 42, 581-8.
2. Wood, W. G.; Weatherall, D. J. Haemoglobin synthesis during human foetal development. *Nature* **1973**, 244, (5412), 162-5.
3. Passarge, E. Color atlas of genetics. 2nd ed.; *Thieme: Stuttgart ; New York*, 2001; p xi, 457 p.
4. Blouquit, Y.; Bardakdjian, J.; Lena-Russo, D.; Arous, N.; Perrimond, H.; Orsini, A.; Rosa, J.; Galacteros, F. Hb Bruxelles: alpha 2A beta (2)41 or 42(C7 or CD1)Phe deleted. *Hemoglobin* **1989**, 13, (5), 465-74.
5. Bradley, T. B., Jr.; Wohl, R. C.; Rieder, R. F. Hemoglobin Gun Hill: deletion of five amino acid residues and impaired heme-globin binding. *Science* **1967**, 157, (3796), 1581-3.
6. Ingram, V. M. Gene mutations in human haemoglobin: the chemical difference between normal and sickle cell haemoglobin. *Nature* **1957**, 180, (4581), 326-8.
7. Bunn, H. F.; Schmidt, G. J.; Haney, D. N.; Dluhy, R. G. Hemoglobin Cranston, an unstable variant having an elongated beta chain due to nonhomologous crossover between two normal beta chain genes. *Proceedings of the National Academy of Sciences of the United States of America* **1975**, 72, (9), 3609-13.
8. Embury, S. H. Sickle cell disease : basic principles and clinical practice. *Raven Press (NY)*, 1994;.
9. OMS. La drépanocytose dans la région africaine: situation actuelle et perspectives. *Rapport de l'OMS, Comité Régionale de l'Afrique* **2006**.
10. Benson, J. M.; Therrell, B. L., Jr. History and current status of newborn screening for hemoglobinopathies. *Seminars in perinatology* **2010**, 34, (2), 134-44.
11. Clarke, G. M.; Higgins, T. N. Laboratory investigation of hemoglobinopathies and thalassemias: review and update. *Clinical chemistry* **2000**, 46, (8 Pt 2), 1284-90.
12. Bisse, E.; Hovasse, A.; Preisler-Adams, S.; Epting, T.; Wagner, O.; Kogel, G.; Van Dorsselaer, A.; Schaeffer-Reiss, C. Hb Riesa or beta93 (F9) Cys-->Ser, a new electrophoretically silent haemoglobin variant interfering with haemoglobin A1c measurement. *J Chromatogr B Analyt Technol Biomed Life Sci* **2011**, 879, (28), 2952-6.
13. Ryan, K.; Bain, B. J.; Worthington, D.; James, J.; Plews, D.; Mason, A.; Roper, D.; Rees, D. C.; de la Salle, B.; Streetly, A. Significant haemoglobinopathies: guidelines for screening and diagnosis. *British journal of haematology* **2010**, 149, (1), 35-49.
14. Miyazaki, A.; Nakanishi, T.; Kishikawa, M.; Shimizu, A.; Ohba, Y.; Tanaka, A.; Sasaki, H. Post-translational modification from methionine to aspartic acid-residue on a variant hemoglobin, Hb Bristol, a proof by ESI-MS-MS. *Journal of mass spectrometry* **1996**, 31, (11), 1311-3.
15. Gevaert, K.; Goethals, M.; Martens, L.; Van Damme, J.; Staes, A.; Thomas, G. R.; Vandekerckhove, J. Exploring proteomes and analyzing protein processing by mass spectrometric identification of sorted N-terminal peptides. *Nature biotechnology* **2003**, 21, (5), 566-9.
16. Lindskog, S. Structure and mechanism of carbonic anhydrase. *Pharmacology & therapeutics* **1997**, 74, (1), 1-20.

Chapitre 4 :

Caractérisation des glycosylations de l'invertase vacuolaire de raisin

Cette étude a été réalisée en collaboration avec le Dr. S. Jégou de l'Université de Reims Champagne-Ardenne.

Ce sujet s'inscrit dans la continuité d'une première thèse réalisée au laboratoire par Agnès Hovasse¹. Au cours de ce travail de thèse, une stratégie d'analyse par LC-MS/MS de glycopeptides intacts a été développée et appliquée à la caractérisation d'une partie des N-glycosylations de l'invertase vacuolaire de raisin.

L'objectif du travail présenté dans ce chapitre de thèse était de réaliser une caractérisation complète de la glycosylation de l'invertase vacuolaire du raisin en utilisant une combinaison d'outils de spectrométrie de masse.

I. Contexte de cette étude

1. Invertase vacuolaire: fonction et rôle

L'une des principales caractéristiques des plantes est leur capacité à réduire le CO₂ en présence de la lumière du soleil et de l'eau en sucres, et le transport ultérieur du carbone assimilé dans les tissus non photosynthétiques. Dans la plupart des plantes, le sucre transporté est un suc, un disaccharide (saccharose) non-réduit dans lequel glucose (Glc) et fructose (Fru) sont liés. Selon le besoin des activités physiologiques et biochimiques des tissus hétérotrophes, le suc est véhiculé dans différents compartiments subcellulaires. Il peut entrer dans la glycolyse et le cycle de Krebs pour la production d'ATP (adénosine triphosphate) et NADH (nicotinamide adénine dinucléotide). Les atomes de carbone du disaccharide peuvent également être utilisés pour la biosynthèse des métabolites primaires importants pour la croissance et le développement des tissus de végétaux. De même, le suc peut être converti en polymères comme l'amidon, ou en composés secondaires de stockage des plantes, qui vont permettre à la plante de faire face à des prédateurs et à des parasites ou d'autres défis environnementaux.

L'utilisation du suc comme source de carbone et d'énergie dépend de son clivage en hexoses et chez les plantes il existe deux types d'enzymes qui catalysent cette réaction. Ces deux enzymes sont, la sucrose synthase (EC 2.4.1.13) et les invertases (EC 3.2.1.26)². L'enzyme EC 2.4.1.13 est une glycosyltransférase qui convertit le suc en présence de l'UDP (uridine diphosphate) en UDP-Glc et UDP-Fru³. L'invertase est l'enzyme (hydrolase, β -fructofuranosidase) qui hydrolyse le saccharose en deux monosaccharides, le glucose et le fructose (Figure 1).

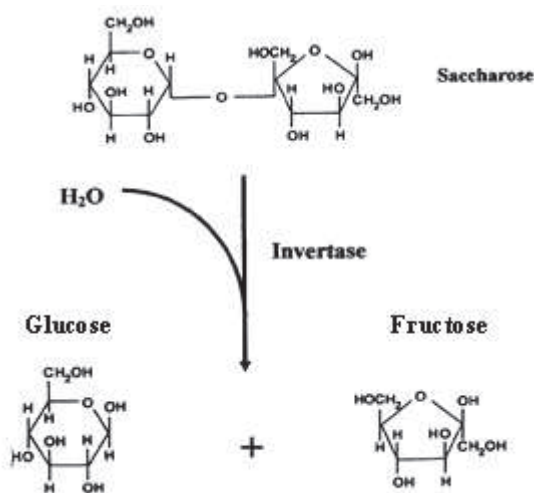
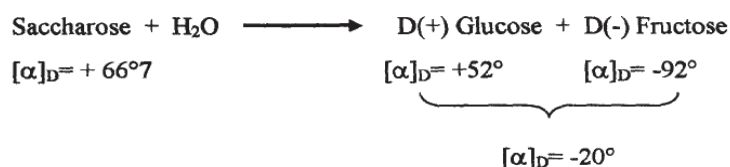


Figure 1. Schéma de l'hydrolyse du saccharose par l'invertase⁴.

Par cette hydrolyse, l'invertase provoque l'inversion du pouvoir rotatoire entre le substrat (saccharose, dextrogyre) et le produit (mélange de glucose et le fructose, lévogyre), d'où le nom de sucre inverti pour ce mélange et par extrapolation invertase pour l'enzyme.



L'invertase existe sous plusieurs isoformes, qui ont des propriétés biochimiques différentes et des localisations subcellulaires différentes⁵. Chez les plantes, trois formes d'invertase ont été rapportées et sont classées selon le pH optimal de leur activité *in vitro*, leur solubilité et leur localisation cellulaire. Ainsi, il existe, une forme pariétale insoluble avec une activité optimale à pH 4,5-5, et deux formes solubles, l'une localisée dans le cytoplasme et la dernière dans la vacuole, présentant respectivement des valeurs de pH optimal de 7-8 et 4,5-5,5. Les fonctions spécifiques des différents isoformes d'invertase ne sont pas claires,

mais semblent réguler l'utilisation du saccharose dans les différentes voies métaboliques. Parce que les sucres dans les plantes ne sont pas seulement des nutriments, mais aussi d'importants régulateurs de l'expression des gènes⁶, l'invertase peut donc être indirectement impliquée dans le contrôle de la différenciation cellulaire et le développement des plantes^{7,8}.

Chez la vigne, l'accumulation des hexoses dans les baies de raisin s'accompagne d'une augmentation de l'expression des gènes des invertases vacuolaires (GIN1 et GIN2) dans les 8 premières semaines de pré-maturation⁹. L'invertase vacuolaire du raisin est une glycoprotéine et constitue l'une des protéines majeures du jus de raisin ou moût et du vin^{4,10}. De plus des études biochimiques ont montré que cette glycoprotéine serait impliquée dans les propriétés moussantes des vins de Champagne^{4,11}.

2. Problématique et objectifs de l'étude

Pour améliorer la compréhension du rôle cette enzyme clé du métabolisme des sucres du raisin qu'est l'invertase vacuolaire de raisin, nous avons développé une stratégie basée sur les outils de protéomique pour caractériser cette glycoprotéine.

L'invertase vacuolaire du raisin est une glycoprotéine d'environ 60-70 kDa qui possède 12 sites potentiels de N-glycosylation dont seulement 2 sont décrites¹². La caractérisation complète des glycosylations de l'invertase de raisin est essentielle pour mieux comprendre leur impact sur sa stabilité, sa solubilité, son hydrophobie, sa conformation, son activité biologique, sa résistance potentielle aux protéases et son rôle dans les propriétés moussantes des vins effervescents dont le Champagne. Une étude glycomique portée sur l'analyse des chaînes de glycanes de l'invertase libérés après déglycosylation à la PNGase A⁴ a montré que l'invertase vacuolaire du raisin porte des glycanes de type paucimannosidique (Figure 2).

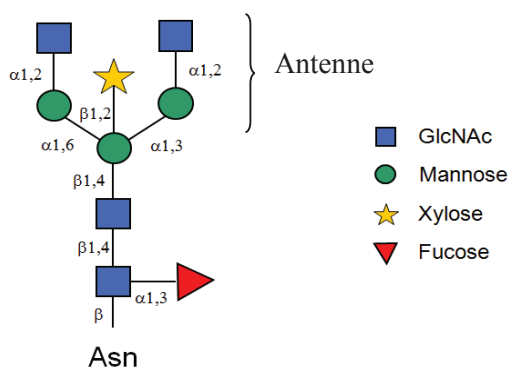


Figure 2. Structure d'une glycosylation de type paucimannosidique observé chez l'invertase vacuolaire de raisin (Dambrouck 2004).

Ce type de glycosylation comprend un 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. Cette glycosylation est caractéristique des glycoprotéines de plantes.

Dans un mélange protéique complexe (contenant des centaines de protéines), la quantité d'une glycoprotéine peut représenter une infime proportion de l'ensemble des protéines. La digestion enzymatique de ce mélange va générer des milliers de peptides qui rendent le milieu encore plus complexe. Les peptides provenant d'une glycoprotéine mineure peuvent donc être masqués par les autres peptides issus des autres protéines lors d'une analyse par nanoLC-MS/MS. Pour mener une étude de caractérisation fine de glycoprotéine, il est donc nécessaire de disposer de quantité plus importante et suffisante de la protéine d'intérêt purifiée que pour une analyse protéomique (de l'ordre du μg pour pouvoir réaliser plusieurs types d'analyse) pour étudier la macro-hétérogénéité (l'hétérogénéité d'occupation des sites) et la micro-hétérogénéité (l'hétérogénéité de structures des glycanes). En effet, une glycoprotéine peut souvent comporter plusieurs sites de glycosylation avec une occupation plus ou moins partielle et chacun de ces sites peut porter également plusieurs oligosaccharides différents donc plusieurs glycoformes. Le signal du glycopeptide portant un site de glycosylation est donc réparti entre les différentes glycoformes, ce qui réduit l'intensité de signal et donc la possibilité de détecter ces glycopeptides. Pour ces multiples raisons, la caractérisation fine d'une glycoprotéine passe, soit par une étape de purification de cette protéine d'intérêt, et/ou soit par l'enrichissement des glycopeptides de celle-ci dans un milieu complexe à l'aide de différentes techniques d'enrichissement ou de purification: affinité à la lectine (concanavaleine A, ConA), phase ZIC-HILIC), le TiO_2 , l'hydrazide, phases greffées avec l'acide phénylborique¹³, chromatographie échangeuse d'anion etc.

Parmi ces techniques, la chromatographie d'interaction hydrophile sur phase ZIC-HILIC (Zwitterionic Hydrophilic Interaction Liquid Chromatography) est largement utilisée pour l'enrichissement et l'analyse des composés polaires et hydrophiles comme les glycoprotéines^{14, 15}. Les propriétés physico-chimiques des glycopeptides leur confèrent un caractère plus hydrophile et polaire que les peptides non glycosylés dans un milieu complexe à cause des oligosaccharides qu'ils portent (par la présence de groupements OH des oligosaccharides). La sélectivité de l'interaction hydrophile des phases ZIC-HILIC pour les composés polaires et hydrophiles lui permet de retenir dans un milieu complexe toutes sortes de glycopeptides qui pourront être élués séparément contrairement à la lectine qui par exemple n'est spécifique qu'à certains glycanes donnés de type mannose et glucose¹⁶. De plus, il a été démontré que les techniques d'enrichissement des glycopeptides communément utilisées à l'instar des lectines ou l'hydrazide sont incomplets, car des quantités considérables d'espèces glycosylées restent toujours dans les flux à travers les colonnes^{17, 18}. Quant aux techniques d'enrichissement chimiques comme le TiO₂ ou utilisant les liaisons covalentes avec les phases hydrazides quoique beaucoup plus spécifiques aux glycopeptides, elles sont beaucoup plus difficiles à mettre en œuvre et ne permettent pas de conserver l'information sur la nature de l'oligosaccharide sur les glycopeptides. L'enrichissement sur phases ZIC-HILIC a donc été préféré aux autres techniques d'enrichissement pour les multiples raisons présentées précédemment.

Nous avons donc réalisé une étude structurale fine des sites de glycosylation de l'invertase vacuolaire de raisin et de leur micro-hétérogénéité en utilisant une combinaison de différentes stratégies de spectrométrie de masse sur l'invertase vacuolaire purifiée à partir de moût¹⁹.

Le travail préliminaire sur la caractérisation de l'invertase vacuolaire du raisin menée au laboratoire¹ avait permis d'identifier un site de glycosylation sur l'asparagine en position 430 (Asn 430). Sur ce site, 11 glycoformes ont été identifiées grâce à l'analyse de glycopeptides, tandis que 5 sites (Asn 225, 227, 289, 430, 499) de glycosylation ont pu être identifiés grâce à l'analyse des peptides déglycosylés.

Nous avons donc poursuivi ce travail de caractérisation des N-glycosylations de l'invertase en combinant plusieurs digestions enzymatiques avec les stratégies basées sur l'analyse de glycopeptides et des peptides déglycosylés, et en tirant partie des performances des différents spectromètres de masse de type Q-TOF et trappe ionique.

II. Identification et caractérisation de la glycosylation de l'invertase vacuolaire de raisin

1. Stratégie d'analyse protéomique

a. Description des échantillons

L'invertase vacuolaire de raisin a été purifiée en une seule étape à partir de moût par chromatographie échangeuse d'anions¹⁹. La pureté et l'identité de la protéine ont été vérifiées par analyse nanoLC-MS/MS. L'invertase vacuolaire de raisin purifiée a été lyophilisée et conservée à -20°C pour éviter son éventuelle dégradation.

L'invertase a été par la suite mise en solution et soumise à des étapes de réduction/alkylation et de digestion en milieu liquide suivant les protocoles adoptés.

b. Stratégie d'analyse par spectrométrie de masse

Le schéma de stratégie d'analyse nanoLC-MS/MS pour l'identification et la caractérisation des sites et glycoformes de l'invertase vacuolaire est représentée sur la Figure 3.

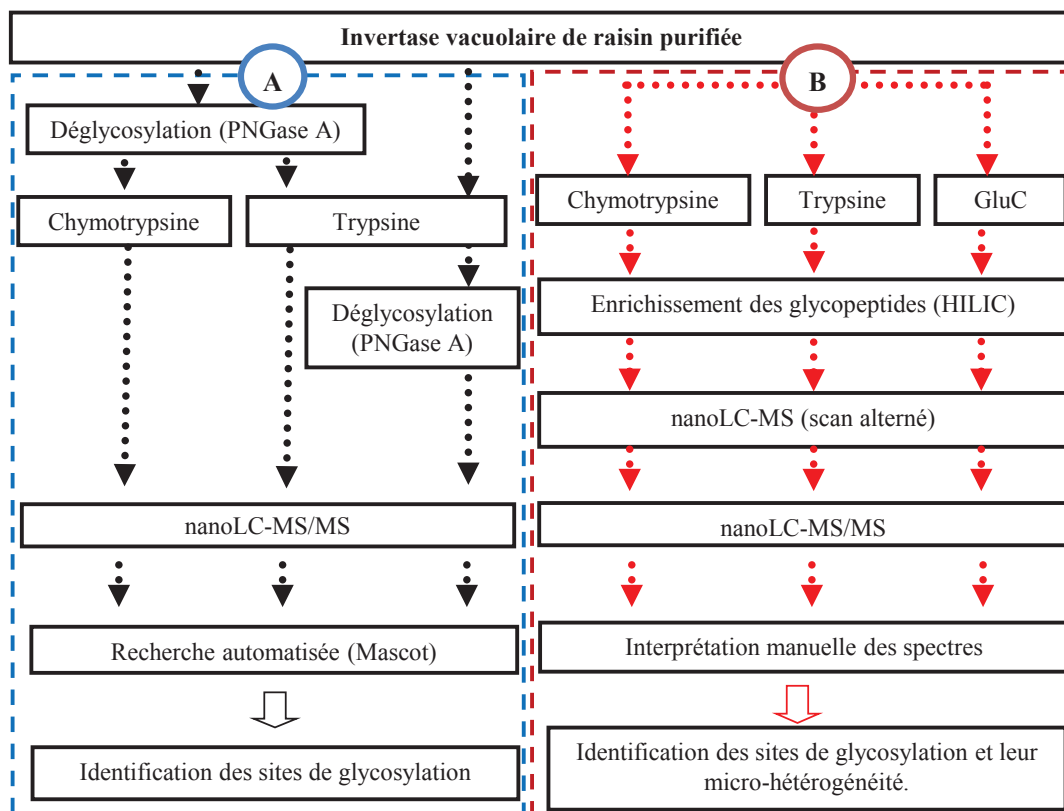


Figure 3. Stratégie globale d'analyse nanoLC-MS et MS/MS adoptée pour l'identification et caractérisation des sites et micro-hétérogénéité de l'invertase vacuolaire de raisin. En Bleu, (A) protocole d'analyse de peptides déglycosylés, en rouge (B) protocole d'analyse de glycopeptides.

Le détail de l'ensemble des protocoles expérimentaux (digestions, enrichissement ZIC-HILIC, analyses MS et identification des glycopeptides) est décrit dans la partie expérimentale générale.

i. Protocole d'analyse A (peptides déglycosylés)

Pour la stratégie basée sur l'analyse de peptides déglycosylés (Protocole A, Figure 3, en bleu), l'invertase a été digérée par la trypsine ou chymotrypsine en parallèle avant ou après une digestion à la PNGase A (Peptide : N-glycosydase A). La PNGase A est une enzyme (amidase) qui clive entre l'asparagine de la chaîne

de protéine ou un peptide et le GlcNac proximal de la chaîne glycanique des glycoprotéines avec ou sans fucose en position $\alpha(1,3)$.

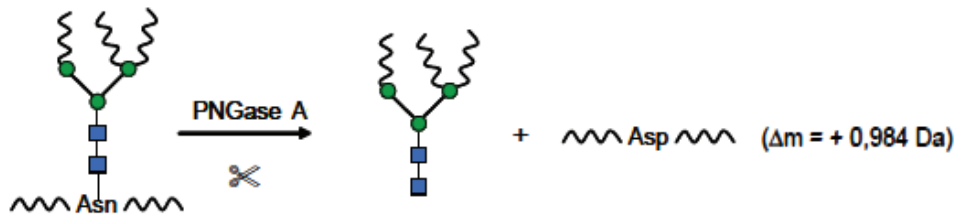


Figure 4. Schéma de la déglycosylation d'un glycopeptide ou d'une protéine par la PNGase A. La PNGase A permet de décrocher l'oligosaccharide, et l'asparagine du site de glycosylation est transformée en acide aspartique sur le peptide nu, ce qui entraîne une incrémentation de + 0,984 Da sur le peptide nu.

Cette déglycosylation entraîne une déamidation de l'asparagine portant la glycosylation en acide aspartique ($N \rightarrow D$) et provoque ainsi une différence de masse de +1 Da sur les asparagines. Cette différence de masse est suffisante pour être détectée sur le peptide ayant porté une glycosylation avec les spectromètres de masse utilisés dans cette étude. Cette stratégie permet d'obtenir de façon indirecte l'information sur les sites de N-glycosylation mais l'information des glycanes qui sont attachés est perdue pour une glycoprotéine qui comporte plusieurs sites de N-glycosylation comme l'invertase. C'est pourquoi, nous avons mis au point en parallèle à cette stratégie une autre basée sur l'analyse de glycopeptides qui permet de conserver l'information sur le site de N-glycosylation et des glycanes.

ii. Protocole d'analyse B (glycopeptides intacts)

Pour la stratégie basée sur l'analyse des glycopeptides intacts, l'invertase a été digérée parallèlement avec la trypsine, la chymotrypsine ou encore la GluC (Figure 5).

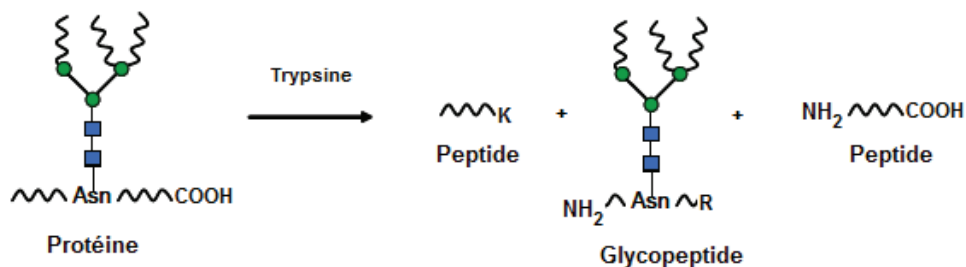


Figure 5. Schéma de la digestion d'une protéine à la trypsine. La trypsine permet de rompre les liaisons peptides après la lysine (K) ou l'arginine (R) dans la séquence protéique. Le digeste obtenu est composé de peptides et glycopeptides. C'est ce digeste qui sera enrichi en glycopeptide sur phase ZIC-HILIC pour l'analyse nanoLC-MS et nanoLC-MS/MS.

Les digestes obtenus ont ensuite été enrichis en glycopeptides par chromatographie d'interaction hydrophile (phase ZIC-HILIC) et analysés par nanoLC-MS suivies nanoLC-MS/MS.

Différents enzymes de digestion ont été utilisées (la trypsine, la chymotrypsine et la GluC). Ces enzymes ont été choisies après des tests de digestion *in silico*, dans le but d'obtenir des peptides et glycopeptides de tailles homogènes, et ne portant pas plus d'un ou deux sites de glycosylation. Les digestes obtenus ont été analysés par nanoLC-MS/MS sur le Q-TOF maXis et la trappe ionique amaZon (Bruker).

Avec une précision de masse de $\pm 0,2$ Da en mode MS et MS/MS pour la trappe ionique amaZon, et 5 ppm en MS et 0,02 Da en MS/MS pour le maXis, les deux spectromètres de masse possèdent la précision de masse suffisante pour mesurer la différence de masse de 1 Da sur le peptide déglycosylé.

2. Principaux résultats

a. Identification des sites de glycosylation des peptides déglycosylés

Afin d'identifier les sites d'attachement des glycanes, nous avons procédé à la déglycosylation de l'invertase vacuolaire de raisin à l'aide de la PNGase A pour cliver les glycanes (Figure 6). Cette figure montre une perte (shift) de masse de l'invertase déglycosylée, lié à la perte des oligosaccharides de l'invertase ce qui prouve une déglycosylation de la protéine. L'identification des peptides déamidés par la digestion à la PNGase A a permis de mettre en évidence l'asparagine ayant porté les chaînes glycaniques clivées.

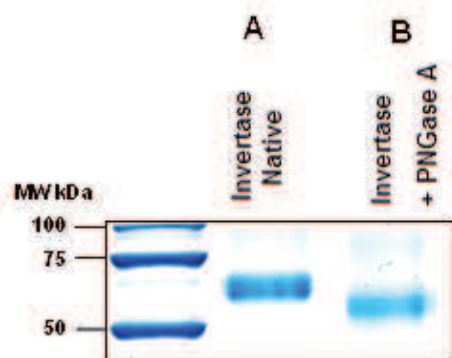


Figure 6. Gel SDS-PAGE coloré au bleu de Coomassie. De gauche à droite, piste des marqueurs de masse moléculaires, piste (A) invertase vacuolaire de raisin native purifiée, piste (B) invertase vacuolaire de raisin déglycosylée avec la PNGase A. La déglycosylation a conduit à une perte de masse (shift) dans la migration de l'invertase vacuolaire déglycosylée d'environ 10 kDa. L'invertase a bien donc été déglycosylée par la PNGase A.

Résultats

Ces analyses ont permis d'identifier une isoforme de l'invertase vacuolaire annotée dans NCBI : PREDICTED: beta-fructofuranosidase, soluble isoenzyme I-like (gi|225466093) avec une couverture de séquence de 70% pour la chymotrypsine et de 50 % pour la trypsine. Cette isoforme de l'invertase a été identifiée avec 4 peptides discriminants (digestions tryptique et chymotrypsique confondues) par rapport à l'invertase GIN1 (gi|1839578) (Figure 7).

β-fruc	1MDTHHATSHD	LLHDEAYAPL	PYSPLPDHPV	PAGAPALGQR	RPLKVF ^{AV} TL	ASLMFLASLV	ALVIHQSPQT	PVTVDNDGPS
GIN1	MDTHHATSHD	LLHDEAYAPL	PYSPLPDHPV	PAGAPALGQR	RPLKV CVTL	SLMFLVSLV	SFI RV	HVTVDNDGPS
	*****	*****	*****	*****	*****	*****	*****	*****
β-fruc	81MEREGRGVAQ	GVSEKSFLGF	SGRRLSY ^{NWT}	TAMFAWQ ^{ETA}	FHFQPEKNWM	NDPDGPLFHM	GWYHLFYQYN	PDSAVWGNIT
GIN1	MEREGRGVAQ	GVSEKSFLGF	SGRRLSY ^{NWT}	TAMFAWQ ^{ETA}	FHFQPEKNWM	NDPDGPLFHM	GWYHLFYQYN	PDSAVWGNIT
	*****	*****	*****	*****	*****	*****	*****	*****
β-fruc	161WGHAVSRDMI	HWLYLPLAMV	PDRWFDLNGV	WTGSATILPN	GQIIMLYTGD	T ^{NDSVQVQNL}	AYPANLSDPL	LLHWIKYENN
GIN1	WGHAVSRDMI	HWLYLPLAMV	PDRWFDLNGV	WTGSATILPN	GQIIMLYTGD	T ^{NDSVQVQNL}	AYPANLSDPL	LLHWIKYENN
	*****	*****	*****	*****	*****	*****	*****	*****
β-fruc	241PVMVPPAGIG	SDDFRDPTTM	WVGADGNWRV	AVGSLV ^{NTTG}	IVLVFQT ^{TNF}	TDFELLDGEL	HGVPGTGMWE	CVDFYPVSIN
GIN1	PVMVPPAGIG	SDDFRDPTTM	WVGADGNWRV	AVGSLV ^{NTTG}	IVLVFQT ^{TNF}	TDFELLDGEL	HGVPGTGMWE	CVDFYPVSIN
	*****	*****	*****	*****	*****	*****	*****	*****
β-fruc	321GVYGLDTSAH	GPGIKHVLKA	SMDDNRHDY	ALGEYDPMTD	TWTPDDPELD	VGIGRLDY ^G	KYASKTFYD	QVKKRRILYG
GIN1	GVYGLDTSAH	GPGIKHVLKA	SMDDNRHDY	ALGEYDPMTD	TWTPDDPELD	VGIGRLDY ^E	RYYASKTFYD	QVKKRRILYG
	*****	*****	*****	*****	*****	*****	*****	*****
β-fruc	401WISEGDIESD	DLKKGWASLQ	SIPRTVLHD ^N	KTGTYL ^{LLWP}	IEEVESL ^{RTN}	STEFEDV ^{LLE}	PGSV ^{VPLD} IG	SASQLDIVAE
GIN1	WISEGDIESD	DLKKGWASLQ	SIPRTVLHD ^N	KTGTYL ^{LLWP}	IEEVESL ^{RTN}	STEFEDV ^{LLE}	PGSV ^{VPLD} IG	SASQLDIVAE
	*****	*****	*****	*****	*****	*****	*****	*****
β-fruc	481FEVDNETLEA	MVEADVIY ^{NC}	STSAGAAGRG	ALGPFGL ^{LV}	ADDTLSELTP	IYFYIAKDTD	GSYKTFCTD	LSRSSLAVDD
GIN1	FEVDNETLEA	MVEADVIY ^{NC}	STSAGAAGRG	ALGPFGL ^{LV}	ADDTLSELTP	IYFYIAKDTD	GSYKTFCTD	LSRSSLAVDD
	*****	*****	*****	*****	*****	*****	*****	*****
β-fruc	561VDQR ^{LYGSIV}	P ^{VL} DDEKPTM	RVLVDHSIVE	GFSQGRSCI	TTRVYPTEAI	YGAARLFLFN	NATGV ^{NVTAS}	IKIWEMASAD
GIN1	VDQR ^{LYGSIV}	P ^{AL} DDEKPTM	RVLVDHSIVE	GFSQGRSCI	TTRVYPTEAI	YGAARLFLFN	NATGV ^{NVTAS}	IKIWEMASAD
	*****	*****	*****	*****	*****	*****	*****	*****
β-fruc	641IHPYPLDQP							
GIN1	IHPYPLDQP							

Figure 7. Alignement de séquences de la PREDICTED: beta-fructofuranosidase, soluble isoenzyme I-like (gi|225466093) identifiée contre la GIN 1 (gi|1839578). Les sites consensus de glycosylation sont indiqués en rouge, les acides aminés surlignés en gris sont différents selon les isoformes d'invertase. Les peptides discriminants identifiés pour la PREDICTED: beta-fructofuranosidase, soluble isoenzyme I-like (gi|225466093) sont encadrés (□).

L'interprétation des données a été effectuée à l'aide de l'algorithme de recherche Mascot, dans une banque restreinte à la taxonomie *Vitis vinifera* extraite de la banque NCBI non redondante (NCBIInr). Lors de

cette recherche, en plus de la carbamidométhylation (C) et oxydation de la méthionine (M), la modification de (N) → (D) a été prise en compte comme une modification variable. Les spectres MS/MS de peptides modifiés qui sont identifiés comme portant la modification (N) → (D) ont été validés avec un score d'identification Mascot supérieur à 30 (ex. Figure 8 et 9).

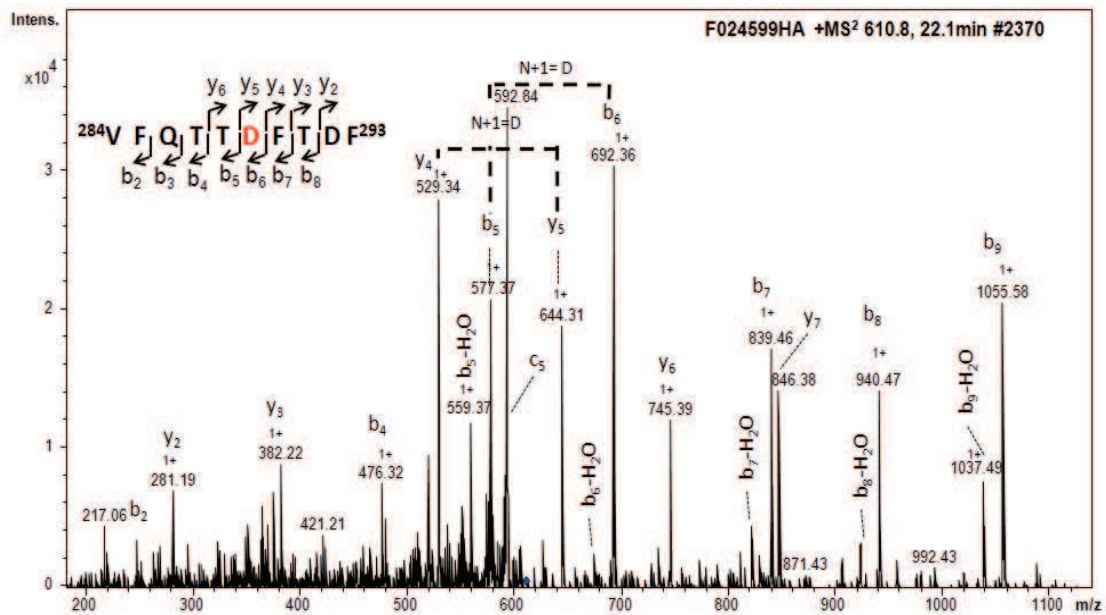


Figure 8. Spectre MS/MS annoté du peptide chymotrypsique de $m/z = 610,8 (+2)$ portant le site de glycosylation N^{289} transformé en D^{289} après déglycosylation par la PNGase A et ayant permis l'identification du site de glycosylation N^{289} . La différence de masse entre y_4 et y_5 ou b_5 et b_6 correspond à $N+1 = D$.

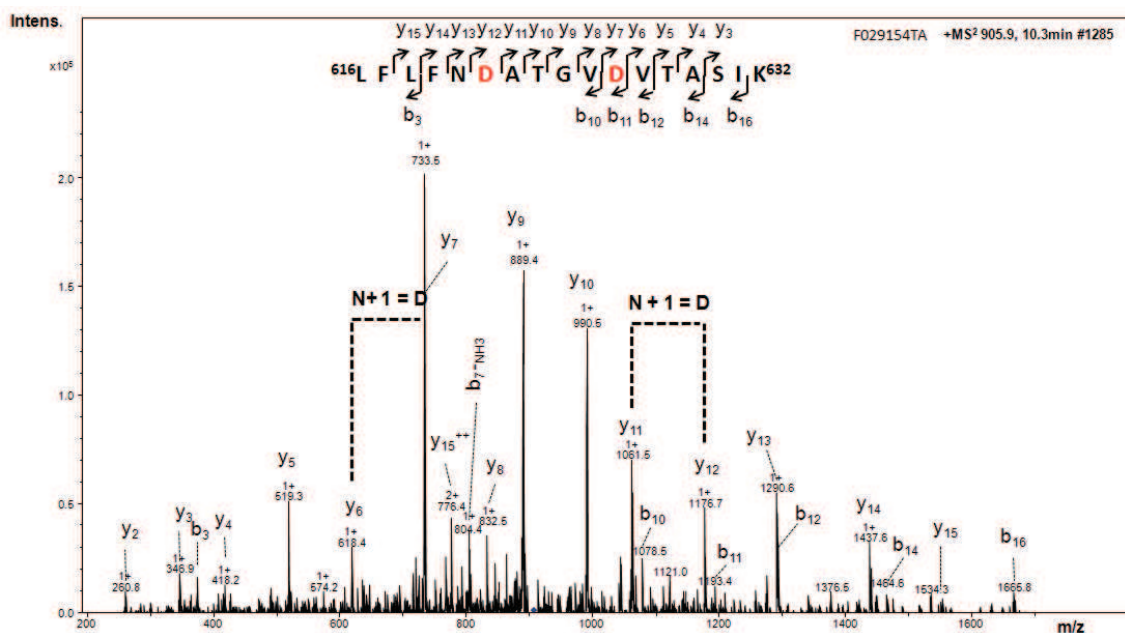


Figure 9. Spectre MS/MS annoté du peptide trypsinique $m/z = 905,9 (2+)$ portant les deux sites de glycosylation N^{621} et N^{626} transformés respectivement en D^{621} et D^{626} après déglycosylation par la PNGase A et qui ont permis l'identification des sites de glycosylation N^{621} et N^{626} . La différence de masse entre y_6 et y_7 ou y_{11} et y_{12} correspond à $N+1 = D$.

Ainsi, 9 sites de glycosylation (sites Asn : 108 ; 225 ; 277 ; 289 ; 430 ; 450 ; 499 ; 621 ; 626) ont pu être identifiés comme portant une déamidation sur les asparagines des peptides identifiés (Tableau 1). De plus, un site potentiel de N-glycosylation a été identifié sans déamidation (Asn 212). Ce qui indique que tous les sites potentiels ne sont pas occupés par des chaînes glycaniques.

Les sites Asn 621 et Asn 626 n'ont pu être identifiés que par cette approche de déglycosylation des glycopeptides après digestion trypsinique de l'invertase vacuolaire de raisin. La proximité des deux sites de glycosylation et l'encombrement stérique créé par les glycanes présents sur ces deux sites pourraient expliquer

que la déglycosylation à la PNGase A est plus efficace lorsqu'elle est réalisée sur les glycopeptides plutôt que sur la protéine entière. De plus, la proximité de ces sites de glycosylation pourrait expliquer pourquoi seul ce peptide, comportant les deux sites de glycosylation uniquement, a pu être identifié dans l'étude du glycoprotéome de vin ¹².

Tableau 1. Résumé des résultats d'identification des sites de glycosylation obtenus par déglycosylation et digestion chymotrypsique (C) ou trypsique (T) et analyse nanoLC-MS/MS sur trappe ionique amaZon (IT) ou sur Q-TOF maXis. A chaque peptide est associé un score d'identification Mascot, le m/z et l'état de charge de l'ion parent sélectionné pour la MS/MS, les masses expérimentales et théoriques. (C(CAM)), peptide avec une cystéine carbamidométhylée.

Sites de Glycosylation	Séquences identifiées	Enzyme	Instrument MS	Mascot ion score	Ion parent (m/z) - état de charge	Masse expérimentale (Da)	Masse calculée du peptide déglycosylé (Da)
Peptide avec site consensus de N-glycosylation non déamidé							
212	²⁰⁸ TGDTNDSVQVQNL ²²⁰	C	IT	47	695.80 (2+)	1389.59	1389.65
Peptide avec site consensus de N-glycosylation déamidé							
108	¹⁰⁵ LSYDWTAMFAWQR ¹¹⁸	T	IT	88	888.45 (2+)	1774.88	1773.83
225	²²¹ AYPADLSDPL ²³⁰	C	IT	74	531.25 (2+)	1060.49	1059.52
	²²¹ AYPADLSDPLLL232	C	IT	52	644.35 (2+)	1286.68	1285.69
	²²¹ AYPADLSDPLLLHW ²³⁴	C	IT	43	805.84 (2+)	1609.67	1608.83
	²⁰⁸ TGDTNDSVQVQNLAYPADLSDPLLL ²³²	C	Q-TOF	39	1330.16 (2+)	2658.3	2657.32
277	²⁸⁵ RVAVGSLVDTTGIVL ²⁸³	C	IT	63	750.43 (2+)	1498.84	1497.89
		C	Q-TOF	53	750.44 (2+)	1498.87	
289	²⁸⁴ VFQTTDFDTDF ²⁹³	C	IT	66	610.78 (2+)	1219.52	1218.56
430	⁴²⁰ QSIPRTLHDDKTGT ⁴³⁵	C	IT	45	610.99 (3+)	1829.97	1828.94
	⁴²⁵ TVLHDDKTGTYLLWPIEEVESLR ⁴⁴⁸	T	IT	32	943.12 (3+)	2826.34	2825.5
450	⁴⁴⁷ RTDSTEFEDVLEPGSVVPL ⁴⁶⁸	C	Q-TOF	72	735.04 (3+)	2202.1	2201.12
499	⁴⁹⁹ DC(CAM)STSAGAAGRGAALGPF ⁵¹⁵	C	IT	135	797.83 (2+)	1593.65	1592.74
		C	Q-TOF	64	797.87 (2+)	1593.72	
		C	IT	60	939.40 (2+)	1876.78	
621 and 626	⁶¹⁶ LFLFNATGVDVTASIK ⁶³²	T*	IT	106	905.94 (+2)	1809.89	1807.98

Conclusion

Grâce à la stratégie d'identification des sites de glycosylation basée sur l'analyse des peptides déglycosylés de l'invertase vacuolaire de raisin, nous avons pu identifier au total 9 sites de glycosylation (sites Asn: 108 ; 225 ; 277 ; 289 ; 430 ; 450 ; 499 ; 621 ; 626), donc 4 sites supplémentaires par rapport à l'étude préliminaire de caractérisation des glycosylations de cette glycoprotéine (Hovasse 2010). De plus, nous avons montré que tous les sites potentiels de N-glycosylation comme par exemple le site Asn 212, peuvent être ou non partiellement glycosylés. Cependant, même si dans notre cas (pH 7-8) nous n'étions pas dans des conditions de pH basiques extrêmes, il faut noter que dans ces conditions (pH 10-11), les déamidations naturelles ont été mises en évidence ²⁰. Une différence de masse de 1 Da peut donc être due à une déamidation naturelle qui provoque des faux positifs ^{20, 21}. C'est pourquoi certaines études utilisant cette stratégie d'identification de site de glycosylation préconisent aussi l'étude de peptides déglycosylés dans de l'eau lourde. En effectuant la déglycosylation avec la PNGase A dans de l'eau lourde (H₂O¹⁸), ceci entraînerait une différence de masse de 3 Da lors de la transformation de l'asparagine (N) en acide aspartique (D) ^{20, 22}, ce qui permettrait de lever l'ambiguïté sur l'origine de la déamidation.

L'étape d'interprétation de données de spectres MS/MS a été facilitée par la prise en compte de la modification de (N) → (D) lors de la recherche à l'aide de l'algorithme Mascot. Cette recherche a permis, en effet, de retrouver les spectres d'intérêt parmi des centaines de spectres. Cette méthodologie ne permet cependant pas d'avoir les informations sur la nature des chaînes glycaniques portées par les protéines. Pour déterminer la micro-hétérogénéité de la glycosylation sur chacun des sites occupés, nous avons donc développé une deuxième approche basée sur l'analyse de glycopeptides.

b. Identification et caractérisation des glycopeptides

Afin de déterminer la micro-hétérogénéité de chaque site, nous avons utilisé une combinaison de trois enzymes protéolytiques (trypsine, chymotrypsine et la GluC) pour générer des glycopeptides avec des séquences peptidiques de tailles différentes. Les digestes après enrichissement en glycopeptides sur phase ZIC-HILIC ont été ensuite analysés par nanoLC-MS/MS.

Pour déterminer les zones d'éluion des glycopeptides, dans les chromatogrammes, une première analyse nanoLC-MS a été réalisée en "scan alterné". Ce mode d'acquisition de MS, consiste à faire varier l'énergie dans la cellule de collision du Q-TOF entre des valeurs de basse énergie (10 eV) et de haute énergie (28 eV). L'utilisation d'une énergie élevée permet de générer des ions dits diagnostiques (Figure 10) qui sont des ions oxoniums spécifiques aux glycopeptides avec une intensité relativement importante afin d'augmenter la détection²³.

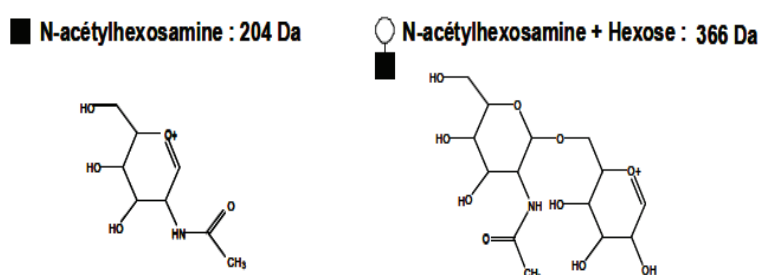


Figure 10. Structure chimique des ions oxoniums servant d'ions diagnostiques pour localiser les glycopeptides de l'invertase vacuolaire dans le chromatogramme.

En effet l'augmentation de l'énergie dans la cellule de collision (haute énergie 28 eV) provoque une fragmentation partielle des extrémités des antennes de la chaîne glycanique portée par les glycopeptides, ce qui donne des ions fragments oxoniums mono-chargés m/z 204 (N-acétylhexosamide) et 366 (N-acétylhexosamide + hexose) qui sont caractéristiques (Figure 10) des glycopeptides. L'extraction du courant d'ions MS des ions diagnostiques (204, 366) obtenus à haute énergie permettra donc ultérieurement de localiser les glycopeptides dans l'analyse nanoLC-MS/MS. Cette méthode a été développée spécifiquement pour l'analyse des glycopeptides lors d'une précédente thèse effectuée au laboratoire (Hovasse 2010). L'intérêt de cette méthode est la génération d'ions diagnostiques suffisamment intenses pour permettre de localiser dans les chromatogrammes, les temps d'éluion de glycopeptides afin de faciliter l'identification et l'annotation des spectres MS/MS de l'analyse suivante (Figure 11). Une deuxième analyse nanoLC-MS/MS a été réalisée ensuite pour fragmenter les glycopeptides enrichis sur phase ZIC-HILIC. La superposition du chromatogramme nanoLC-MS en "scan alterné" et nanoLC-MS/MS a permis de localiser dans l'analyse nanoLC-MS/MS les temps d'éluion des glycopeptides ($tr = 12,9$ min, 13,4 min ; 15,8 min ; 16,9 min ; Figure 11).

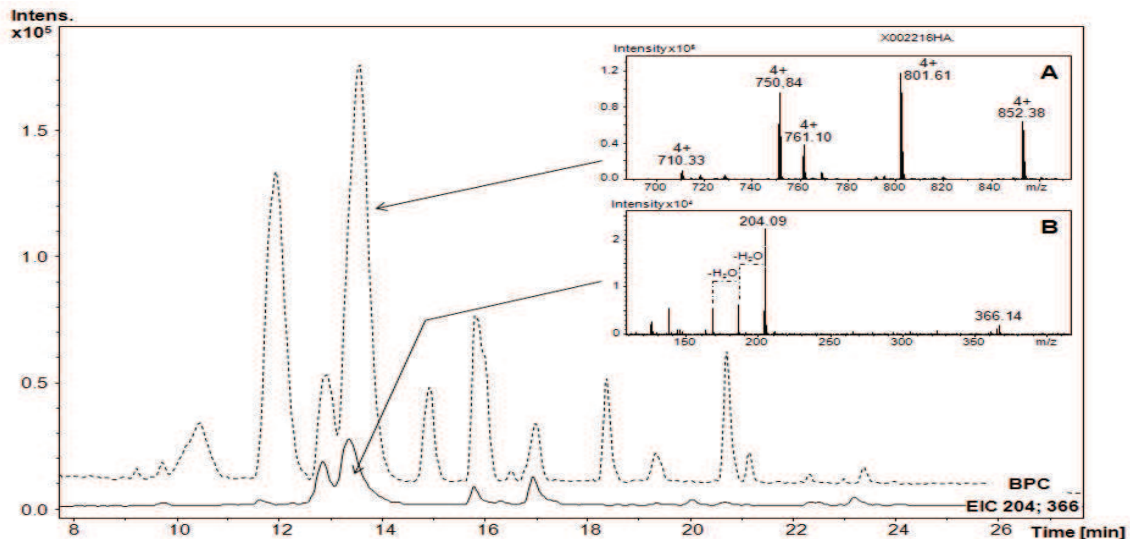


Figure 11. Chromatogramme (--BPC (Base Peak Chromatogram), nanoLC-MS scan alterné) du digeste enrichi en glycopeptides de l'invertase vacuolaire de raisin. (A) spectre MS base énergie (10 eV) des glycopeptides élués à 13,5 min, et en (B) spectre MS haute énergie (28 eV) avec les ions fragments mono chargés (m/z 204 et m/z 366) caractéristiques de glycopeptides. Les courants d'ions 204 et 366 (– ECI) des MS haute énergie ont permis de localiser dans le chromatogramme les pics des glycopeptides (t_r = 12,9 min ; 13,4 min ; 15,8 min ; 16,9 min).

Résultats

La stratégie mise en place a permis d'identifier les spectres de fragmentation MS/MS correspondant aux glycopeptides qui ont été ensuite annotés manuellement (Figures 12 et 13). Par exemple le glycopeptide de m/z 1019,4 (2+) correspondant au peptide tryptique $^{425}\text{TVLHDN}^{430}\text{K}^{431}$ avec $\text{FucXylMan}_2\text{GlcNAc}_3$ sur l'Asn 430 (Figure 12) a été sélectionné et fragmenter en mode CID (Collision Induced Dissociation). Les spectres MS/MS CID des glycopeptides sont caractérisés dans les bas m/z par la présence des ions oxoniums (m/z 204 et 366,1) spécifiques de la glycosylation (Figures 12 et 13). Dans les hauts m/z , la totalité des ions fragments observés correspond à la fragmentation successive de la chaîne d'oligosaccharide portée par le glycopeptide. La fragmentation conduit à une perte successive de résidus correspondant à chacun des glycanes composant l'oligosaccharide du glycopeptide (Xyl 132 Da, Fuc 146 Da, Man 162 Da, GlcNAc 203 Da) jusqu'au peptide + GlcNAc (m/z 1029,8 (1+)) suffisamment intense ou le peptide nu (m/z 826,7 (1+)) peu intense (Figure 12). Le peptide nu ne se fragmente pas en CID. L'identification des deux ions, peptide + GlcNAc et/ou peptide nu qui sont parfois sous plusieurs états de charge permet d'identifier le peptide et donc l'Asn portant la glycosylation par comparaison de ces m/z avec des centaines de m/z théoriques (peptides + GlcNAc et/ou peptides nus) obtenus par digestion tryptique *in silico* de l'invertase vacuolaire de raisin.

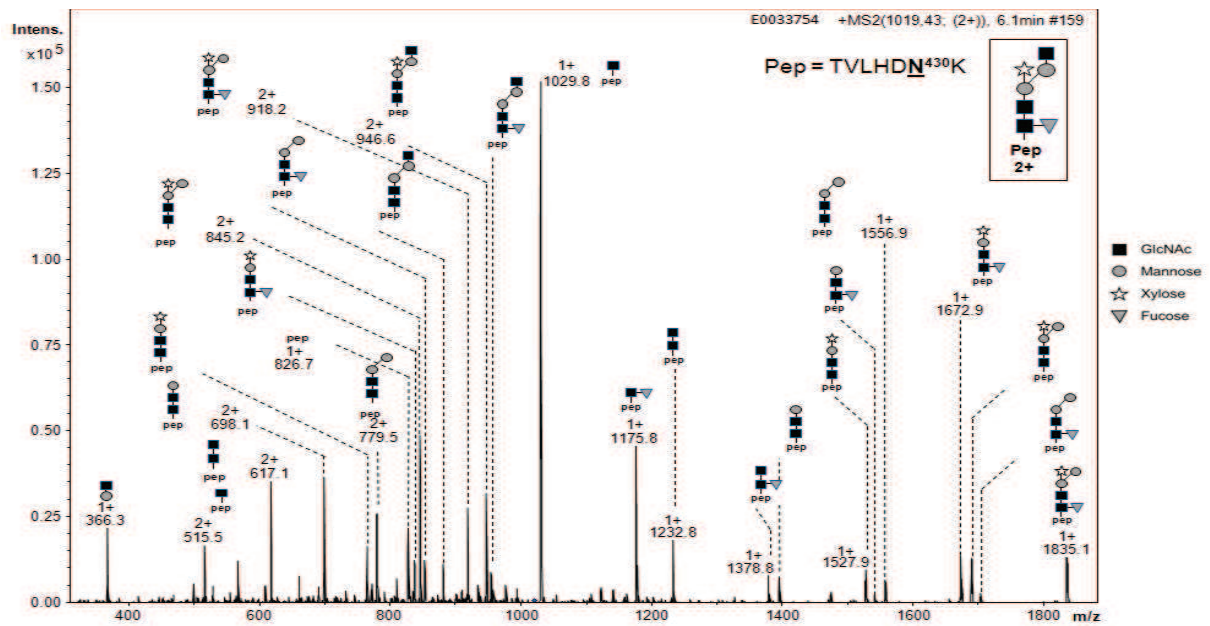


Figure 12. Spectre MS/MS CID (obtenu sur une trappe ionique) annoté d'un glycopeptide tryptique de $m/z = 1019,43$ (2) du site Asn 430 avec FucXylMan₂GlcNAc₃.

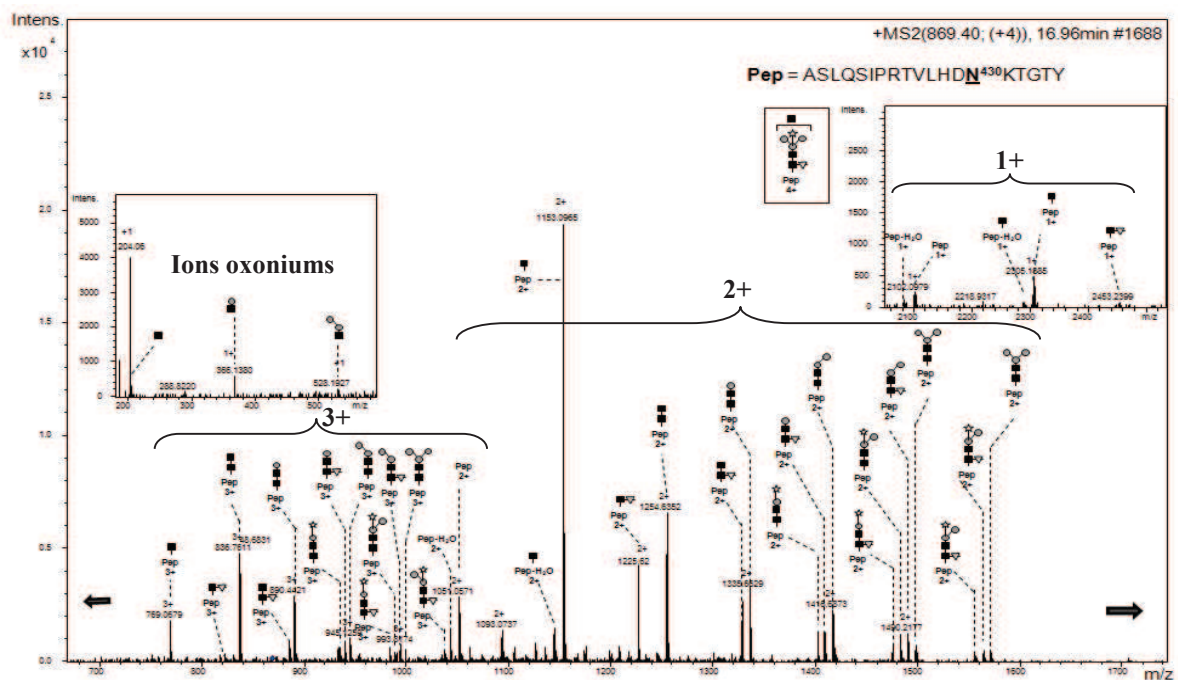


Tableau 2. Résumé des 6 sites de glycosylation identifiés par analyse des glycopeptides. A chaque site de glycosylation est associé le nombre de glycoformes, les glycoformes, le peptide identifié, l'enzyme utilisée pour la digestion et le rapport m/z de l'ion parent correspondant au glycopeptide qui a été sélectionné pour la fragmentation MS/MS.

Site de glycosylation	Glycoformes	Peptides associés aux glycoformes	Enzyme	ion parent (m/z) - état de charge	Nombre de glycoformes identifiées
158	FucXylMan ₃ GlcNac ₂	⁴⁷ YQYNPDSAVWGN*ITW ¹⁶	Chymotrypsine	995.1 (3+)	3
	FucXylMan ₃ GlcNac ₃	⁴⁷ YQYNPDSAVWGN*ITW ¹⁶	Chymotrypsine	1062.8 (3+)	
	FucXylMan ₃ GlcNac ₄	⁴⁷ YQYNPDSAVWGN*ITW ¹⁶	Chymotrypsine	1130.5 (3+)	
212	FucXylMan ₃ GlcNac ₂	²⁰⁸ TGDTN*DSVQVQNL ²²⁰	Chymotrypsine	1280,48 (2+)	1
277	FucXylMan ₃ GlcNac ₂	²⁶⁹ RVAVGSLVNTTGIVL ²⁸³	Chymotrypsine	958,1 (3+)	4
	FucXylMan ₃ GlcNac ₃			1025,8 (3+)	
	FucXylMan ₂ GlcNac ₂			904,1 (3+)	
	FucMan ₃ GlcNac ₂			914,1 (3+)	
430	FucXylMan ₃ GlcNac ₂	⁴²⁵ TVLHDN*K ⁴³¹	Trypsine	998.9 (2+)	11
		⁷ ALSQSIPRTLVDHN*KTGT ⁴	Chymotrypsine	818,6 (4+)	
		⁴²⁰ QSIPRTLVDHN*KTGT ⁴³⁵	Chymotrypsine	750.8 (4+)	
	FucXylMan ₃ GlcNac ₃	⁴²⁵ TVLHDN*K ⁴³¹	Trypsine	1100.9 (2+)	
		⁷ ALSQSIPRTLVDHN*KTGT ⁴	Chymotrypsine	869.4 (4+)	
		⁴²⁰ QSIPRTLVDHN*KTGT ⁴³⁵	Chymotrypsine	801.6 (4+)	
	FucXylMan ₃ GlcNac ₄	²⁰ QSIPRTLVDHN*KTGT ⁴³	Chymotrypsine	829.9 (4+)	
		⁴²⁵ TVLHDN*K ⁴³¹	Trypsine	1202.0 (2+)	
		⁷ ALSQSIPRTLVDHN*KTGT ⁴	Chymotrypsine	920.2 (4+)	
	XylMan ₃ GlcNac ₄	⁴²⁰ QSIPRTLVDHN*KTGT ⁴³⁵	Chymotrypsine	852.4 (4+)	
		⁴²⁵ TVLHDN*K ⁴³¹	Trypsine	1128.9 (2+)	
		⁴²⁵ TVLHDN*K ⁴³¹	Trypsine	1027.4 (2+)	
		⁴²⁵ TVLHDN*K ⁴³¹	Trypsine	1019.4 (2+)	
		⁴²⁵ TVLHDN*K ⁴³¹	Trypsine	925.8 (2+)	
⁴²⁵ TVLHDN*K ⁴³¹		Trypsine	932.9 (2+)		
⁴²⁵ TVLHDN*K ⁴³¹		Trypsine	9179 (2+)		
⁴²⁵ TVLHDN*K ⁴³¹		Trypsine	844.8 (2+)		
FucMan ₂ GlcNac ₂	⁴²⁵ TVLHDN*K ⁴³¹	Trypsine	851.9 (2+)		
450	FucXylMan ₃ GlcNac ₂	⁴⁴⁶ SLRTN*STEFE ⁴⁵⁵	GluC	1177.5 (+2)	3
		⁴⁴⁴ VESLRTN*STEFE ⁴⁵⁵	GluC	1291.5 (2+)	
	FucXylMan ₃ GlcNac ₃	⁴⁴⁶ SLRTN*STEFE ⁴⁵⁵	GluC	1279.0 (2+)	
		⁴⁴⁴ VESLRTN*STEFE ⁴⁵⁵	GluC	1393.0 (2+)	
FucXylMan ₃ GlcNac ₄	⁴⁴⁶ SLRTN*STEFE ⁴⁵⁵	GluC	1380,5 (+2)		
485	FucXylMan ₃ GlcNac ₂	⁴⁸¹ FEVDN*E ⁴⁸⁶	GluC	961.9 (2+)	2
		⁴⁸¹ FEVDN*ETLE ⁴⁸⁹	GluC	1133.4 (2+)	
	FucXylMan ₃ GlcNac ₃	⁴⁸¹ FEVDN*E ⁴⁸⁶	GluC	1063.4 (2+)	

Conclusion

Grâce à la stratégie basée sur l'analyse de glycopeptides, 6 sites de glycosylation (Asn : 158, 212, 277, 430, 450, 485) ont pu être identifiés et leurs glycoformes ont été identifiées et caractérisées.

Au total, nous avons pu identifier 11 oligosaccharides. Le site de glycosylation Asn 430 est particulièrement hyper glycosylé puisque les 11 oligosaccharides ont été identifiés sur ce site.

La comparaison des résultats du protocole B (analyse des glycopeptides) avec les résultats d'analyse du protocole A (analyse de peptides déglycosylés), nous a permis de confirmer 3 sites de glycosylation (Asn : 277 ; 430 ; 450).

En outre l'identification des glycopeptides portant les sites de glycosylation (Asn : 158 ; 212 ; 285), qui antérieurement n'avaient pas été identifiés par la stratégie d'analyse des peptides déglycosylés montre le caractère partiel de la glycosylation de ces sites en particulier.

De plus, nous avons mis en évidence 3 chaînes d'oligosaccharides supplémentaires (FucXylMan₃GlcNac₂, XylMan₃GlcNac₃ et XylMan₃GlcNac₄) par rapport à l'étude des chaînes glycaniques libres (étude glycomique) de l'invertase ⁴.

L'analyse de glycopeptides intacts a donc permis de déterminer l'hétérogénéité de 6 sites de glycosylation et de donner des informations sur les oligosaccharides de ces glycopeptides.

Cependant, l'interprétation des spectres MS/MS a été fastidieuse, puisqu'aucun outil bioinformatique d'interprétation de spectres MS/MS des glycopeptides n'a été utilisé. En réalité, les outils d'identification des glycopeptides actuels fonctionnent sur la base d'une bonne connaissance des glycoformes existant sur la protéine étudiée ce qui n'était pas notre cas.

Conclusion et perspectives

Afin de caractériser les N-glycosylations de l'invertase, nous avons développé une méthodologie basée sur la spectrométrie de masse. Ainsi grâce à cette méthodologie basée sur deux approches d'analyse nanoLC-MS et nanoLC-MS/MS de peptides déglycosylés et glycopeptides intacts, nous avons pu caractériser la micro- et macro-hétérogénéité de l'invertase qui comporte 12 sites consensus de N-glycosylation.

La combinaison des 2 approches, protocole A (l'analyse de peptides déglycosylés) et protocole B (analyse des glycopeptides intacts), nous a permis de montrer le caractère partiel de la glycosylation de l'invertase et de déterminer les différents oligosaccharides présents sur l'invertase vacuolaire de raisin. Grâce à l'analyse des glycopeptides (protocole B), nous avons déterminé 11 oligosaccharides présents sur l'invertase.

En résumé la Figure 14 présente l'ensemble des résultats obtenus avec la combinaison des deux approches. En vert (■) les 6 sites de glycosylation ont été identifiés exclusivement par l'approche protocole A ; en vert orangé (■), les 3 sites de glycosylation qui ont été identifiés par les deux approches protocole A et protocole B. En orange (■), les 2 sites de glycosylation qui ont été identifiés exclusivement par l'approche protocole B. En revanche en orange claire (■) le site partiellement glycosylés a été identifié sans déamidation par l'approche protocole A, alors que par l'approche B, le site a été identifié avec la glycosylation. Au total 12 des 12 sites potentiels de glycosylation de l'invertase vacuolaire de raisin ont été identifiés comme étant glycosylés dont l'un partiellement glycosylé. Ces résultats suggèrent le caractère partiel de la glycosylation de l'invertase vacuolaire, et le fort taux de glycosylation (N-glycosylation) de la protéine.

```

1 MDTHHATSHD LLHDEAYAPL PYSPLPDHPV PAGAPALGQR RPLKVFVAVTL ASLMFLASLV ALVIHQSPQT PVTVDNDGPS
81 MEREGRGVAQ GVSEKSEFLGF SGRRLSYNWT TAMFAWQRTA FHFQPEKNWM NDPDGPLFHM GWYHLFYQYN PDSAVWGNIT
161 WGHAVSRDMI HWLYLPLAMV PDRWFDLNGV WTGSATILPN GQIIMLYTGD TNSVQVQNL AYPANLSDPL LLHWIKYENN
241 PVMVPPAGIG SDDFRDPTM WVGADGNWRV AVGSLVNTTG IVLVFQTTNF TDFELLDGEL HGVPGTGMWE CVDFYPVSIN
321 GVGGLDTSAH GPGIKHVLKA SMDDNRHDYY ALGEYDPMTD TWPDDPELD VGIGLRLDYG KKYASKTFYD QVKKRRILYG
401 WISEGDIESD DLKKGWASLQ SIPRTVLHDN KTGTYLLLWP IEEVESLRTN STEFEDVLE PGSVVPLDIG SASQLDIVAE
481 FEVDNETLEA MVEADVIYNC SFSAGAARG ALGPFGLVLV ADDTLSELTP IYFYIAKDTD GSYKTFEFTD LSRSSSLAVDD
561 VDQRIYGSIV PVLDDKEPTM RVLVDHSIVE GFSQGGRSCT TTRVYPTAEI YGAARLFLFN NATGVNVTAS IKIWEMASAD
641 IHPYPLDQP

```

Figure 14. Séquence de l'invertase vacuolaire de raisin et des sites identifiés. Sites identifiés avec la stratégie basée sur l'analyse: (■) des glycopeptides, (■) des peptides déglycosylés. (■) Sites identifiés avec les deux stratégies. (■) Site partiellement glycosylé.

A partir de nos résultats d'analyse par spectrométrie de masse, des études biochimiques, physico-chimiques et structurales ciblant spécifiquement certains sites de glycosylation identifiés, sont en cours de réalisation chez les collaborateurs afin de mieux comprendre l'impact de la glycosylation de l'invertase vacuolaire de raisin sur les propriétés biochimique et structurale de la protéine, et notamment sur les propriétés moussantes des vin effervescents en particulier le Champagne.

Si l'interprétation des spectres MS/MS de peptides déglycosylés à la PNGase A a été rapide a cause de l'assistance de l'outil bioinformatique qu'est l'algorithme de recherche Mascot, l'interprétation des résultats de l'analyse des glycopeptides a été quant à elle fastidieuse. Il n'y a pas à ce jour d'algorithme de recherche pour l'interprétation des données de spectres MS/MS de glycopeptides assez performant. Les difficultés bioinformatiques résident dans l'écriture de l'algorithme capable de prendre en compte les différentes enzymes de digestion, la combinaison de centaines de fragments de peptides (ions a, x, b, y, c, z)²⁴ et toutes les combinaisons de centaines de fragments des oligosaccharides (A, X, B, Y, C, Z)²⁵ portés par les glycopeptides. Certains des logiciels actuellement utilisés pour l'identification de glycopeptides nécessitent la connaissance préalable des oligosaccharides présents sur les glycoprotéines étudiées²⁶, or ce n'est pas toujours le cas comme l'exemple de notre étude. C'est pourquoi, même avec l'avantage de fournir des informations à la fois sur les sites de glycosylation (macro-hétérogénéité) et des informations sur les oligosaccharides portés par les sites (micro-hétérogénéité), l'approche par analyse des glycopeptides est très peu utilisée.

Le développement des algorithmes beaucoup plus performants permettant d'interpréter automatiquement des données massives de spectres MS/MS de glycopeptides seront donc nécessaires, et devraient permettre dans le futur une automatisation de l'interprétation de ces données.

Les résultats obtenus au cours de notre étude constituent de nouvelles informations de caractérisation biochimique et structurale de l'invertase vacuolaire de raisin. L'hyper-glycosylation de l'invertase vacuolaire du raisin pourrait avoir un rôle dans sa stabilité, son hydrophobie, sa solubilité, sa conformation structurale. Ces données pourraient expliquer par exemple la résistance de cette enzyme aux protéases notamment au cours de la vinification, son implication éventuelle dans les troubles ou "casses" protéiques des vins, et son rôle dans les propriétés moussantes de vins effervescents donc le Champagne.

Références bibliographiques du chapitre 4

1. Hovasse, A. Protéomique: maîtrise de l'instrumentation, application et étude de glycosylations. *Thèse de l'Université de Strasbourg*, Strasbourg, 2010.
2. Nguyen-Quoc, B.; Foyer, C. H. A role for 'futile cycles' involving invertase and sucrose synthase in sucrose metabolism of tomato fruit. *Journal of experimental botany* **2001**, 52, (358), 881-9.
3. Alonso, A. P.; Raymond, P.; Hernould, M.; Rondeau-Mouro, C.; de Graaf, A.; Chourey, P.; Lahaye, M.; Shachar-Hill, Y.; Rolin, D.; Dieuaide-Noubhani, M. A metabolic flux analysis to study the role of sucrose synthase in the regulation of the carbon partitioning in central metabolism in maize root tips. *Metabolic engineering* **2007**, 9, (5-6), 419-32.
4. Dambrouck, T. Contribution à l'étude des priéines des moûts et des vins de Champagne: Détermination de leur origine; Etude d'une glycoprotéine majeure, l'invertase du raisin. Thèse de l'Université de Reims, Reims, 2004.
5. Sturm, A. Molecular characterization and functional analysis of sucrose-cleaving enzymes in carrot (*Daucus carota* L.). *Journal of experimental botany* **1996**, 47 Spec No, 1187-92.
6. Koch, K. E. Carbohydrate-Modulated Gene Expression in Plants. *Annual review of plant physiology and plant molecular biology* **1996**, 47, 509-540.
7. Li, Z.; Palmer, W. M.; Martin, A. P.; Wang, R.; Rainsford, F.; Jin, Y.; Patrick, J. W.; Yang, Y.; Ruan, Y. L. High invertase activity in tomato reproductive organs correlates with enhanced sucrose import into, and heat tolerance of, young fruit. *Journal of experimental botany* **2012**, 63, (3), 1155-66.
8. Proels, R. K.; Roitsch, T. Extracellular invertase LIN6 of tomato: a pivotal enzyme for integration of metabolic, hormonal, and stress signals is regulated by a diurnal rhythm. *Journal of experimental botany* **2009**, 60, (6), 1555-67.
9. 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), 275-83.
10. Cilindre, C.; Jegou, S.; Hovasse, A.; Schaeffer, C.; Castro, A. J.; Clement, C.; Van Dorselaer, A.; Jeandet, P.; Marchal, R. Proteomic approach to identify champagne wine proteins as modified by *Botrytis cinerea* infection. *Journal of proteome research* **2008**, 7, (3), 1199-208.
11. 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. *Journal of agricultural and food chemistry* **2005**, 53, (22), 8782-9.
12. Palmisano, G.; Antonacci, D.; Larsen, M. R. Glycoproteomic profile in wine: a 'sweet' molecular renaissance. *J Proteome Res* **2010**, 9, (12), 6148-59.
13. Ozohanics, O.; Turiak, L.; Drahos, L.; Vekey, K. Comparison of glycopeptide/glycoprotein enrichment techniques. *Rapid communications in mass spectrometry* **2012**, 26, (2), 215-7.
14. Neue, K.; Mormann, M.; Peter-Katalinic, J.; Pohlentz, G. Elucidation of glycoprotein structures by unspecific proteolysis and direct nanoESI mass spectrometric analysis of ZIC-HILIC-enriched glycopeptides. *Journal of proteome research* **2011**, 10, (5), 2248-60.
15. Takegawa, Y.; Ito, H.; Keira, T.; Deguchi, K.; Nakagawa, H.; Nishimura, S. Profiling of N- and O-glycopeptides of erythropoietin by capillary zwitterionic type of hydrophilic interaction chromatography/electrospray ionization mass spectrometry. *Journal of separation science* **2008**, 31, (9), 1585-93.
16. Alvarez-Manilla, G.; Warren, N. L.; Atwood, J., 3rd; Orlando, R.; Dalton, S.; Pierce, M. Glycoproteomic analysis of embryonic stem cells: identification of potential glyco-biomarkers using lectin affinity chromatography of glycopeptides. *Journal of proteome research* **2010**, 9, (5), 2062-75.

17. Lee, A.; Nakano, M.; Hincapie, M.; Kolarich, D.; Baker, M. S.; Hancock, W. S.; Packer, N. H. The lectin riddle: glycoproteins fractionated from complex mixtures have similar glycomic profiles. *Omics : a journal of integrative biology* **2010**, 14, (4), 487-99.
18. Lee, A.; Kolarich, D.; Haynes, P. A.; Jensen, P. H.; Baker, M. S.; Packer, N. H. Rat liver membrane glycoproteome: enrichment by phase partitioning and glycoprotein capture. *Journal of proteome research* **2009**, 8, (2), 770-81.
19. Jegou, S.; Conreux, A.; Villaume, S.; Hovasse, A.; Schaeffer, C.; Cilindre, C. Van Dorsselaer, A.; Jeandet, P., One step purification of the grape vacuolar invertase. *Anal Chim Acta* **2009**, 638, (1), 75-8.
20. Palmisano, G.; Melo-Braga, M. N.; Engholm-Keller, K.; Parker, B. L.; Larsen, M. R. Chemical deamidation: a common pitfall in large-scale N-linked glycoproteomic mass spectrometry-based analyses. *Journal of proteome research* **2012**, 11, (3), 1949-57.
21. Robinson, N. E. Robinson, A. B., Deamidation of human proteins. *Proceedings of the National Academy of Sciences of the United States of America* **2001**, 98, (22), 12409-13.
22. Kaji, H.; Saito, H.; Yamauchi, Y.; Shinkawa, T.; Taoka, M.; Hirabayashi, J.; Kasai, K.; Takahashi, N.; Isobe, T. Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify N-linked glycoproteins. *Nature biotechnology* **2003**, 21, (6), 667-72.
23. Huddleston, M. J.; Bean, M. F.; Carr, S. A. Collisional fragmentation of glycopeptides by electrospray ionization LC/MS and LC/MS/MS: methods for selective detection of glycopeptides in protein digests. *Analytical chemistry* **1993**, 65, (7), 877-84.
24. Johnson, R. S.; Martin, S. A.; Biemann, K.; Stults, J. T.; Watson, J. T. Novel fragmentation process of peptides by collision-induced decomposition in a tandem mass spectrometer: differentiation of leucine and isoleucine. *Analytical chemistry* **1987**, 59, (21), 2621-5.
25. Domon, B.; Costello, C. E. Structure elucidation of glycosphingolipids and gangliosides using high-performance tandem mass spectrometry. *Biochemistry* **1988**, 27, (5), 1534-43.
26. Deshpande, N.; Jensen, P. H.; Packer, N. H.; Kolarich, D. GlycoSpectrumScan: fishing glycopeptides from MS spectra of protease digests of human colostrum sIgA. *J Proteome Res* **2010**, 9, (2), 1063-75.

CONCLUSION GÉNÉRALE

CONCLUSION GÉNÉRALE

L'objectif de ces travaux de thèse était de mettre au point des méthodologies et des stratégies d'analyse protéomique pour identifier et caractériser des protéines issues d'échantillons complexes d'organismes très variés de génomes séquencés ou non: végétaux, homme, parasites et champignons et ceci, même dans les cas où les échantillons n'étaient disponibles qu'en très faible quantité et ne pouvaient être fournis qu'une seule fois (préparation d'échantillons particulièrement longue, rareté des cas pathologiques étudiés,...). Au-delà de la caractérisation maximale des séquences en acides aminés des protéines, nous avons comme objectif de caractériser aussi finement que possible des PTMs qui affectaient certaines protéines identifiées, en particulier la N-glycosylation.

Dans l'**Introduction bibliographique**, nous avons passé en revue les outils, les techniques et les principales méthodes couramment utilisés pour identifier, voire caractériser, les protéines. Ceci nous a permis de mettre en lumière certaines limites de l'analyse protéomique pour l'étude des organismes non ou mal séquencés, ce qui a mis en évidence la nécessité de développer les approches par séquençage *de novo*. Cette partie bibliographique a également montré la nécessité de développer des approches pour mieux caractériser les N-glycosylations et leurs hétérogénéités, notamment par l'étude des glycopeptides qui permet d'accéder directement à la description des micro-hétérogénéités.

Dans la partie **Résultats**, nous avons présenté quatre chapitres correspondant aux quatre contextes d'études différentes qui ont été abordés au cours de ces travaux de thèse pour lesquels nous avons mis en place des méthodologies analytiques basées sur les techniques de LC, gels 1D et 2D et de la spectrométrie de masse. Différentes stratégies d'interprétation des données LC-MS/MS ont été mises en place selon que les génomes des organismes étudiés soient séquencés ou non.

Le **chapitre 1** a porté sur la biologie du parasite *T. gondii* responsable de la toxoplasmose, avec comme objectif l'identification des complexes protéiques impliqués dans trois problématiques : la motilité du parasite (partenaires de TgGAP50), l'interconversion de la forme dormante en forme virulente du parasite (partenaires de TgAlba), et le trafic intracellulaire du parasite (partenaires de TgSORTLR). Pour les trois complexes concernés, des stratégies de purification de complexes protéiques par immuno-affinité, séparation de protéines sur gels 1D et LC-MS/MS ont été utilisées pour parvenir à l'identification des différents partenaires.

Pour TgGAP50 (Glideosome-Associated Protein 50), nous avons pu identifier en plus de TgGAP 50 d'autres constituants majeurs connus du glidéosome (l'actine, la MyoA, la TgMLC1, TgGAP 45, etc.), des enzymes de la glycolyse, des protéines intervenant dans le trafic intracellulaire et des protéines hypothétiques possédant des segments transmembranaires putatifs. De plus, nous avons montré que, selon que TgGAP50 était glycosylé ou non, certains partenaires étaient présents ou non. La glycosylation de TgGAP50 est donc essentielle pour le maintien du complexe. Ces résultats mettent en évidence le rôle clé des glycosylations de TgGAP50 dans le mécanisme d'invasion des cellules hôtes.

Pour les facteurs de transcription TgAlba qui présentent de fortes homologies avec des protéines d'archaebactéries (Acetylation Lowers Binding Affinity), nous avons montré que ces Alba interagissent avec un large complexe de protéines (une centaine de protéines ont pu être identifiées appartenant à plusieurs sous unités). Parmi ces protéines partenaires, plusieurs sont impliquées dans la régulation de la traduction comme, par exemple, les sous-unités du facteur d'initiation de la traduction eiF3 ou encore sont décrites comme liées à des motifs d'ADN comme la protéine TgSsoSB, décrite chez les archéobactéries comme partenaire de l'ADN monobrin. Nous avons ainsi pu établir que le complexe TgAlba joue un rôle important dans la régulation de la virulence de *T. gondii*.

Pour TgSORTLR (Sortilin-Like Receptor), un récepteur membranaire putatif homologue de la sortiline humaine décrite notamment pour jouer un rôle dans la maladie d'Alzheimer et le métabolisme du cholestérol, nous avons identifié les protéines recrutées par cette sortiline. L'analyse protéomique des partenaires des différents domaines de TgSORTLR (domaine luminal N-terminal de la protéine et cytoplasmique C-terminal) ont permis de montrer que les protéines des micronèmes et des rhoptries sont spécifiquement liées au N-terminal de TgSORTLR tandis que l'extrémité cytoplasmique C-terminale de TgSORTLR, qui se fixe spécifiquement aux protéines endosomales (Mu1, sec 23, Vps 9, Vps 26, Vps 35, etc.), joue un rôle capital dans le trafic antérograde et rétrograde du récepteur, dans sa localisation subcellulaire et ses fonctions biologiques.

Le **chapitre 2** a apporté des résultats sur les interactions vigne-pathogènes de la maladie de l'esca, causée par des champignons dont les génomes ne sont pas séquencés. L'analyse différentielle par gel 2D des

tiges herbacées et des troncs de plants de vignes (Chardonnay) sains et infectés a permis d'identifier de protéines de la vigne dont l'expression était modulée (une trentaine de protéines sur les tiges herbacées et une centaine de protéines sur les troncs). Parmi ces protéines présentant des variations d'expression d'après les analyses protéomiques, une dizaine de protéines, apparaissant comme clefs de plusieurs voies métaboliques, ont été étudiées par quantification PCR, ce qui a confirmé en grande partie nos résultats. D'autre part, l'approche de séquençage *de novo* développée a permis l'identification de protéines de champignons (dans les troncs de vignes où siègent les champignons) dont les fonctions biologiques sont en cours d'étude. L'approche méthodologique développée au cours de ce projet a également pu être appliquée avec succès à d'autres variétés de *Vitis vinifera* (Mourvèdre et Gewurztraminer) atteints de la maladie de l'esca. Les protéines de la vigne et des champignons identifiées au cours de ce travail devraient apporter des informations sur les mécanismes d'infection.

Le **chapitre 3** a présenté des résultats portant sur l'étude de variants d'hémoglobine humaine difficilement identifiable par les techniques conventionnellement utilisées en hématologie telles que le séquençage d'ADN. Une stratégie basée sur la mesure de masse des chaînes d'hémoglobine intactes suivie par une analyse des peptides de digestion par LC-MS/MS a permis d'identifier et de caractériser complètement deux variants d'hémoglobine chez deux patients différents. C'est le séquençage *de novo* manuel des peptides non identifiés par approche classique (PFF) qui a permis de déterminer la position et l'acide aminé muté chez les deux variants. La mutation de la chaîne de globine delta décrite dans ce chapitre de thèse est une mutation de la sérine en position 50 en thréonine. La seconde mutation est confidentielle pour l'instant puisqu'elle est associée à une fonctionnalité très particulière de l'hémoglobine.

Le **chapitre 4** a traité de la caractérisation de l'invertase vacuolaire du raisin, une glycoprotéine majeure présentant un fort taux de glycosylation (12 sites particulièrement hétérogènes). La stratégie d'analyse mise en place, basée sur une combinaison de différentes méthodes d'analyse et de digestion, a permis de montrer la présence de glycosylation sur les 12 sites, dont un des sites qui est également présent sous une forme non glycosylée. Nous avons pu décrire au total 11 structures N-glycaniques différentes sur l'ensemble des sites, et pour l'un des sites particulièrement complexe, les 11 formes ont été caractérisées. Les résultats de notre étude ont montré l'hyper-glycosylation de l'invertase qui pourrait avoir un rôle dans sa stabilité, son hydrophobie, sa solubilité et sa conformation structurale.

Ces travaux de thèse montrent combien il reste nécessaire pour les biologistes d'avoir une description toujours plus détaillée des protéines, les modifications post-traductionnelles comprises. Nous avons montré que les méthodes de l'analyse protéomique étaient très efficaces pour identifier les partenaires d'une protéine utilisée pour isoler un complexe. Ces résultats ont pu être obtenus grâce au développement de méthodologies analytiques adaptées à chaque contexte d'étude.

Nous pensons qu'il reste encore une très grande marge de progression du point de vue de la caractérisation complète des protéines en termes de séquence et de modifications post-traductionnelles. Ceci pourra se faire principalement grâce à une série d'améliorations qui sont, à notre avis, les suivantes :

- Une préparation des échantillons réalisée en concertation très étroite avec les biologistes pour fournir des fractions protéiques toujours adaptées à l'étape de LC-MS/MS;
- Des colonnes de chromatographies plus résolutive (colonnes plus longues et gradients plus lents) ;
- L'apparition de la MS/MS de type DIA (Data Independent Acquisition) qui devrait permettre d'obtenir des spectres MS/MS sur tous les peptides contenus dans un échantillon ;
- Une exploitation plus complète de toutes les données LC-MS/MS grâce à de nouveaux outils bioinformatiques.

Notre sentiment est que le potentiel de développement de l'analyse protéomique est encore considérable et que celui-ci demandera encore de nombreuses années.

PARTIE EXPÉRIMENTALE

PARTIE EXPÉRIMENTALE

I. Partie résultats: chapitre 1 et 2

Les expériences et les protocoles analytiques des chapitres 1 et 2 de la Partie II des résultats sont décrits dans les publications respectives des deux chapitres.

II. Partie résultats: chapitre 3

1. Séparation et mesure de masse des chaînes de globines entières

Afin de mesurer la masse des chaînes de globines mutées, les fractions d'hémoglobines anormales ont été analysées par LC-MS, sur un système HPLC Agilent 1100 series (Agilent Technologies, Palo Alto, USA) couplé à un spectromètre de masse micrOTOF-Q (Bruker Daltonics, Brême, Allemagne).

Les chaînes de globines ont été séparées sur une colonne C18, 250 × 2 mm, particules de 5 µm, porosité 200Å (MultoHigh Bio, Langerwehe, Allemagne). La voie A de la pompe HPLC est composée d'eau avec 0,1 % TFA (acide trifluoroacétique) et la voie B composée d'acétonitrile avec 0,1 % TFA. Chaque solution de chaînes de globines est diluée au tiers dans du 0,1 % TFA, et 30 µL (~ 30 µg) de chaque solution sont chargées sur la colonne à un débit de 0,6 ml / min avec 40% de B et éluées avec un gradient linéaire de 40 - 55 % B en 35 min suivi d'un palier de 4 min à 70% de B et d'un rééquilibrage de la colonne à 40 % de B pendant 7 min. Les différents pics détectés ont été collectés pour des analyses LC-MS/MS.

Le spectromètre de masse micrOTOF-Q a été étalonné avec une solution de référence (Tuning Mix ESI-LC, Agilent Technologies) sur une gamme de balayage 50-3000 m/z. Les paramètres généraux du spectromètre de masse sont les suivants : tension du capillaire, -4500 V ; gaz de séchage, 10 L/min ; température de séchage 200 °C. Un total de 3 spectres est moyenné pour obtenir les spectres MS finaux. L'ensemble du système LC-MS est contrôlé par le logiciel HyStar 3.2 (Bruker Daltonics).

Les données brutes collectées pendant les analyses de LC-MS ont été traitées avec le logiciel DataAnalysis 4.0 Build 281 (Bruker Daltonics). Les masses des protéines ont été calculées et comparées aux masses théoriques des différentes chaînes de globine afin de calculer les différences de masse Δm .

2. Analyse nanoLC-MS/MS du digeste tryptique de la chaîne de globine mutée

Digestion des chaînes de globine

Les fractions collectées des globines mutées ont été évaporées pour éliminer l'acétonitrile, et les chaînes de globines ont été reprises dans 20 µL de NH_4HCO_3 à 25 mM (tampon) avant d'être dénaturées par la suite avec 0,5 µL de TFE (Trifluoroéthanol). Puis 100 µL du tampon ont été ajoutés pour diluer l'échantillon avant la digestion avec la trypsine porcine (1 µL d'une solution de trypsine de 0.05 µg/µL) à 37°C à l'obscurité pendant toute une nuit. La digestion a été par la suite arrêtée avec l'ajout de 1 µL solution de TFA à 10%. Les digestes obtenus ont été dilués au dixième pour l'analyse LC-MS/MS.

Analyse LC-MS/MS du digeste

Les analyses nanoLC-MS/MS ont été réalisées sur un système nanoLC-Chip/MS (Agilent Technologies, Palo Alto, USA) couplé à une trappe ionique amaZon (Bruker Daltonics, Brême, Allemagne).

La puce microfluidique contenant une colonne Zorbax 300SB-C18 (150 mm de longueur × 75 µm de diamètre interne, particules 5 µm) et une pré-colonne d'enrichissement Zorbax 300SB-C18 (40 nL) a été utilisée. Le système de solvants est composé de : solvant (A) (acétonitrile 2% et 0,1% d'acide formique dans de l'eau) et solvant (B) (0,1% d'acide formique dans l'acétonitrile et 2% d'eau). Trois µl du mélange peptidique de chaque échantillon ont été injectés sur la précolonne d'enrichissement, à un débit fixé à 3,75 µL/min avec le solvant (A). L'éluion des peptides a été faite à un débit de 300 nL/min avec un gradient linéaire de 8-40%

solvant (B) en 30 min suivi d'un palier de 3 min à 70% de (B) avant le reconditionnement de la colonne à 8% de (B).

La trappe ionique a été étalonnée avec des composés de référence Tuning Mix ESI (Agilent Technologies). Les paramètres suivants ont été utilisés: la température de la source a été fixée à 135 °C pour un débit de gaz de séchage de 3 L/min. La tension capillaire a été fixée à -1850 V. Les spectres ont été acquis en mode positif sur une gamme de masse de 250-1500 m/z en MS en utilisant le mode de résolution standard amélioré («standard enhanced resolution mode», 8100 m/z/s). Le contrôle du nombre de charge d'ions dans la trappe « Ion Charge Control, ICC » a été fixé à 200000 avec un temps d'accumulation maximale de 200 ms, et le nombre de spectres moyennés pour les spectres finaux fixé à 2. Lors des MS/MS, le système a opéré de façon automatique en basculant du mode MS au mode MS/MS. Les 6 ions de peptides les plus abondants au cours de chaque MS, ont été sélectionnés, isolés et fragmentés en mode CID avec une préférence pour les ions doublement chargés. Pour la fragmentation, l'argon a été utilisé comme gaz de collision. 2 spectres de MS/MS ont été acquis pour chaque ion sélectionné et ces derniers d'exclus de sélection pendant une durée de 0.12 min. Les spectres MS/MS ont été acquis sur la gamme de masse 100-2000 m/z. Le nombre de charge d'ion dans la trappe a été fixé à 300000 et 2 scans ont été moyennés pour obtenir chaque spectre MS/MS. Le système au complet (LC et MS) a été entièrement sous contrôle du logiciel Hystar 3.2 (Bruker Daltonics).

Interprétation des données LC-MS/MS

L'ensemble des données MS/MS acquises a été converti en liste de masses de format ".mgf". Pour l'interprétation des données et l'identification des chaînes de globine, la recherche a été effectuée dans la banque Swissprot à l'aide de l'algorithme de recherche Mascot (Matrix Science) et les paramètres de recherche suivants ont été utilisés : l'erreur de masse ions parents et fragments de 0.25 Da; nombre d'omission de clivage tryptique fixé à 1 ; et comme modifications variables, l'oxydation de la méthionine, la carbamidométhylation des cystéines, l'acétylation N-terminale. Les peptides mutés ont été localisés dans les chromatogrammes par extraction des courants d'ions de l'ensemble des peptides tryptiques (T) augmenté du Δm déterminé lors de l'analyse par LC-MS. Après avoir localisé dans le chromatogramme le peptide muté, le spectre MS/MS correspondant a été annoté par *de novo* manuellement pour identifier l'acide aminé muté dans la séquence du peptide.

III. Partie résultats: chapitre 4

1. Digestions enzymatique de l'invertase

L'invertase vacuolaire du raisin a été solubilisée dans de l'eau milliQ et chauffée à 95 °C pendant 5 min. Dans une première expérience, l'invertase vacuolaire a été incubée avec de la PNGase A d'amande douce (5 mU) à 37 °C pendant la nuit puis soumise à des digestions avec de la trypsine et la chymotrypsine. Dans une seconde expérience, la protéine native a été digérée avec la trypsine, la chymotrypsine ou la Glu-C pour obtenir des glycopeptides, puis les N-glycanes ont été libérés par incubation des glycopeptides avec PNGase A comme décrit ci-dessus.

Les digestions protéolytiques ont été effectuées en parallèle avec différentes enzymes sur la protéine native ainsi que sur l'invertase vacuolaire déglycosylée.

L'étape de réduction/alkylation de l'invertase vacuolaire native et déglycosylée a été effectuée comme suit. Pour chaque expérience, 50 pg de protéine (natif ou déglycosylée) ont été dissous dans 50 μ l d'urée 8 M dans 0,1 M de tampon NH_4HCO_3 . La protéine a été réduite avec 2 μ l de 700 mM de DTT (dithiothréitol) pendant 60 min, puis alkylée avec 6 μ l de 700 mM d'iodoacétamide pendant 45 min à 60 °C et enfin 500 μ l de tampon 0,1 M NH_4HCO_3 sont ajoutés pour ramener la concentration de l'urée à 1 M environ avant la digestion protéolytique.

L'invertase vacuolaire déglycosylée a été digérée avec 1 μ g de chymotrypsine (25 °C) ou 0,4 μ g de trypsine (37 °C) pendant une nuit (rapport enzyme/protéine, respectivement 1/20 et 1/50). Les solutions finales ont été acidifiées pour arrêter le processus de digestion avec 2 μ l d'acide formique. Pour éliminer l'urée, deux digestes ont été dessalés sur une cartouche SPE (Sep-Pak C18, Waters) selon la procédure suivante: les digestes ont été chargés sur les cartouches de Sep-Pak conditionnés, puis lavés avec de l'acide formique à 0,1% et l'élution a été réalisée avec 600 μ l d'eau/acétonitrile/acide formique, (50/50/0,1 ; v/v/v) et le volume a été réduit à 150 μ l par évaporation avant injection.

L'invertase vacuolaire a été digérée soit avec de la trypsine, de la chymotrypsine ou de la Glu C (rapport enzyme/protéine de 1/100) et 100 mM de tampon phosphate à pH 8,0 a été utilisée pour la digestion avec la Glu C. Le digeste a ensuite été dessalé et enrichi en glycopeptides sur une phase ZIC-HILIC (Sequant, ??) selon le protocole de Hägglund et al (2004).

2. Analyse nanoLC-MS et MS/MS des peptides et glycopeptides

Les expériences de nanoLC-MS/MS ont été effectuées sur deux systèmes de couplage LC-MS: un système nanoACQUITY Ultra-performance-LC système (Waters, Milford, MA, USA) couplé à un spectromètre de type Q-TOF maXis (Bruker Daltonics) et un système nanoLC-Chip/MS (Agilent Technologies) couplé à une trappe ionique amaZon (Bruker Daltonics). L'ensemble des systèmes est contrôlé par le logiciel Hystar 3.2 (Bruker Daltonics).

Système UPLC / Q-TOF

Les échantillons sont chargés sur une pré-colonne Symmetry C18 (20 x 0,18 mm, particules de 5 µm, Waters). Les peptides sont séparés sur une colonne ACQUITY UPLC® BEH130 C18 (75 µm x 200 mm, particules de 1,7 µm, 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 µL/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 à 45°C suivit d'un saut à 90% de B et un retour aux conditions initiales (1% B).

Le spectromètre de masse Q-TOF utilise les paramètres suivants : la température de la source est réglée à 200°C, le gaz de séchage à un débit de 4 L/min et la tension sur le nanospray est de -4,5 kV. L'étalonnage du TOF est réalisé en utilisant le mélange Tuning Mix ESI (Agilent Technologies) en mode positif sur la gamme de balayage 50-2200 m/z.

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 en mode MS/MS automatique. La gamme de balayage est m/z 50 à 2200. En MS la durée d'accumulation est de 0,2 s. La MS/MS des ions 2, 3 et 4 fois chargés sont préférentiellement 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 0,6 min.

Système HPLC / Trappe ionique

La puce microfluidique contenant une colonne Zorbax 300SB-C18 (150 mm de longueur × 75 µm de diamètre interne, particules 5 µm) et une pré-colonne d'enrichissement Zorbax 300SB-C18 (40 nL) a été utilisée. Le système de solvants est composé de : solvant (A) (acétonitrile 2% et 0,1% d'acide formique dans de l'eau) et solvant (B) (0,1% d'acide formique dans l'acétonitrile et 2% d'eau). Trois µl du mélange peptidique de chaque échantillon ont été injectés sur la précolonne d'enrichissement, à un débit fixé à 3,75 µL/min avec le solvant (A). L'élution des peptides a été faite à un débit de 300 nL/min avec un gradient linéaire de 8-40% solvant (B) en 30 min suivi d'un palier de 3 min à 70% de (B) avant le reconditionnement de la colonne à 8% de (B).

La trappe ionique a été étalonnée avec des composées de référence Tuning Mix ESI (Agilent Technologies). Les paramètres suivants ont été utilisés: la température de la source a été fixée à 135 °C pour un débit de gaz de séchage de 3 L/min. La tension capillaire a été fixée à -1850 V. Les spectres ont été acquis en mode positif sur une gamme de masse de 500-1500 m/z en MS en utilisant le mode de résolution standard amélioré («standard enhanced resolution mode », 8100 m/z/s). Le contrôle du nombre de charge d'ions dans la trappe « Ion Charge Control, ICC » a été fixé à 100000 avec un temps d'accumulation maximale de 200 ms, et le nombre de spectres moyennés pour les spectres finaux fixé à 2. Lors des MS/MS, le système a opéré de façon automatique en basculant du mode MS au mode MS/MS. Les 6 ions de peptides les plus abondants au cours de chaque MS, ont été sélectionnés, isolés et fragmentés en mode CID avec une préférence pour les ions doublement chargés. Pour la fragmentation, l'argon a été utilisé comme gaz de collision. 2 spectres de MS/MS

ont été acquis pour chaque ion sélectionné et ces derniers d'exclus de sélection pendant une durée de 0.12 min. Les spectres MS/MS ont été acquis sur la gamme de masse 150-2000 m/z. Le nombre de charge d'ion dans la trappe a été fixé à 300000 et 2 scans ont été moyennés pour obtenir chaque spectre MS/MS.

Interprétation des données LC-MS/MS

Pour l'interprétation des analyse réalisées sur les digests de la protéine déglycosylée, les données recueillies au cours des analyses nanoLC-MS/MS ont été converties en liste de masses ".mgf" avec le logiciel DataAnalysis (Bruker Daltonics). Les listes de masse ont ensuite été interprétés avec l'algorithme de recherche Mascot (Matrix Science) dans une banque protéique composée de séquences de protéines de *Vitis vinifera* issue de NCBI (1 mai 2010) concaténé avec des copies inversées de toutes les séquences (au total 51 258 entrées). Les recherches ont été effectuées avec l'enzyme indiquée, avec une erreur de masse sur les ions parents et fragments de $\pm 0,25$ Da pour les données de la trappe ionique et de 5 ppm pour les ions parents et 0,02 Da pour les ions fragment pour les données du spectromètre de type Q-TOF. Deux omissions de clivage par peptide tryptique, quatre par peptide chymotryptique et cinq pour la Glu-C ont été autorisés. Quelques modifications variables ont été prises en compte pour la recherche: carbamidométylation des cystéines, acétyle N-terminale de la protéine et l'oxydation (M). Pour les peptides déglycosylés, la modification de l'asparagine (Asn: N) en acide aspartique (Asp: D) a été ajoutée pour identifier le site de glycosylation modifiée par clivage des groupements glycaniques au cours de la digestion par la PNGase A. Pour valider les peptides, le score Mascot minimum requis était de 25.

L'interprétation des données LC-MS et MS/MS des glycopeptides a été ensuite réalisée manuellement et validées par assignation manuelle des ions fragments des glycopeptides comme décrit par Castro et al. (2010). Les spectres MS/MS des glycopeptides sont recherchés au niveau des zones d'élution des ions diagnostiques. Les fragments majoritaires permettent de déterminer la structure du glycane et la masse du peptide non glycosylé sont identifiés. 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).

LISTE DE PUBLICATIONS ET COMMUNICATIONS

Liste des publications

Gissot M, Walker R, Delhaye S, **Alayi T D**, Huot L, Hot D, Callebaut I, Schaeffer-Reiss C, Van Dorsselaer A, S T: Toxoplasma gondii Alba proteins are involved in translational control of gene expression. *J. Molecular Biology* 2013, in press.

Sloves PJ, Delhaye S, Mouveaux T, Werkmeister E, Slomianny C, Hovasse A, **Dilezitoko Alayi T**, Callebaut I, Gaji RY, Schaeffer-Reiss C, et al: Toxoplasma Sortilin-like Receptor Regulates Protein Transport and Is Essential for Apical Secretory Organelle Biogenesis and Host Infection. *Cell Host Microbe* **2012**, 11:515-527.

Spagnolo A, Magnin-Robert M, **Alayi TD**, Cilindre C, Mercier L, Schaeffer-Reiss C, Van Dorsselaer A, Clement C, Fontaine F: Physiological changes in green stems of Vitis vinifera L. cv. Chardonnay in response to esca proper and apoplexy revealed by proteomic and transcriptomic analyses. *J Proteome Res* **2012**, 11:461-475.

Fauquenoy S, Hovasse A, Sloves PJ, Morelle W, **Dilezitoko Alayi T**, Slomianny C, Werkmeister E, Schaeffer C, Van Dorsselaer A, Tomavo S: Unusual N-glycan structures required for trafficking Toxoplasma gondii GAP50 to the inner membrane complex regulate host cell entry through parasite motility. *Mol Cell Proteomics* **2011**, 10:M111 008953.

Communications orales

Alayi T D., Jégou S., Hovasse A., Conreux A., Villaume S., Jeandet P., Van Dorsselaer A. et Schaeffer-Reiss C. Characterization of a highly glycosylated protein by combination of mass spectrometry strategies: the grape vacuolar invertase. *XVIIème rencontres du CJSM (Club Jeune de la Société Française de Spectrométrie de Masse SFSM)*, **avril 2012**, La Bresse (France).

Communications par affiche

Alayi T D., Schaeffer-Reiss C., Hovasse A, Sloves P-J., Delhaye S., Mouveaux T., Werkmeister E., Slomianny C., Y. Gaji R., Carruthers V B., Van Dorsselaer A. and Tomavo S. Proteomic study of a key protein Sortilin-Like Receptor essential for the biogenesis of secretory organelles in Toxoplasma gondii. *9th Siena Meeting From Genome to Proteome: Open Innovations*, **août 2012**, Siena (Italie).

Alayi T D., Schaeffer-Reiss C., Hovasse A., Sloves P-J., Delhaye S., Mouveaux T., Werkmeister E., Slomianny C., Y. Gaji R., Carruthers V B., Van Dorsselaer A. and Tomavo S. Étude protéomique de la Sortiline, un récepteur déterminant pour la toxoplasmose causée par le parasite Toxoplasma gondii. *29èmes Journées Françaises de Spectrométrie de Masse (SFSM)*, **septembre 2012**, Orléans (France).

T D. Alayi, S. Jégou, A. Hovasse, A. Conreux, S. Villaume, P. Jeandet, A. Van Dorsselaer et C. Schaeffer-Reiss. Identification des sites de coupure protéolytique de l'invertase vacuolaire de raisin par la stratégie de marquage N-TOP. *Congrès de la Société Française d'Electrophorèse et d'Analyse Protéomique (SFEAP)*, **octobre 2012**, Rouen (France).

Alayi TD., Carapito C., Spagnolo A., Magnin-Robert M., Cilindre C., Van Dorsselaer A., Clément C., Fontaine F. et Schaeffer-Reiss C. Etude protéomique de la vigne infectée par l'Esca : Identification des protéines fongiques dans les troncs par une approche de novo/MS-BLAST. *Congrès de la Société Française d'Electrophorèse et d'Analyse Protéomique (SFEAP)*, **octobre 2012**, Rouen (France).

RÉSUMÉ

L'analyse protéomique par spectrométrie de masse s'est imposée comme une méthode incontournable pour la caractérisation des protéines. Grâce aux progrès de l'instrumentation et de la bioinformatique, l'interprétation automatisée des spectres MS/MS permet aujourd'hui d'identifier des milliers de protéines dans un type cellulaire. Cependant, cette méthodologie s'applique encore difficilement aux organismes dont les génomes n'ont pas été séquencés, et donc pour lesquels il n'existe pas de banques de séquences peptidiques de référence.

Notre travail a porté sur le développement et l'application d'une méthodologie d'interprétation des données MS/MS pour les organismes à génomes non séquencés. Cette méthodologie est basée sur le séquençage *de novo* suivi de recherche MS-BLAST. Ainsi nous avons pu :

- Identifier les différents partenaires de complexes protéiques tels que les protéines des complexes TgGAP50, TgAlba, TgSORTLR impliqués dans la motilité, la virulence ou le trafic intracellulaire des protéines du parasite *Toxoplasma gondii*,
- Identifier et caractériser des variants d'hémoglobine humaine,
- Identifier les protéines différenciellement exprimées lors des interactions vigne et champignons à génomes non séquencés dans la maladie de l'esca,
- Caractériser finement la N-glycosylation de l'invertase vacuolaire du raisin.

Nous avons pu réaliser nos études sur des échantillons d'origines très différentes : homme, plantes, champignons, parasites et nous avons apporté des éléments de réponses moléculaires aux questions biologiques.

Mots clés: protéomique, spectrométrie de masse, séquençage *de novo*, MS-BLAST, *Toxoplasma gondii*, esca, variants d'hémoglobine, N-glycosylation.

ABSTRACT

The proteomic analysis by mass spectrometry is now an essential method for the characterization of proteins. Thanks to advances in instrumentation and bioinformatics, automated interpretation of MS/MS spectra can now identify thousands of proteins in a cell type. However, this methodology remains poorly applied to the organisms that genomes are not sequenced and therefore where there is no database of reference for peptides sequences.

Our work has focused on the development and application of a methodology for the interpretation of MS/MS data for the organisms that genomes are not sequenced. This methodology is based on the *de novo* sequencing followed by MS-BLAST search. Thus we have:

- Identify different partners of protein complexes such as proteins TgGAP50, TgAlba and TgSORTLR complex, involved in motility, virulence or intracellular protein trafficking of *Toxoplasma gondii*,
- Identify and characterize human hemoglobin variants,
- Identify the proteins differentially expressed during interaction of vines and fungi that genomes are not sequenced in esca disease,
- Finely characterize the N-glycosylation of the grape vacuolar invertase.

We have achieved our studies on samples of very different origins: human, plants, fungi, parasites, and we provided evidence of molecular responses to biological questions.

Keywords: proteomics, mass spectrometry, *de novo* sequencing, MS-BLAST, *Toxoplasma gondii*, esca, hemoglobin variants, N-glycosylation.