



**HAL**  
open science

**Vers une étude approfondie des protéomes :  
caractérisation des extrémités N-terminales des  
protéines**

Daniel Ayoub

► **To cite this version:**

Daniel Ayoub. Vers une étude approfondie des protéomes : caractérisation des extrémités N-terminales des protéines. Sciences agricoles. Université de Strasbourg, 2012. Français. NNT : 2012STRAF035 . tel-00855301

**HAL Id: tel-00855301**

**<https://theses.hal.science/tel-00855301v1>**

Submitted on 29 Aug 2013

**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.

ÉCOLE DOCTORALE DES SCIENCES CHIMIQUES

Institut Pluridisciplinaire Hubert Curien  
UMR 7178

# THÈSE

présentée par

**Daniel AYOUB**

soutenue le : **25 Septembre 2012**

pour obtenir le grade de

**DOCTEUR DE L'UNIVERSITÉ DE STRASBOURG**

Discipline / Spécialité : Chimie Analytique

**VERS UNE ÉTUDE APPROFONDIE DES PROTÉOMES :  
CARACTERISATION DES EXTREMITÉS N-TERMINALES DES PROTÉINES**

**THÈSE dirigée par :**

**M. VAN DORSELAER Alain**

Directeur de recherche, IPHC, Université de Strasbourg

**RAPPORTEURS :**

**Mme TORTAJADA Jeanine**

Professeur, Université D'Evry Val d'Essonne

**M. SCHMITTER Jean-Marie**

Professeur, Université Bordeaux Segalen

---

**MEMBRES DU JURY :**

**M. WAGNER Alain**

Directeur de recherche, Université de Strasbourg

**M. LANZER Michael**

Professeur, Université de Heidelberg



# THÈSE

Présentée pour l'obtention du grade de  
DOCTEUR DE L'UNIVERSITÉ DE STRASBOURG

par

Daniel AYOUB

## **VERS UNE ÉTUDE APPROFONDIE DES PROTÉOMES : CARACTÉRISATION DES ÉXTREMITÉS N-TERMINALES DES PROTÉINES**

Soutenue le 25 Septembre 2012 devant la commission d'examen :

Dr. Alain VAN DORSSELAER

Directeur de thèse

Pr. Jeanine TORTAJADA

Rapporteur

Pr. Jean-Marie SCHMITTER

Rapporteur

Dr. Alain WAGNER

Examineur

Pr. Michael LANZER

Examineur







***A mes parents, à mes sœurs***

***A ma chérie***

***A la mémoire de Ruby***

*« Ne jamais réprimer ni te taire à toi-même une objection que l'on peut faire à ta pensée ! Fais-en le vœu ! Cela fait partie de la loyauté première de la pensée. Tu dois chaque jour mener aussi campagne contre toi-même. Une victoire ou une redoute conquise ne sont plus ton affaire, mais celle de la vérité,... mais ta défaite aussi n'est plus ton affaire ! »*

***Friedrich Nietzsche***



## MERCI !

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

En premier lieu, mes plus sincères remerciements s'adressent naturellement à Alain Van Dorselaer, non seulement pour m'avoir accueilli, soutenu et offert un cadre de travail magnifique mais aussi pour l'amitié qu'il m'a montrée et toutes les choses qu'il m'a appris.

Je tiens également à remercier Christine Schaeffer-Ries pour ses conseils, son encadrement et l'immense aide qu'elle m'a prêtée dans les moments les plus difficiles de la rédaction et de la préparation pour la soutenance. Un grand merci aussi à Fabrice Bertile pour son amitié d'abord, pour le travail que nous avons pu réaliser ensemble sur le jeûne du rat (ah ces pauvres petites bêtes !) et pour toutes les discussions philosophiques que nous avons eues en fumant devant la porte du labo ! Mes remerciements vont aussi à Christine Carapito avec qui c'était un grand plaisir de travailler sur les mitochondries. Merci pour ta bonne humeur, ta disponibilité et ton amitié. Et bien sûr je n'oublie pas de remercier Sarah Sanglier-Cianféran pour m'avoir appris une autre face de la spectrométrie de masse, celle de la SM supramoléculaire, merci pour tous tes conseils et pour ton support.

Je remercie vivement les membres du jury, Jeanine Tortajada, Jean Marie Schmitter et Alain Wagner pour avoir accepté d'évaluer ce travail de thèse.

Un très grand merci à tous ceux avec qui j'ai partagé ces années au LSMBO : Les anciens qui sont partis : Audrey Bednarczyk (ça fait un moment déjà mais maintenant que je suis près de Genève je te croiserai peut être), Sébastien Gallien le chef de notre bureau des pouilleux et mon mentor dans l'analyse des données protéomiques (bonne continuation dans ta carrière luxembourgeoise!), Cédric Atmanène (mon coach de spectrométrie de masse supramoléculaire et grand gagnant du tournoi poker et victime de la FC LSMBO. Bonne continuation chez NovaliX!), Nicolas Barthélémy arrivé presque en même temps que moi, grand spécialiste du cheveu humain et partenaire de voyage et d'excursion dans le wild west Américain, Thierry Wasselin notre chti spécialiste du dépouillage des gels 2D et grand amateur de gâteaux Rem et Prince. Du côté des précieux, Guillaume Béchade le plus précieux des précieux parti au Benelux, Laetitia Fouillen bretonne-picarde devenue bordelaise, quand est-ce que tu m'invites pour déguster quelques grands crus ? Chrystel Husser ma partenaire pour les gels d'électrophorèse pleine de bonne humeur et Magali Rompais qui s'aventure maintenant dans l'enseignement à Lyon. Je remercie également Véronique Trimbou grande aide dans les problèmes administratifs et Jennifer Jund notre qualitiennne préférée qui est partie pour de nouvelles aventures.

Les anciens qui sont encore au LSMBO, le dernier des précieux et membre de l'académie Françoise Jean-Michel Saliou, Jean Marc Strub notre spécialiste de presque tout, même les grues et les gros engins Liebherr et Caterpillar, la surveillante des bonnes pratiques de laboratoire et ma mentor de l'électrophorèse Danièle Thiersé et ses petits schnokeles, François Delalande grand écolo du labo et notre guide pour les randonnées Vosgiennes, Kevin Jeanpert, spécialiste de l'administration et relève footballistique de la FC LSMBO, notre miss quality assurance ISO 9001

Hélène Dimer avec qui je continue de collaborer, Laurence Sabatier et Veronique Delval qui ont enfin rejoint pleinement le LSMBO après l'annexion d'une partie du R2, Patrick Guterl qui s'est converti en bioinformaticien et notre MSDA man Alexandre Burel (merci pour tous les problèmes de banques de données que tu as résolus), les bientôt docteurs Elie Alayi, Stéphanie Petiot (devenue récemment maman, félicitation !) et Amandine Bœuf aussi partenaire dans l'enseignement des TPs, Diego Bertacini la relève dans les projets Plasmodium (bonne chance pour la suite), Sarah Lenon (ça rime avec bretonne et John Lenon), Marine Plumel spécialiste de la DiGE. Merci aussi à mon partenaire OptimAbs François Debaene grand gagnant du prix poster de la SFSM. Je remercie très spécialement Agnès Hovasse ma partenaire cigarette et merveilleuse amie avec qui j'ai eu un énorme plaisir de partager ces années au LSMBO, et bien sûr je n'oublie pas Fabrice Varrier mon informaticien préféré, merci pour ton amitié et ta bonne humeur et tous les moments et rires que nous avons partagés ensemble.

Je souhaite aussi bonne chance à ceux qui ont rejoint récemment le LSMBO et que je n'ai pas eu la chance de bien connaître.

Je remercie également l'ensemble des personnes avec qui j'ai eu un grand plaisir à collaborer : Amos Bairoch, Lydie Lane, T. Rabilloud, M. Lanzer, C. Sanchez, A. Podjarny, A. Cousido-Siah, H. Schaller. Je remercie spécialement Dimitri Heintz pour son amitié, pour tous les projets sur lesquelles nous avons collaborés et pour son grand enthousiasme.

Je remercie mes parents, et mes sœurs qui m'ont soutenu tout au long de ma vie et qui malheureusement n'ont pas pu venir assister à ma soutenance. Merci à Ghassan qui a traversé la Manche pour écouter mon discours de 45 mins en Français. Un grand merci à mon oncle Michel qui m'a toujours soutenu et sans qui je ne serai pas là. Merci à Richard, Bachir, Lena, Patrick et tous mes amis qui se reconnaîtront. Je remercie du fond de mon cœur ma petite chérie qui a été à mes côtés pendant tous ces moments difficiles de rédaction et qui m'a beaucoup soutenu. A tous ceux que j'ai manqués de remercier ici, Merci !

# Plan Général

## Introduction générale

---

INTRODUCTION GÉNÉRALE	3
-----------------------	---

## Partie I Introduction bibliographique

---

CHAPITRE 1 LES OUTILS ET TECHNOLOGIES AU SERVICE DE L'ANALYSE PROTÉOMIQUE	9
CHAPITRE 2 LES STRATÉGIES D'IDENTIFICATION EN ANALYSE PROTÉOMIQUE	33
CHAPITRE 3 LES STRATÉGIES DE QUANTIFICATION EN ANALYSE PROTÉOMIQUE	45
CHAPITRE 4 L'ANALYSE DES SITES DE CLIVAGE PROTÉOLYTIQUES DES PROTÉINES : ÉTAT ACTUEL DE L'ART	63
RÉFÉRENCES BIBLIOGRAPHIQUES	73

## Partie II Développements méthodologiques

---

INTRODUCTION	93
CHAPITRE 1 MISE AU POINT DE LA DÉRIVATION N-TERMINALE DES PROTÉINES DANS LE GEL : LA STRATÉGIE « INGEL N-TOP »	97
CHAPITRE 2 MISE AU POINT DE LA STRATÉGIE qN-TOP : DÉRIVATION AVEC DU TMPP <sup>13</sup> C POUR LA QUANTIFICATION DU PROTÉOME N-TERMINAL	103
CHAPITRE 3 APPORT DU MARQUAGE AVEC UN NOUVEAU TMPP PLUS HYDROPHOBE POUR ENRICHIR D'AVANTAGE EN PEPTIDES N-TERMINAUX	111
RÉFÉRENCES BIBLIOGRAPHIQUES	117

## **Partie III Applications de l'Analyse Protéomique en Biologie**

---

<b>CHAPITRE 1. ETUDE DU PROTÉOME MITOCHONDRIAL HUMAIN : APPLICATION DE LA STRATÉGIE N-TOP POUR LA CARACTÉRISATION DES PEPTIDES DE TRANSIT</b>	<b>121</b>
<b>CHAPITRE 2. ÉTUDE DU PROTÉOME DU PARASITE PLASMODIUM <i>FALCIPARUM</i> ET DE L'EXPORT DES PROTÉINES PARASITAIRES À L'ÉRYTHROCYTE INFECTÉ. APPLICATION DE LA STRATÉGIE N-TOP ET INGEL N-TOP</b>	<b>137</b>
<b>CHAPITRE 3 ANALYSE PROTÉOMIQUE DU TISSU ADIPEUX DU RAT EN RÉPONSE À UNE PRIVATION DE NOURRITURE PLUS OU MOINS PROLONGÉE</b>	<b>167</b>
<b>CHAPITRE 4 APPLICATION DE LA STRATEGIE qN-TOP À L'ÉTUDE DU PROTÉOME DU FOIE DU RAT SOUMIS AU JEÛNE PROLONGÉ</b>	<b>179</b>
<b>CHAPITRE 5 ÉTUDE FONCTIONNELLE DE REP1-NCXSQ, UNE PROTÉINE QUI RÉGULE L'ÉCHANGEUR Na<sup>+</sup>/Ca<sup>2+</sup> DU NERF DU CALAMAR</b>	<b>187</b>
<b>RÉFÉRENCES BIBLIOGRAPHIQUES</b>	<b>197</b>

## **Conclusion générale**

---

<b>CONCLUSION GÉNÉRALE</b>	<b>209</b>
----------------------------	------------

## **Partie expérimentale**

---

<b>PARTIE EXPÉRIMENTALE</b>	<b>215</b>
-----------------------------	------------

## **Travaux supplémentaires**

---

<b>TRAVAUX SUPPLÉMENTAIRES</b>	<b>227</b>
--------------------------------	------------

# Plan détaillé

## Introduction générale

---

### INTRODUCTION GÉNÉRALE

3

## Partie I Introduction bibliographique

---

### CHAPITRE 1 LES OUTILS ET TECHNOLOGIES AU SERVICE DE L'ANALYSE PROTÉOMIQUE

9

1. La spectrométrie de masse pour l'analyse protéomique 10
  - 1.1. *Les sources électrospray* 10
    - 1.1.1. La source d'ionisation électrospray 11
    - 1.1.2. La source MALDI 13
  - 1.2. *Les analyseurs de masse* 14
    - 1.2.1. L'analyseur quadripolaire 15
    - 1.2.2. L'analyseur Trappe Ionique 15
    - 1.2.3. Analyseur à temps de vol 17
    - 1.2.4. Les analyseurs à transformé de Fourier 20
      - 1.2.4.1. Analyseur à Résonance Cyclotronique des Ions à Transformée de Fourier (FT-ICR) 20
      - 1.2.4.2. Analyseur Orbitrap<sup>TM</sup> 21
  - 1.3. *Les détecteurs* 22
    - 1.3.1. Le détecteur multiplicateur électrostatique 22
    - 1.3.2. Le détecteur MCP (MicroChannel Plate) 22
    - 1.3.3. Le détecteur magneTOF 23
  - 1.4. *La fragmentation peptidique* 23
    - 1.4.1. La nomenclature des fragmentations peptidiques 23
    - 1.4.2. Le mode de fragmentation CID 24
      - 1.4.2.1. Mécanisme de fragmentation 24
    - 1.4.3. Le mode de fragmentation par transfert d'électron (ETD) 26
2. Les techniques de séparation des protéines et des peptides 27
  - 2.1. *La séparation au niveau des protéines* 27
    - 2.1.1. L'électrophorèse sur gel monodimensionnelle 1D SDS PAGE (Sodium DodecylSulfate PolyAcrylamide Gel Electrophoresis) 27
    - 2.1.2. L'électrophorèse bidimensionnelle sur gel (gel 2 D) 28
    - 2.1.3. L'électrophorèse « off gel » 29
    - 2.1.4. La séparation des protéines par chromatographie liquide 30
  - 2.2. *La séparation au niveau peptidique* 30

2.2.1.	Les faibles débits et les systèmes nanoHPLC (ou nanoLC) et nanoUHPLC	30
2.2.2.	La chromatographie multidimensionnelle	31
<b>CHAPITRE 2</b>	<b>LES STRATÉGIES D'IDENTIFICATION EN ANALYSE PROTÉOMIQUE</b>	<b>33</b>
1.	L'empreinte peptidique massique PMF	34
2.	Les stratégies par analyses MS/MS	35
2.1.	<i>Identification des protéines en mode d'acquisition données-dépendants (DDA)</i>	35
2.1.1.	Identification par recherche dans les banques de données protéiques	36
2.1.1.1.	Attribution de scores par les moteurs de recherche aux identifications	37
2.1.1.1.1.	Le score de corrélation	38
2.1.1.1.2.	Le score probabiliste	38
2.1.1.2.	La combinaison de plusieurs moteurs de recherche	39
2.1.1.3.	Evaluation des résultats d'identifications et du taux de fausses découvertes	39
2.1.1.3.1.	La stratégie Target-Decoy	40
2.1.1.3.2.	Les approches empiriques de Bayes	41
2.1.2.	Identification par séquençage de novo	42
2.1.3.	Identification par recherche dans les banques de spectres	43
2.1.4.	Les approches de recherches hybrides	44
2.2.	<i>Identification en mode d'acquisition données-indépendant</i>	44
<b>CHAPITRE 3</b>	<b>LES STRATÉGIES DE QUANTIFICATION EN ANALYSE PROTÉOMIQUE</b>	<b>45</b>
1.	Quantification basées sur les gels d'électrophorèse 2D	45
1.1.	<i>Protéomique quantitative à l'aide de l'électrophorèse bidimensionnelle classique</i>	46
1.2.	<i>L'électrophorèse différentielle 2D DiGE</i>	46
2.	Les méthodes sans marquage (« Label Free »)	48
2.1.	<i>La quantification basée sur la comparaison des intensités des signaux MS</i>	48
2.2.	<i>La quantification « Spectral Counting » basée sur les données d'identification</i>	49
3.	Les méthodes de marquage par isotopes stables pour la quantification des protéines par spectrométrie de masse	51
3.1.	<i>Le marquage métabolique</i>	51
3.2.	<i>Le marquage par réaction enzymatique</i>	53
3.3.	<i>Le marquage chimique</i>	54
3.3.1.	Le marquage isotopique	54
3.3.2.	Le marquage isobarique	55
3.4.	<i>Quantification absolue ciblée avec les stratégies de dilution isotopique</i>	56
3.5.	<i>La quantification absolue combinant la dilution isotopique à la SRM</i>	57
4.	Les trois grandes stratégies d'analyse protéomique quantitative par spectrométrie de masse	59
4.1.	<i>Les stratégies « shotgun » ou découverte</i>	59
4.2.	<i>Les stratégies basées sur la MS dirigée</i>	60
4.3.	<i>Les stratégies de MS ciblée</i>	60
4.4.	<i>Les facteurs clés à prendre en compte dans la construction d'une expérience de protéomique quantitative</i>	60

<b>CHAPITRE 4 L'ANALYSE DES SITES DE CLIVAGE PROTÉOLYTIQUE DES PROTÉINES : ÉTAT ACTUEL DE L'ART</b>	<b>63</b>
1. Méthode basée sur une sélection positive des peptides N-terminaux	64
1.1. Stratégie d'enrichissement des peptides N-terminaux par biotinylation des $\alpha$ -amino-groupes	64
1.2. Stratégie de dérivation par le réactif du séquençage d'Edman	65
1.3. La stratégie PICS (« Proteomic Identification of Cleavage Sites ») d'identification des sites de clivage à partir d'une librairie de peptides naturels	65
2. Méthodes basées sur la sélection négative des peptides N-terminaux	67
2.1. La méthode COFRADIC (« COmbined FRActional Dlagonal Chromatography »)	67
2.2. Stratégie de sélection négative des peptides N-terminaux par biotinylation des peptides internes	68
2.3. Stratégie d'enrichissement <i>in silico</i> en peptides N-terminaux à l'aide du marquage iTRAQ	68
2.4. La stratégie TAILS (« Terminal Amine Isotope Labeling of Substrates »)	69
2.5. Synthèse	70
<b>RÉFÉRENCES BIBLIOGRAPHIQUES</b>	<b>73</b>

## Partie II Développements méthodologiques

---

<b>INTRODUCTION</b>	<b>93</b>
1. Le TMPP-Ac-OSu et son utilisation en analyse de protéines par spectrométrie de masse	93
2. Les avantages de la dérivation N-terminale au TMPP pour l'identification des peptides N-terminaux	94
<b>CHAPITRE 1 MISE AU POINT DE LA DÉRIVATION N-TERMINALE DES PROTÉINES DANS LE GEL : LA STRATÉGIE « INGEL N-TOP »</b>	<b>97</b>
1. Développement de la stratégie Ingel N-TOP	97
2. Rendement de la réaction	99
3. La régiosélectivité et spécificité du marquage au TMPP dans le gel	100
4. Conclusion	102
<b>CHAPITRE 2 MISE AU POINT DE LA STRATÉGIE qN-TOP : DÉRIVATION AVEC DU TMPP <sup>13</sup>C POUR LA QUANTIFICATION DU PROTÉOME N-TERMINAL</b>	<b>103</b>
1. Réactifs et marquage différentiel qN-TOP	103
2. Propriétés du réactif [ <sup>13</sup> C <sub>9</sub> ]TMPP-Ac-OSu	104

3. Déroulement de la stratégie qN-TOP	105
4. Choix et paramétrages instrumentaux pour la stratégie qN-TOP	105
4.1. <i>La bonne définition du pic chromatographique</i>	105
4.2. <i>La linéarité de la réponse en analyse de masse</i>	106
5. Conclusion	108
<b>CHAPITRE 3 APPORT DU MARQUAGE AVEC UN NOUVEAU TMPP PLUS HYDROPHOBE POUR ENRICHIR D'AVANTAGE EN PEPTIDE N-TERMINAUX</b>	<b>111</b>
1. Addition d'un groupement méthylène : le TMPP-Propanoyl-OSu	112
2. Ajout d'un groupement propyl : le TMPP-Pentanoyl-OSu	112
2.1. <i>Application à des protéines modèles</i>	113
3. Augmentation de l'hydrophobie des peptides marqués avec le TMPP-Pentanoyl-OSu	114
4. Conclusion	115
<b>RÉFÉRENCES BIBLIOGRAPHIQUES</b>	<b>117</b>

## **Partie III Applications de l'Analyse Protéomique en Biologie**

---

<b>CHAPITRE 1. ÉTUDE DU PROTEOME MITOCHONDRIAL HUMAIN : APPLICATION DE LA STRATÉGIE N-TOP POUR LA CARACTÉRISATION DES PEPTIDES DE TRANSIT</b>	<b>121</b>
1. Résumé	121
2. Cadre de l'étude dans le Projet Protéome Humain (HPP, Human Proteome organization (HuPO))	122
3. Etat des connaissances au début du projet	123
4. Objectif de l'étude	126
5. Stratégie d'analyse	127
5.1. <i>Préparation de l'échantillon</i>	127
5.2. <i>Dérivation des extrémités N-terminales et analyses LC-MS/MS</i>	128
6. Résultats et discussion	130
7. Conclusion et perspectives	135

<b>CHAPITRE 2. ÉTUDE DU PROTEOME DU PARASITE PLASMODIUM FALCIPARUM ET DE L'EXPORT DES PROTEINES PARASITAIRES À L'ÉRYTHROCYTE INFECTÉ.</b>	
<b>APPLICATION DE LA STRATÉGIE N-TOP ET INGEL N-TOP</b>	<b>137</b>
1. Résumé	137
2. Le paludisme en bref	138
3. Etude du mécanisme d'export des protéines à motif PEXEL du parasite Plasmodium falciparum à l'érythrocyte infecté. Application de la stratégie Ingel N-TOP	140
3.1. Contexte de l'étude	140
3.2. Etat des connaissances concernant le motif PEXEL	142
3.3. Stratégie d'analyse : Application de la stratégie Ingel N-TOP	145
3.4. Résultats et discussion	146
3.5. Conclusion	151
4. Etude du protéome N-terminal du parasite P. falciparum dans son stade érythrocytaire	151
4.1. Contexte de l'étude	151
4.2. Stratégie analytique	152
4.2.1. Culture et purification des protéines de chaque compartiment	152
4.2.2. Marquage N-TOP des protéines	152
4.2.3. Analyse LC-MS/MS et identification des peptides et des protéines	153
4.3. Résultats et discussion	154
4.4. Conclusion	156
5. Etude de la voie de translocation de la PfCRT vers la membrane de la vacuole digestive de Plasmodium falciparum	156
5.1. Contexte de l'étude	156
5.2. Etats des connaissances au début de l'étude	157
5.3. Stratégie analytique mise en œuvre	158
5.4. Résultats et discussion	158
5.5. Conclusion	162
Publication	165
<b>CHAPITRE 3 ANALYSE PROTÉOMIQUE DU TISSU ADIPEUX DU RAT EN RÉPONSE À UNE PRIVATION DE NOURRITURE PLUS OU MOINS PROLONGÉE</b>	<b>167</b>
1. Résumé	167
2. Contexte de l'étude	167
3. Le modèle du jeûne prolongé	168
4. Position du problème	168
5. Objectif	169

6. Stratégie analytique	169
7. Matériels et méthode	170
7.1. <i>Traitement des animaux et prélèvements d'échantillons</i>	170
7.2. <i>Extraction des protéines</i>	171
7.3. <i>Marquage des protéines</i>	171
7.4. <i>Electrophorèse 2D</i>	172
7.5. <i>Analyse des images de gels 2D</i>	172
7.6. <i>Analyse LC-MS/MS et identification des protéines</i>	173
8. Résultats et discussion	173
8.1. <i>Validation du statut métabolique des animaux</i>	173
8.2. <i>Protéines du tissu adipeux différenciellement exprimées en réponse au jeûne prolongé</i>	174
9. Conclusion	178
<b>CHAPITRE 4 APPLICATION DE LA STRATÉGIE qN-TOP À L'ÉTUDE DU PROTÉOME DU FOIE DU RAT SOUMIS AU JEÛNE PROLONGÉ</b>	<b>179</b>
1. Résumé	179
2. Contexte de l'étude	179
3. Stratégie analytique	180
3.1. <i>Préparation des échantillons</i>	180
3.2. <i>Marquage différentielle et analyse par nanoLC-MS/MS</i>	181
3.3. <i>Identification des protéines et caractérisation des peptides N-terminaux</i>	181
4. Résultats et discussion	182
4.1. <i>Identification des protéines et des processus protéolytiques</i>	182
4.2. <i>Quantification des peptides N-terminaux</i>	182
5. Conclusion	186
<b>CHAPITRE 5 ÉTUDE FONCTIONNELLE DE REP1-NCXSQ, UNE PROTÉINE QUI RÉGULE L'ÉCHANGEUR Na<sup>+</sup>/Ca<sup>2+</sup> DU NERF DU CALAMAR</b>	<b>187</b>
1. Résumé	187
2. Contexte de l'étude	187
3. Brève introduction à l'étude de complexe non-covalent par spectrométrie de masse supramoléculaire	188
3.1. <i>La préparation de l'échantillon</i>	188
3.2. <i>Les paramètres instrumentaux</i>	188
4. Stratégie analytique mise en œuvre	189
4.1. <i>Caractérisation des complexes par spectrométrie de masse supramoléculaire</i>	189

4.2.	<i>Caractérisation des acides gras par LC-HRMS</i>	190
4.3.	<i>Résultats et discussion</i>	190
5.	Conclusion	193
	Publication	195
	<b>RÉFÉRENCES BIBLIOGRAPHIQUES</b>	<b>197</b>

## **Conclusion générale**

---

<b>CONCLUSION GÉNÉRALE</b>	<b>209</b>
----------------------------	------------

## **Partie expérimentale**

---

<b>PARTIE EXPÉRIMENTALE</b>	<b>215</b>
-----------------------------	------------

## **Travaux supplémentaires**

---

<b>TRAVAUX SUPPLÉMENTAIRES</b>	<b>227</b>
--------------------------------	------------



## Liste des abréviations

AQUA	:	Absolute QUAntification
ARN	:	Acide RiboNucléique
ARNm	:	ARN méssager
BLAST	:	Basic Local Alignment Search Tool
CAD	:	Collision Activated Dissociation
CEA	:	Commissariat à l'Energie Atomique
CID	:	Collision Induced Dissociation
COFRADIC	:	COmbined FRActional DIagonal Chromatograh
CQR	:	Chloroquine Resistant
CQS	:	Chloroquine Sensitive
DDA	:	Data Dependent Acquisition
DDT	:	DichloroDiphényl(Trichloroéthane)
DIA	:	Data Independent Acquisition
DIC	:	Differential Interference Contrast
DiGE	:	Differential Gel Electrophoresis
DTT	:	Dithiotréitol
ECD	:	Electron Capture Dissociation
EDTA	:	acide Ethylène DiamineTetra-Acétique
EPI	:	tissu adipeux Epididymère
ESI	:	ElectroSpray Ionisation
ETD	:	Electron Transfer Dissociation
FAB	:	Fast Atom Bombardment
FAO	:	Food and Agriculture Organization
FDR	:	False Discovery Rate (Taux de fausse découverte)
FT-ICR	:	Fourrier Transform - Ion Cyclotron Resonance
GFP	:	Green Fluorescent Protein
HCT	:	Host Cell Targeting signal
HEPES	:	acide 4-(2-Hydroxyéthyl)-1-Pipérazine Ethane Sulfonique
Hoe	:	Hoechst stain (colorant de Hoechst)
HPLC	:	High Performance Liquide Chromatography
HPP	:	Human Proteome Project

HR : Haute Résolution  
Hsp : Heat shock protein  
HUPO : HUman Proteome Organization  
ICAT : Isotope-Coded Affinity Tag  
ICPL : Isotope-Coded Protéine Label  
iTRAQ : isobaric Tag for Relative and Absolute Quantification  
LC : Liquid Chromatography  
LOD : Limite de Detection  
LOQ : Limite de Quantification  
MACS : Magnetic-Activated Cell Sorting  
MALDI : Matrix Assisted Laser Desorption Ionization  
MCP : MultiChannel Plate  
MELAS : Mitochondrial Encephalopathy, Lactic Acidosis and Stroke like episodes.  
MERRF : Myoclonic Epilepsy with Ragged Red Fibers  
MgATP : Adénosine TriPhosphate de Magnésium  
MPE : Membrane Plasmique de l'Erythrocyte  
MPP : Membrane Plasmique du Parasite  
MRE : Membrane du Réticulum Endoplasmique  
MS : Mass Spectrometry  
Mthsp : Matrix Heat shock protein  
NCBI : National Center for Biotechnology Information (Etats-Unis)  
NHS : N-HydroxySuccinimide  
N-TOP : N-Terminal Oriented Proteomics  
OMS : Organisation Mondiale de la Santé  
OMSSA : Open Mass Spectrometry Search Algorithm  
PAGE : PolyAcrylamide Gel Electrophoresis  
PAM : Presequence translocase-Associated Motor  
PBS : Phosphate Buffer Solution  
PEXEL : plasmodium Protein EXport ELeMent  
PfCRT : Plasmodium falciparum Chloroquine Resistance Transporter  
PfEMP1 : Plasmodium falciparum Membrane Protein 1  
PFF : Peptide Fragment Fingerprinting  
PfMDR1 : P. falciparum Multi-Drug Resistance Transporter 1  
PfMRP : P. falciparum Multi-drug Resistance associated Protein  
PICS : Proteomic Identification of Cleavage Sites  
PMF : Peptide Mass Fingerprint  
PTEX : Plasmodium Translocon of EXported proteins  
qN-TOP : quantitative N-Terminal Oriented Proteomics  
Q-TOF : Quadrupole-Time of Flight  
RE : Réticulum Endoplasmique

SC : tissu adipeux Sous-Cutané  
SCX : Strong Cation eXchange  
SDS : DodécylSulfate de Sodium  
SIB : Swiss Institute of Bioinformatics, Institut Suisse de Bioinformatique  
SIDA : Syndrome de l'ImmunoDéficiency Acquis  
SILAC : Stable Isotope Labeling by Amino-acids in Cell culture  
SILE : Stable Isotope Labeling Experiments  
SRM : Selected Reaction Monitoring  
TAILS : Terminal Amine Isotope Labeling of Substrates  
TCEP : Tris(2-CarboxyEthyl)Phosphine  
TIM : Translocase of the Inner Membrane  
TMPP : tris(2,4,6-TriMethoxyPhenyl)Phosphonium  
TMPP-Ac-Osu : (N-Succinimidylloxycarbonylmethyl)tris(2,4,6-trimethoxyphenyl)phosphonium bromide  
TNBS : acide 2,4,6-trinitrobenzenesulfonique  
TOF : Time of Flight  
TOM : Translocase of the Outer Membrane  
UHPLC : Ultra High Performance Liquid Chromatography  
VTS : Vacuolar Transport Signal



# Introduction générale

---



## Introduction générale

Depuis le début des années 90, la spectrométrie de masse a progressivement donné accès de plus en plus facilement à l'analyse protéomique à très haut débit, ce qui en a fait une des méthodologies de base en biologie. A présent, l'analyse des protéomes par spectrométrie de masse tend, en plus de la simple identification des protéines présentes, vers la caractérisation des protéines, qui consiste à déterminer l'enchaînement exact de tous les acides aminés, y compris de toutes les modifications post-traductionnelles.

Le clivage protéolytique, ou protéolyse, est une modification post-traductionnelle omniprésente et d'une importance majeure. Elle affecte toutes les protéines à un moment ou un autre de leur cycle de vie, et compte tenu de son ubiquité, elle «sculpte» tout le protéome d'une façon qui ne peut être prédite en général, par la lecture des séquences génomiques. Seule l'analyse directe des séquences des protéines peut déterminer avec exactitude sous quelles formes celles-ci sont présentes dans le milieu biologique et d'établir ainsi des relations structure/activité pertinentes. Dans le cas des maturations et du clivage du peptide signal en N-terminal d'une protéine, ce processus protéolytique permet aux protéases de moduler finement les fonctions des protéines. Au travers de ces processus protéolytiques, les protéases participent à bon nombre de processus biologiques comme la réplication de l'ADN, la translocation des protéines, la différenciation et la migration des cellules, etc. Par conséquent, si les sites de protéolyse ne sont pas déterminés lors de l'analyse protéomique, les relations de structure/activité des protéines ne pourront être établies de façon sûre. Il est donc indispensable de disposer de méthodes analytiques pour déterminer avec précision les sites de clivage et qui soit compatibles avec l'analyse protéomique à haut débit. Malheureusement, il n'existe pas aujourd'hui de telles méthodes qui soient utilisables en routine.

Pour identifier un clivage protéolytique, il est indispensable de déterminer sans ambiguïté la nouvelle extrémité N-terminale ou C-terminale générée par la protéolyse de la protéine. Lorsque les stratégies d'analyse protéomique sont orientées vers la détermination des sites de clivages, on parle de dégradomique et de processomique. Une méthodologie originale de marquage et d'enrichissement en peptides N-terminaux appelée N-TOP (N-Terminal Oriented Proteomics) a été développée au laboratoire et mon travail de thèse a

porté, en partie, sur des améliorations de cette méthode afin d'en élargir le champ d'applications en biologie. Ce manuscrit de thèse s'articule en trois parties, comme suit.

La partie I est une étude bibliographique composée de quatre chapitres, et qui présente les principaux outils et méthodologies de l'analyse protéomique en relation avec notre travail:

- Le premier chapitre décrit les différents types de spectromètres de masse, ainsi que les techniques séparatives utilisés en analyse protéomique.
- Le deuxième chapitre décrit les différentes stratégies d'identification des protéines utilisées en analyse protéomique, ainsi que les approches bioinformatiques utilisées pour la validation des identifications.
- Le troisième chapitre est consacré à la description des principales méthodes de quantification des protéines et des peptides.
- Le quatrième chapitre décrit les différentes méthodes et stratégies d'analyse des processus protéolytiques.

La partie II est consacrée aux développements méthodologiques de la stratégie de marquage et d'enrichissement en peptides N-terminaux N-TOP. Après une description de cette stratégie, je présenterai trois développements méthodologiques que j'ai apportés à cette stratégie au cours de ma thèse.

- Le premier chapitre décrit la stratégie Ingel N-TOP qui consiste à réaliser la réaction de dérivation chimique de la position N-terminale directement dans une bande de gel 1D contenant une protéine d'intérêt, plutôt que sur une solution liquide d'un extrait protéique total. Cette approche de marquage permettrait non seulement de cibler des protéines présentes en faibles quantités mais également de s'affranchir des problèmes de compatibilité des tampons de purification de certains échantillons.
- Le deuxième chapitre présente la mise au point d'une dimension quantitative ajoutée à la stratégie N-TOP. Il s'agit de la stratégie qN-TOP (« quantitative N-terminal Proteomics ») qui, grâce à la synthèse d'un réactif de marquage marqué aux isotopes stables, permettra de quantifier les processus protéolytiques que subissent les protéines à l'échelle des protéomes.
- Le troisième chapitre est consacré à la mise au point d'un nouveau réactif de marquage des amines N-terminales qui vise, grâce à l'augmentation de son hydrophobie, à mieux séparer en chromatographie les peptides N-terminaux marqués des peptides internes améliorant ainsi leur sélection par le spectromètre de masse pour la fragmentation et permettant par la suite d'en identifier un plus grand nombre.

La Partie III est consacrée à une suite d'applications des stratégies décrites dans la partie II ainsi que d'autres stratégies d'analyse protéomique dans le cadre d'études en biologie dans le but d'apporter des éléments de réponse à différentes problématiques.

- Le chapitre 1 présente une étude du protéome mitochondrial humain réalisée dans le cadre du « Human Proteome Project » en collaboration avec le groupe CALIPHO de l'Institut Suisse de Bioinformatique et du CEA de Grenoble. Dans cette étude, nous avons appliqué la stratégie N-TOP dans le but de caractériser les peptides de transit des protéines mitochondriales. Le clivage de ces peptides de transit intervient dans le mécanisme d'import des protéines dans les mitochondries.
- Le chapitre 2 regroupe trois études, réalisées en collaboration avec l'Université d'Heidelberg, sur le parasite *Plasmodium falciparum* responsable du paludisme. La première a consisté à étudier le mécanisme d'exportation des protéines parasitaires vers l'érythrocyte à travers un compartiment de transit, la vacuole parasitophore. La majorité de ces protéines contient un motif conservé qui intervient dans le mécanisme d'export et qui semble subir des PTM, dont les clivages protéolytiques. Nous avons appliqué la stratégie Ingel N-TOP afin d'étudier ces clivages protéolytiques et leur rôle dans la reconnaissance et l'export de ces protéines. La deuxième étude s'agit de l'application de la stratégie N-TOP au protéome du parasite dans son stade érythrocytaire. Cette étude n'est que la première étape d'un projet plus large visant à étudier le protéome des différents compartiments de l'érythrocyte infecté par le *P. falciparum*. Enfin, la troisième étude est consacrée au mécanisme de translocation d'une protéine transporteur membranaire impliquée dans la résistance du parasite aux antibiotiques. La caractérisation de cette protéine par spectrométrie de masse a permis de mettre en évidence une phosphorylation indispensable à la translocation de cette protéine vers sa destination.
- Le chapitre 3 consiste en une étude protéomique différentielle du protéome du tissu adipeux du rat soumis au jeûne prolongé par une approche Gel 2D DiGE et nanoLC-MS/MS.
- Dans le chapitre 4 je présente une application de la stratégie qN-TOP décrite dans la partie II sur le protéome hépatique du rat soumis au jeûne prolongé en mettant plus particulièrement l'accent sur les principaux problèmes rencontrés pour la quantification des processus protéolytiques.
- Le chapitre 5 décrit une étude fonctionnelle de la protéine ReP1 qui régule l'échangeur  $\text{Na}^+/\text{Ca}^{2+}$  du nerf du calamar. Cette protéine interagit avec des acides gras et nous avons mis en place une approche supramoléculaire afin de caractériser ces interactions. Nous avons développé en parallèle une stratégie LC-MS différentielle afin d'identifier l'acides gras fonctionnels qui activent cette protéine.



# Partie I. Introduction bibliographique

---

## **Chapitre 1**

Les outils et technologies au service de  
l'analyse protéomique

## **Chapitre 2**

Les stratégies d'identification en analyse  
protéomique

## **Chapitre 3**

Les stratégies de quantification en analyse  
protéomique

## **Chapitre 4**

L'analyse des sites de clivage protéolytique des  
protéines : Etat actuel de l'art



## Chapitre 1 Les outils et technologies au service de l'analyse protéomique

Jusqu'au début des années 1990, l'analyse des séquences des protéines s'effectuait essentiellement à l'aide du séquençage par la dégradation d'Edman [1]. Cette technique permet de déterminer la séquence primaire des protéines par analyse des acides aminés en Chromatographie Liquide Haute Performance (HPLC). Cependant elle est limitée à l'analyse des protéines dont l'extrémité N-terminale n'est pas bloquée et n'est pas compatible avec les mélanges complexes. De plus, elle nécessite des quantités importantes de protéine (de l'ordre de la nanomole jusqu'à la picomole) et le processus de séquençage est assez lent.

Dans les années 1990, deux nouvelles techniques d'ionisation ont révolutionné la spectrométrie de masse et lui ont permis de supplanter le séquençage d'Edman. Il s'agit de l'ionisation ElectroSpray (ESI) [2] et de l'ionisation par Désorption Laser assistée par Matrice (MALDI) (« Matrix Assisted Laser Desorption Ionisation ») [3,4] qui sont des sources d'ionisation douce permettant l'ionisation des biomolécules telles que les peptides et les protéines et leur transfert en phase gazeuse sans induire de fragmentation. Ces découvertes ont valu à John Fenn et Koichi Tanaka le prix Nobel de chimie en 2002. En effet, avant l'apparition de ces deux nouvelles techniques, la spectrométrie de masse se limitait à l'analyse de petites molécules thermostables du fait de l'absence de source permettant d'ioniser et de volatiliser les biomolécules. De nos jours, la bonne sensibilité de la spectrométrie de masse, sa polyvalence et sa rapidité en font un outil de choix incontestable pour le séquençage, la caractérisation et même la quantification des protéines [5-9].

Une autre avancée technologique importante pour l'analyse protéomique réside dans les développements qu'ont connus les techniques de séparation des protéines et des peptides (gels d'électrophorèse mono et bidimensionnelles et chromatographie liquide). Ces techniques séparatives viennent en amont de la spectrométrie de masse et facilite l'analyse d'échantillons complexes contenant des milliers de protéines.

Enfin, le développement des banques de données protéiques et le séquençage d'un nombre croissant de génomes ont aussi contribué à l'essor de l'analyse protéomique par spectrométrie de masse.

L'analyse protéomique est généralement composée de quatre étapes (Figure 1). Les protéines sont extraites à partir d'un échantillon biologique et sont ensuite séparées ou fractionnées avant et/ou après digestion enzymatique. Les peptides ainsi obtenus sont analysés par spectrométrie de masse en tandem et les spectres générés serviront à l'identification, à la caractérisation et/ou à la quantification des protéines à l'aide d'outils bioinformatiques.

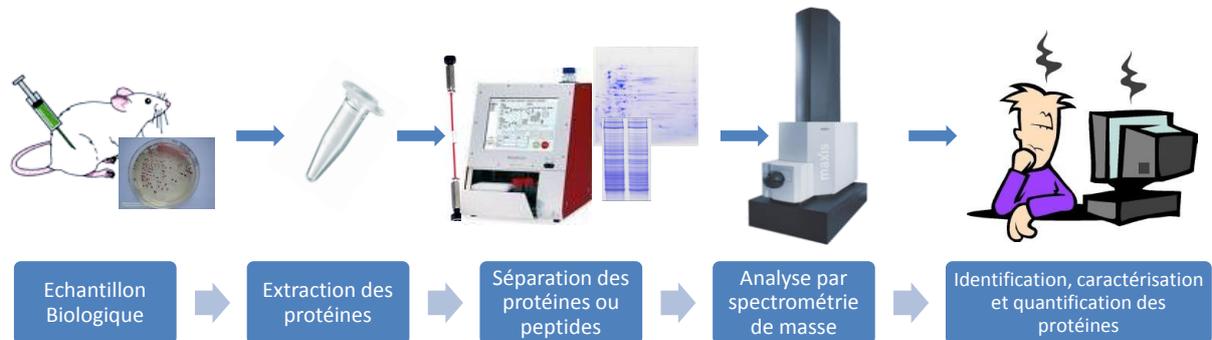


Figure 1 : Les quatre grandes étapes de l'analyse protéomique.

Dans ce chapitre, je présenterai donc les principaux outils de spectrométrie de masse et de techniques séparatives utilisées pour l'analyse protéomique.

## 1. La spectrométrie de masse pour l'analyse protéomique

Généralement, trois principaux éléments constituent un spectromètre de masse : une source d'ionisation, un analyseur de masse et un détecteur de courants d'ions. Les molécules en phase liquide ou solide sont ionisées et transférées en phase gazeuse à l'aide de la source. Les ions sont ensuite conduits par des champs électriques et magnétiques dans l'analyseur où ils seront séparés en fonction de leur rapport masse sur charge ( $m/z$ ). La combinaison de plusieurs analyseurs ainsi que le fonctionnement de certains analyseurs permettent d'effectuer de la spectrométrie de masse en tandem. Ceci permet d'obtenir des informations structurales en fragmentant les ions dans un premier temps et en mesurant les spectres de masse de ces fragments dans un deuxième temps. Enfin, les ions arrivent au détecteur ce qui permet de générer un courant électrique qui est en suite converti en spectres de masse dans lesquels l'intensité du signal est mise en relation avec le rapport  $m/z$ .

### 1.1. Les sources électrospray

Plusieurs sources d'ionisation sont utilisées en spectrométrie de masse. Cependant, pour l'analyse de biomolécules telles que les protéines et les peptides, le choix se limite principalement aux deux sources qualifiées de techniques d'ionisation douce : l'ionisation Electropray (ESI) [2] et l'ionisation MALDI (« Matrix Assisted Laser Desorption Ionisation ») [3]. Ces deux techniques permettent l'ionisation et le transfert de biomolécules intactes en phase gazeuse vers l'analyseur de masse.

### 1.1.1. La source d'ionisation électrospray

La source électrospray est une technique d'ionisation d'analytes en solution à pression atmosphérique ce qui permet de la coupler facilement avec la chromatographie liquide [10]. L'ionisation est réalisée par la nébulisation d'un fin nuage d'ions créé sous l'effet d'un champ électrique de l'ordre de 10 millions de Volts par mètre et de l'assistance pneumatique d'un gaz de nébulisation [2]. La différence de potentiel appliquée entre le capillaire dans lequel se trouve la solution de l'analyte à ioniser et l'entrée du spectromètre de masse provoque la polarisation du liquide, la séparation des espèces de charge opposée. Un excès de charges va se concentrer à la pointe du capillaire entraînant l'étirement du liquide en ce qu'on appelle le cône de Taylor et l'émission de gouttelettes multichargées (Figure 2). Un flux d'azote coaxial appelé gaz de nébulisation sert à stabiliser les gouttelettes. Le chauffage de l'enceinte favorisera l'évaporation du solvant et la densité de charge augmentera au sein des gouttelettes avec la diminution de leur taille jusqu'à ce que la limite de stabilité de Rayleigh [11] soit atteinte : les forces de répulsion coulombiennes dépassent alors la tension de surface des gouttelettes qui subissent alors une succession d'explosion coulombienne au fur et à mesure que le solvant s'évapore jusqu'à la production d'ions entièrement désolvatés.

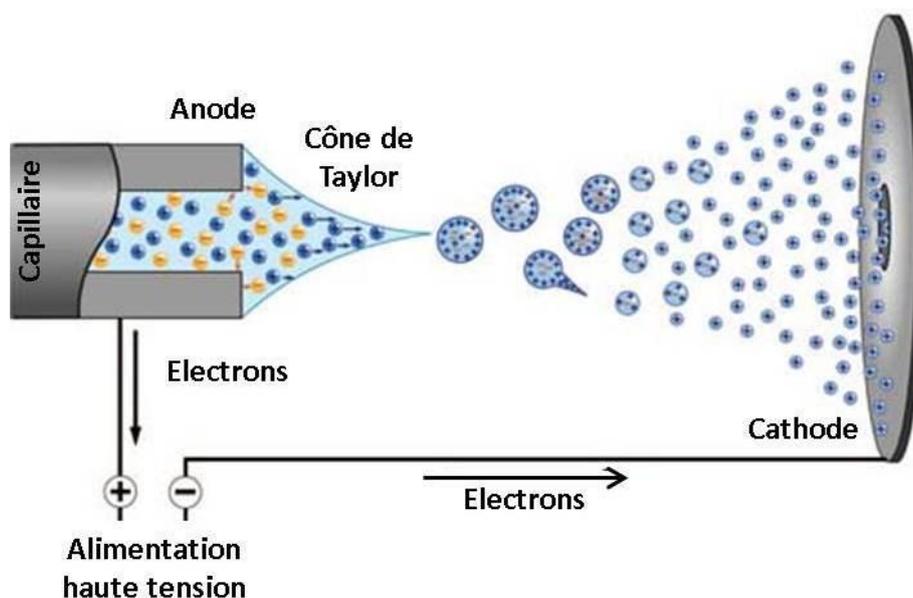


Figure 2 : L'ionisation Electrospray en mode d'ionisation positif.

Deux modèles, encore assez controversés, décrivent la production des ions désolvatés (Figure 3). Le premier est celui proposé par Dole [12] et appelé modèle de la charge résiduelle. Ce modèle propose que l'évaporation du solvant et le processus d'explosions coulombiennes se répètent jusqu'à ce qu'il n'y ait plus qu'une charge par gouttelette. Le deuxième modèle est celui d'Iribarne et Thomson [13]; il propose le mécanisme de l'évaporation ionique selon lequel la densité de charge devient suffisamment importante pour émettre directement les ions en phase gazeuse lors de l'évaporation du solvant. Si ce dernier semble être le plus reconnu pour les petites molécules, le mécanisme

du modèle de Dole semblerait intervenir pour les macromolécules telles que les protéines [14].

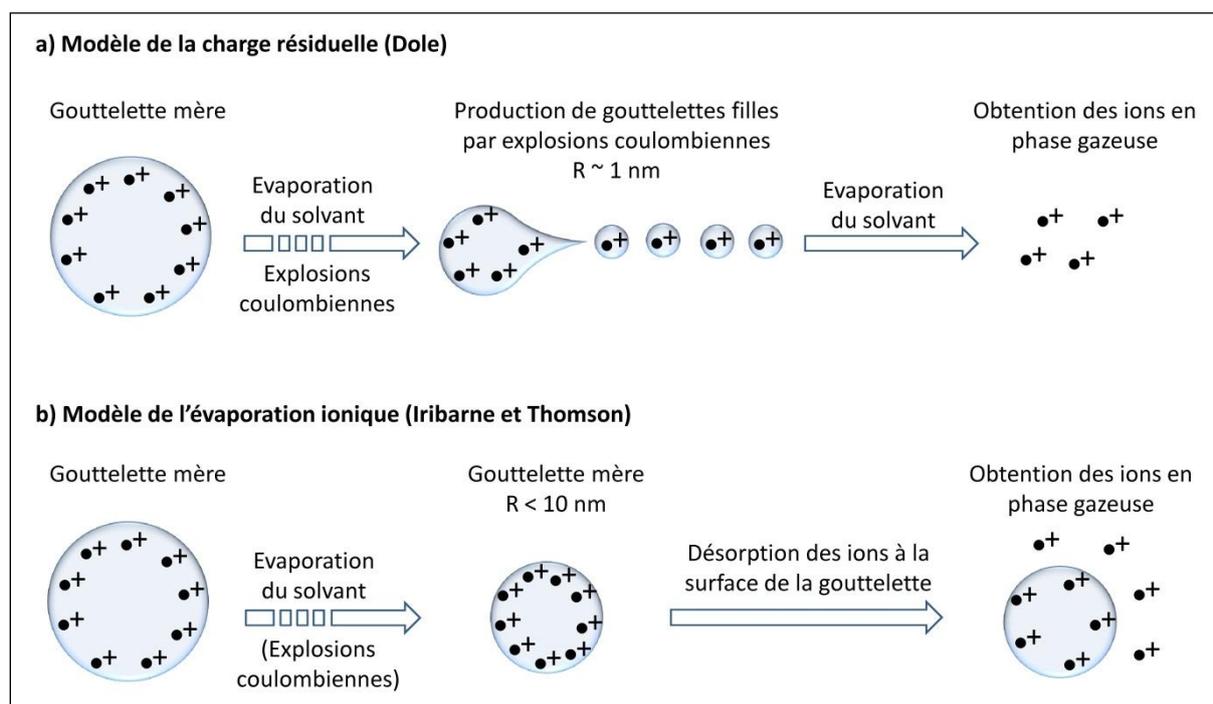


Figure 3 : Emission des ions en phase gazeuse selon a) le mécanisme de la charge résiduelle et b) le mécanisme de l'évaporation ionique (d'après [15]).

A condition qu'elles puissent être ionisées, les molécules en solution sont rapidement transformées en ions en phase gazeuse. Lorsqu'une molécule possède plusieurs fonctions ou sites ionisables, on observe la production d'ions multichargés [16]. Ceci permet d'étendre considérablement la gamme de masse puisque avec un nombre de charges élevé, les valeurs  $m/z$  mesurées seront plus petites [17]. Trois principaux mécanismes d'ionisation peuvent entrer en jeu lors du processus électrospray. Le premier étant l'ionisation par séparation des charges, il intervient pour les molécules possédant des groupements acides ou basiques, ou un groupement phosphonium, ammonium ou oxonium. Le deuxième mécanisme fait intervenir la formation d'adduits ; les analytes polaires ne disposant pas de fonctions acides ou basiques peuvent ainsi s'ioniser en ESI par la formation d'adduits avec d'autres ions présents en solution. Les ions alcalins ( $\text{Li}^+$ ,  $\text{Na}^+$  et  $\text{K}^+$ ) sont souvent utilisés pour ioniser par formation d'adduits. Le troisième mécanisme est celui de l'ionisation par réaction en phase gazeuse. En effet, une fois en phase gazeuse, le fait que le processus ESI soit réalisé à pression atmosphérique et que de grandes quantités de molécules de solvant chargées soient générées en plus de l'analyte, il est possible qu'il y ait des réactions de transfert de protons vers l'analyte [18]. Les molécules protonées en solution peuvent ainsi, en phase gazeuse, céder leurs protons à d'autres molécules qui ont une basicité supérieure en phase gazeuse.

Un développement important des sources électrospray est né de la possibilité de réduire le débit de la solution d'analyte utilisée pour créer le spray. Les premières sources

nanospray (débit de 20 nL.Min<sup>-1</sup> en infusion) sont apparues en 1996. En effet, les modèles de Dole et d'Iribarne et Thomson s'accordent pour faire apparaître l'intérêt d'obtenir des gouttelettes initiales de petite taille. Les travaux qui ont été réalisés dans ce contexte ont mis en évidence l'influence du débit de liquide sur la formation des gouttelettes chargées [19,20]. Lorsque le débit est diminué, aucune assistance pneumatique n'est nécessaire et le diamètre de l'orifice de sortie du capillaire est réduit, le rayon des gouttelettes ainsi formées est réduit à 100 nm soit 10 fois plus petit que celui des gouttelettes formées par une source ESI classique (2 µm environ) [20,21]. Ceci permet une meilleure efficacité d'ionisation et par la suite un gain de sensibilité parce que la spectrométrie de masse est une technique sensible à la concentration des analytes et non à leur quantité. De plus cette source permet une réduction significative de la consommation d'échantillon. Pour tous ces avantages, la source nanospray couplée aux techniques de nano-chromatographie liquide (nanoLC) est particulièrement bien adaptée à l'analyse protéomique.

A côté des systèmes nanoLC standards, de nouveaux dispositifs nanospray sur puces microfluidiques ont été mis au point [22]. Ceux-ci présentent l'avantage d'éviter une manipulation délicate des aiguilles nanospray peu reproductible grâce à des dispositifs intégrés et motorisés et de permettre une conception reproductible des pièces grâce à la technologie laser. Parmi ces dispositifs, on peut citer le système Triversa Nanomate (Advion Biosystem) qui est un automate passeur d'échantillons, le système nanoLC Chip Cube<sup>TM</sup> (Agilent technologies) et le système TRIZAIC UPLC<sup>TM</sup> (Waters) qui sont des systèmes de colonne chromatographique intégrée sur puce microfluidique.

### ***1.1.2. La source MALDI***

La technique d'ionisation MALDI est elle aussi une technique douce d'ionisation bien adaptée pour l'analyse des biomolécules. Elle consiste à co-cristalliser l'échantillon avec une matrice organique à caractère acide (généralement de faible masse moléculaire) et à irradier sous vide le cristal analytes-matrice avec un faisceau laser pulsé de longueur d'onde se situant dans l'ultra-violet. C'est donc la matrice qui absorbe le rayonnement laser plutôt que d'ajuster la longueur d'onde du faisceau laser à la longueur d'onde d'absorption des échantillons [23]. L'excitation de la matrice induit sa dissociation puis sa sublimation. Ceci entrainera donc l'analyte dans le plasma d'expansion ainsi formé où se produira l'ionisation des molécules. Les ions d'analytes ainsi produits sont accélérés à l'aide d'une différence de potentiel et transférés jusqu'à l'analyseur [24]. En général, l'ionisation MALDI produit des ions monochargés et donc des spectres de masse relativement simples à interpréter. Contrairement à l'ESI, cette technique a l'avantage d'être relativement tolérante à la présence des sels. Les phénomènes qui se produisent lors de cette ionisation sont encore sujets à débat mais un modèle en trois étapes semble se dégager [25] (Figure 4):

- Les molécules de la matrice sont chauffées grâce à l'énergie du rayonnement laser pulsé.

- Cet échauffement induit un abaissement des forces de cohésion des cristaux ce qui entraîne la désorption des molécules des analytes.
- Un transfert de protons entre les molécules de la matrice et celles des analytes se produit en phase gazeuse, soit par collisions, soit par réactions acido-basiques et les molécules des analytes sont ainsi ionisées.

Le choix de la matrice ainsi que le mode de dépôt sont des paramètres essentiels qui entrent en jeu pour l'efficacité de l'ionisation. Les matrices les plus utilisées sont l'acide  $\alpha$ -cyano-4-hydroxycinnamique, l'acide 2,5-dihydroxybenzoïque, l'acide sinapique et la 2',4',-dihydroxyacétophénone tandis que les modes de dépôt sont le dépôt en couche mince, en sandwich ou en goutte séchée. Le choix de ces deux paramètres dépend de la nature de l'échantillon.

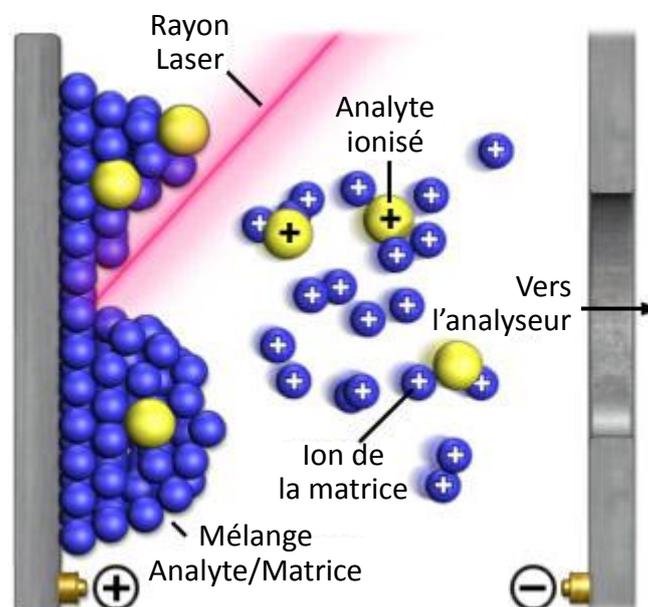


Figure 4: Principe de l'ionisation MALDI (adapté de [26]).

Cette technique d'ionisation est souvent utilisée avec un analyseur à temps de vol (Time Of Flight, TOF). Celui-ci est particulièrement bien adapté à l'analyse des paquets d'ions générés par l'impulsion laser. Le fonctionnement de cet analyseur sera détaillé dans le paragraphe 1.2.3 de ce chapitre.

## 1.2. Les analyseurs de masse

Plusieurs types d'analyseurs de masse ont été développés durant le siècle dernier. Ces différents types d'analyseurs se distinguent par le principe physique mis en jeu pour la mesure du rapport masse sur charge ( $m/z$ ). Les caractéristiques fondamentales d'un analyseur sont sa sensibilité, sa résolution, sa précision de mesure, sa gamme de masse, sa capacité à réaliser de la spectrométrie de masse en tandem, sa vitesse d'acquisition, sa gamme dynamique et sa gamme de linéarité pour les analyses de quantification [27]. Ici, je décrirai les principaux analyseurs utilisés en analyse protéomique en insistant sur ceux que

j'ai utilisés au cours de ma thèse, à savoir les trappes ioniques et les analyseurs hybrides quadripôle-temps de vol.

### 1.2.1. L'analyseur quadripolaire

Un quadripôle est constitué de 2 paires de barres métalliques parallèles. Un champ électrique en opposition de phase est appliqué à l'une et l'autre de ces paires. Ceci lui permet de fonctionner en tant que filtre de masse [28]. La séparation des ions en fonction de leur rapport masse sur charge ( $m/z$ ) dépend de la stabilité de leur trajectoire dans le quadripôle (Figure 5). Les équations de Mathieu décrivent ces trajectoires et permettent de déterminer les conditions de stabilité pour chaque  $m/z$ . Un balayage du potentiel permet donc de transmettre successivement tous les ions selon leur  $m/z$  dans la gamme de balayage donnée. A potentiel fixe, le quadripôle devient un filtre à ions et ne transmet que le  $m/z$  correspondant à ce potentiel. Cet analyseur peut être couplé à un autre analyseur (quadripôle ou autres) pour réaliser des analyses de spectrométrie de masse en tandem.

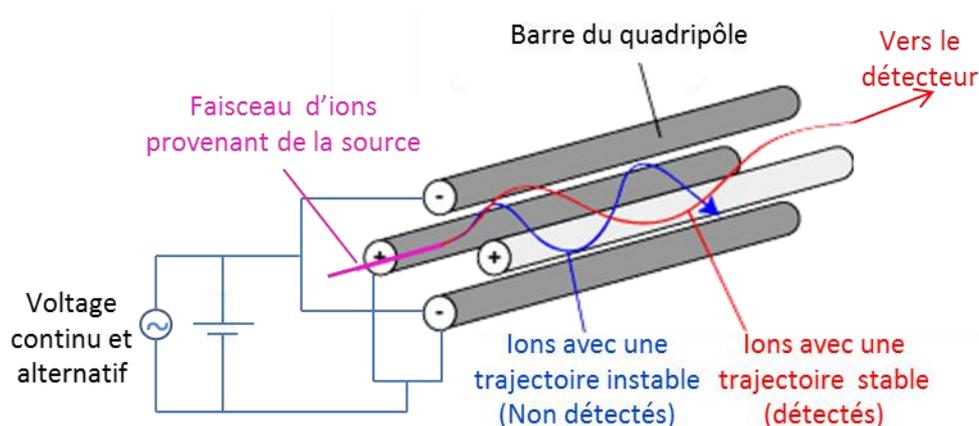


Figure 5 : Trajectoires stable (rouge) et instable (bleu) dans un quadripôle.

### 1.2.2. L'analyseur Trappe Ionique

Nous nous limiterons dans ce paragraphe à la description de l'analyseur trappe ionique cubique (3D) qui a été utilisé dans les travaux décrits dans ce manuscrit. Cependant, une génération récente de trappes d'ions de géométrie linéaire (2D) a été développée et est décrite dans la littérature [29,30]. L'analyseur trappe d'ions a été décrit pour la première fois par W. Paul dans les années 50. Il est constitué de trois électrodes, une centrale de forme annulaire, et deux électrodes chapeaux quasi hyperboliques [31]. L'ensemble forme une cavité d'environ  $1 \text{ cm}^3$  (Figure 6) dans laquelle un gaz tampon, l'hélium, est maintenu à une pression d'environ  $5 \cdot 10^{-3} \text{ mbar}$ . Les ions entrent par l'orifice situé dans l'une des électrodes chapeaux et restent confinés à l'intérieur de la trappe sans la traverser directement grâce à une différence de potentiel oscillante (radiofréquence) appliquée à l'électrode annulaire permettant de générer un champ quadripolaire qui stabilise la trajectoire des ions en les piégeant dans un puits de potentiel [31,32]. Cette trajectoire décrit dans l'espace une courbe de Lissajous et est régie par les équations de Mathieu (Figure 6).

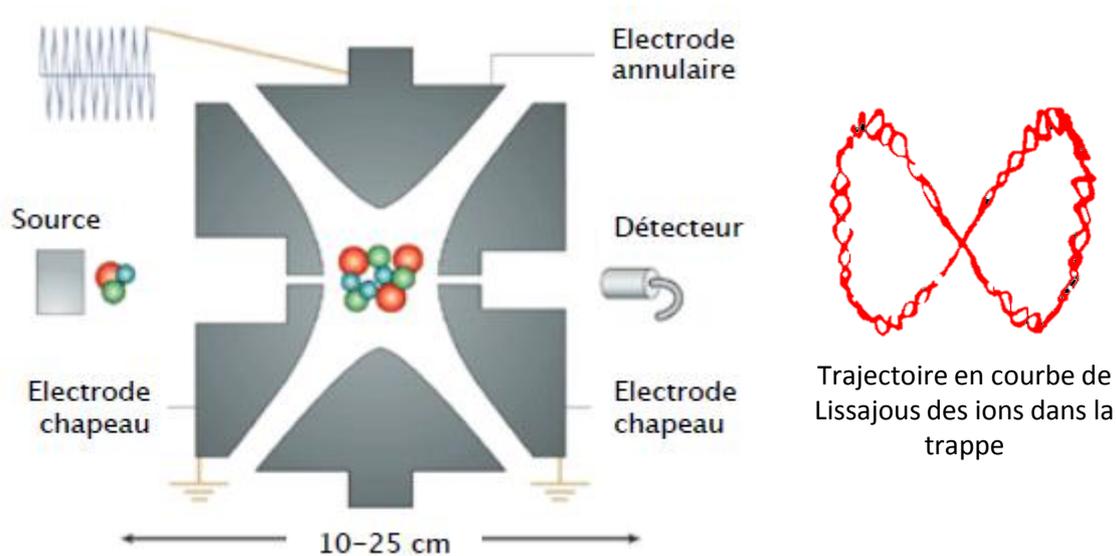


Figure 6 : Représentation schématique de la trappe d'ions. Les ions sont piégés sur une trajectoire en courbe de Lissajous à l'aide des tensions oscillantes (radiofréquence) de l'électrode centrale. La mesure de masse est réalisée en déstabilisant ces trajectoires en fonction du  $m/z$  ce qui permet d'éjecter les ions vers le détecteur (d'après [33]).

L'éjection des ions est réalisée à l'aide de deux méthodes :

- Ejection simple par balayage croissant de l'amplitude de la radiofréquence.
- Ejection résonante par application d'une tension alternative supplémentaire sur les électrodes chapeaux dont on fait varier la fréquence pour faire entrer en résonance les ions de  $m/z$  différents.

Grâce à ses capacités de piégeage et d'éjection sélectifs des ions, cet analyseur offre la possibilité de réaliser dans le temps des analyses de spectrométrie de masse en tandem ( $MS^2$ , ...,  $MS^n$ ). L'isolation d'un ion consiste à éjecter de la trappe tous les ions n'ayant pas le  $m/z$  souhaité. Celle-ci est réalisée par éjection simple des ions de  $m/z$  inférieurs à celui de l'ion à isoler et par éjection résonante des ions de  $m/z$  supérieurs au  $m/z$  de l'ion à isoler. Après isolation, la fragmentation peut être réalisée par ajustement de la fréquence de la tension alternative appliquée aux électrodes chapeaux de façon à ce qu'elle corresponde à la fréquence de résonance de l'ion à fragmenter. L'amplitude de cette tension est réglée pour éviter d'éjecter les ions à fragmenter de la trappe. Les ions ainsi excités entrent en collision avec les molécules d'hélium ce qui transforme leur énergie interne vibrationnelle en énergie cinétique entraînant leur fragmentation. Ces différentes étapes, piégeage, isolation, excitation, fragmentation et éjection des ions fragments, peuvent être répétées plusieurs fois permettant ainsi d'effectuer des analyses  $MS^{n>2}$  (Figure 7). Ce mode de fragmentation induit par la collision (CID – (Collision Induced Dissociation)) sera détaillé ultérieurement avec un autre mode de fragmentation, la fragmentation ETD (Electron Transfer Dissociation).

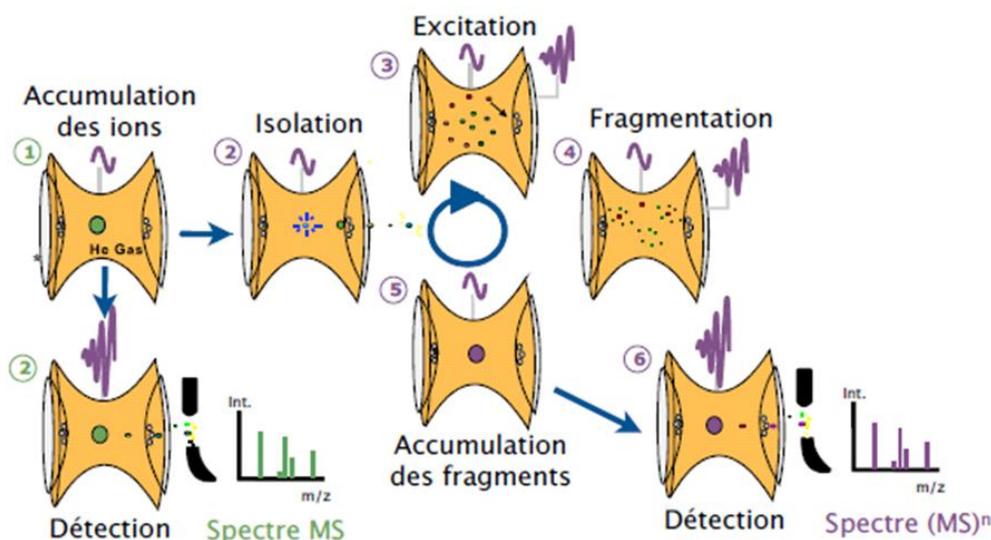


Figure 7 : Illustration de l'isolation, la fragmentation et l'éjection dans une trappe pour des analyses MS, MS/MS et MS<sup>n</sup> (d'après [34]).

Les analyseurs trappe d'ions possèdent généralement une gamme de  $m/z$  restreinte (3000  $m/z$ ) et un faible pouvoir résolutif ; mais celui-ci s'est amélioré avec les instruments de dernière génération (résolution de 6000 à 8000). D'autre part, cet analyseur a l'avantage d'avoir une vitesse d'acquisition élevée et une très bonne sensibilité. Cette rapidité d'acquisition constitue un avantage considérable dans le cadre d'une analyse LC-MS/MS en protéomique. En effet, elle permet l'acquisition d'un nombre plus important de spectres MS/MS au cours de l'élution des peptides dans un pic chromatographique et par conséquent elle permet d'identifier un nombre plus important de peptides.

### 1.2.3. Analyseur à temps de vol

Le premier spectromètre de masse à temps de vol (TOF) commercial est apparu à la fin des années 1950 et sa conception était basée sur celle du spectromètre de Wiley et McLaren publiée en 1955 [35]. Cependant, c'est avec l'apparition des sources ESI et MALDI que cet analyseur est devenu intéressant pour l'analyse des biomolécules. Dans un analyseur TOF, les ions sont accélérés dans une première zone de longueur « s » à l'aide d'une différence de potentiel « E ». Ils traverseront ensuite une deuxième zone libre de champ (de longueur « d ») et sous vide poussé ( $10^{-9} - 10^{-10}$  bar) avant d'atteindre le détecteur. Ayant reçu la même énergie cinétique au départ, les ions mettront plus ou moins de temps à traverser la zone libre de champ en fonction de leur  $m/z$ . En mesurant le temps de vol « t » de chaque ion, on peut déterminer son rapport  $m/z$  suivant l'équation :

$$m/z = 2eEs \left( \frac{t}{d} \right)^2$$

La composante orthogonale de la vitesse des ions en sortie de l'interface étant idéalement nulle, la vitesse d'un ion dans le tube de vol dépendra uniquement de l'énergie cinétique acquise dans la zone d'accélération [36]. Dans la pratique, les ions n'arrivent pas parfaitement focalisés dans cette zone d'accélération ce qui induit une dispersion en énergie

cinétique traduite par une perte considérable de résolution. Cette dispersion en énergie cinétique peut être corrigée grâce à la mise en place d'un réflecteur électrostatique, le « reflectron ». Ce dispositif est constitué d'une série de lentilles annulaires sur lesquelles sont appliqués des potentiels croissants de la même polarité que les ions. Elles font donc office de miroir électrostatique ; lorsque les ions pénètrent dans le champ électrique du réflecteur, ils sont d'abord ralentis puis ré-accelérés en sens opposé vers le détecteur. La profondeur de la pénétration au sein du réflecteur dépend de l'énergie cinétique des ions. Ainsi des ions de même  $m/z$  mais ayant une énergie cinétique différente en sortie de la zone d'accélération, seront refocalisés temporellement de façon à atteindre le détecteur au même moment. Ce dispositif permet donc d'augmenter considérablement la résolution des TOFs [37].

Lorsqu'un TOF est couplé à une source MALDI (Figure 8), le tube de vol est dans l'axe de la source. Les dispersions spatiale et temporelle des ions dans la zone d'accélération peuvent être limitées en utilisant le principe de l'extraction retardée « Delay Extraction » en plus de l'utilisation du réflecteur. Cette extraction retardée est réalisée en ajoutant une lentille supplémentaire entre la source MALDI et l'analyseur. L'application d'un potentiel à cette lentille permet de freiner un court instant les ions formés après le tir laser. Les ions sont alors focalisés et auront un départ synchronisé pour pénétrer dans le tube de vol. Dans un instrument MALDI-TOF, la fragmentation des ions est possible dans la source (« In Source Decay »). Une cellule « LIFT » est introduite avant le tube de vol pour trier les ions parents et leurs fragments métastables formés après la source. Ce tri est effectué à l'aide de leur temps de vol dans cette cellule LIFT avant leur pénétration dans le tube de vol. L'ion parent et ses fragments sont alors réaccélérés et un spectre MS/MS peut être enregistré. On parle alors de MALDI-TOF/TOF.

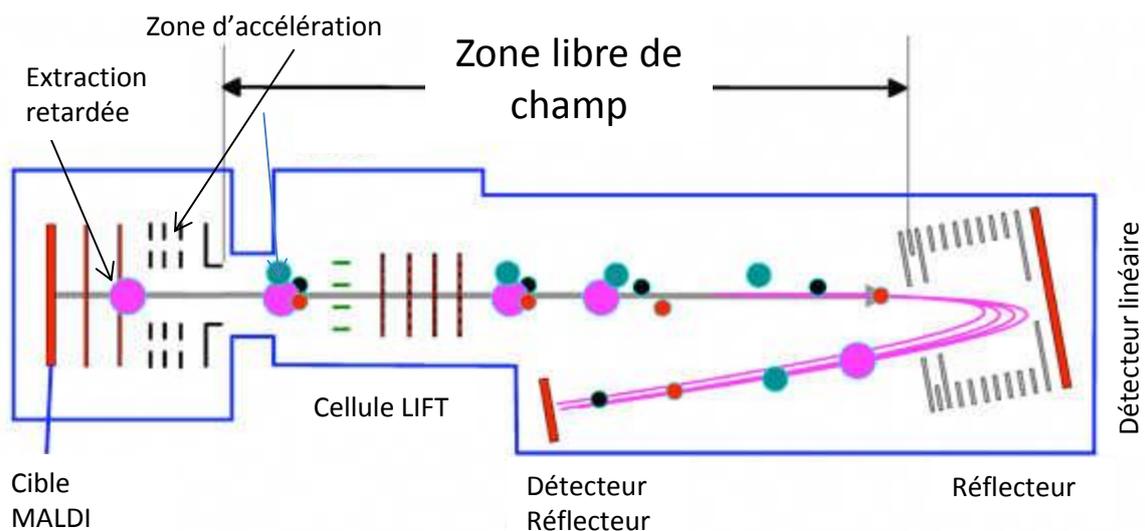


Figure 8 : MALDI-TOF/TOF avec un système d'extraction retardé, une cellule LIFT et un réflecteur électrostatique.

Lorsqu'il est couplé à une source à injection continue comme l'ESI, le tube de vol est en position orthogonal par rapport au trajet des ions dans l'interface entre la source et

l'analyseur [38,39]. En effet, les ions arrivant en flux continu parallèlement à l'axe de l'interface, une accélération perpendiculaire permet d'échantillonner les ions par paquets pour chaque cycle d'analyse du TOF et de définir précisément le temps de départ des ions. Ceux-ci sont généralement refocalisés avant d'être accélérés. Pour pouvoir effectuer de la spectrométrie de masse en tandem (MS/MS), le TOF est couplé à un autre analyseur, un quadripôle. Dans ce type de spectromètre de masse hybride appelé Q-TOF, le quadripôle peut fonctionner soit comme guide d'ions, soit comme analyseur :

- En mode guide d'ions, un balayage de l'amplitude de la radiofréquence du quadripôle permet de transférer les ions vers le TOF qui effectuera la mesure de masse. Cependant, la transmission des ions dans le quadripôle reste dépendante de l'amplitude de la tension oscillante appliquée sur les barres. La transmission de différentes gammes de masse peut donc être favorisée.
- En mode analyseur, un ion peut être sélectionné dans le quadripôle et transmis vers la cellule de collision où il est fragmenté. Les ions fragments sont ensuite transférés dans l'analyseur TOF qui mesure leur  $m/z$ .

La figure 9 présente à titre d'exemple le spectromètre de masse ESI-Q-TOF « MaXis » de la société Bruker. La source ESI est désaxée par rapport à l'entrée du spectromètre ce qui permet d'éliminer les molécules neutres. L'interface de transfert est composée de deux guides d'ions de type « Ion Funnel » et d'un multipôle. Les « ions funnels » sont des guides d'ions en forme d'entonnoir constitués par des lentilles annulaires dont le diamètre interne diminue progressivement dans le sens de progression des ions [40,41]. Le quadripôle permet soit de transférer les ions vers le TOF soit de sélectionner un ion pour le fragmenter dans la cellule de collision. Lors de la fragmentation, un gaz de collision ( $N_2$  ou Ar) est injecté dans la cellule de collision, et celui-ci, en entrant en collision avec les ions transforme leur énergie cinétique en énergie interne provoquant leur dissociation. Après la cellule de collision se trouve aussi un multipôle qui permet de thermaliser (refroidir) les ions et de les refocaliser avant de les injecter dans l'analyseur TOF. L'accélérateur orthogonal ou « Pusher » envoie les ions par paquets dans le tube de vol vers le réflecteur.

Ce type d'instrument permet d'obtenir des mesures de masse avec une haute résolution (de l'ordre de 60 000 pour les plus récents) et une très bonne précision (1 à 5 ppm). Un système « Lock Mass » de correction de la courbe d'étalonnage est utilisé pour améliorer cette précision. Ces systèmes consistent à diffuser des composés volatils qui servent d'étalons et permettent de corriger la dérive de la courbe d'étalonnage du TOF dans le temps. La haute résolution permet de faciliter la détermination des états de charges et l'attribution du pic monoisotopique des ions. Ceci, combiné à la justesse de la mesure de masse, constitue un avantage important lors de l'interprétation des résultats d'identification et permet de diminuer le risque d'identifier des faux positifs lors de l'analyse protéomique (voir le chapitre 2 de cette partie I).

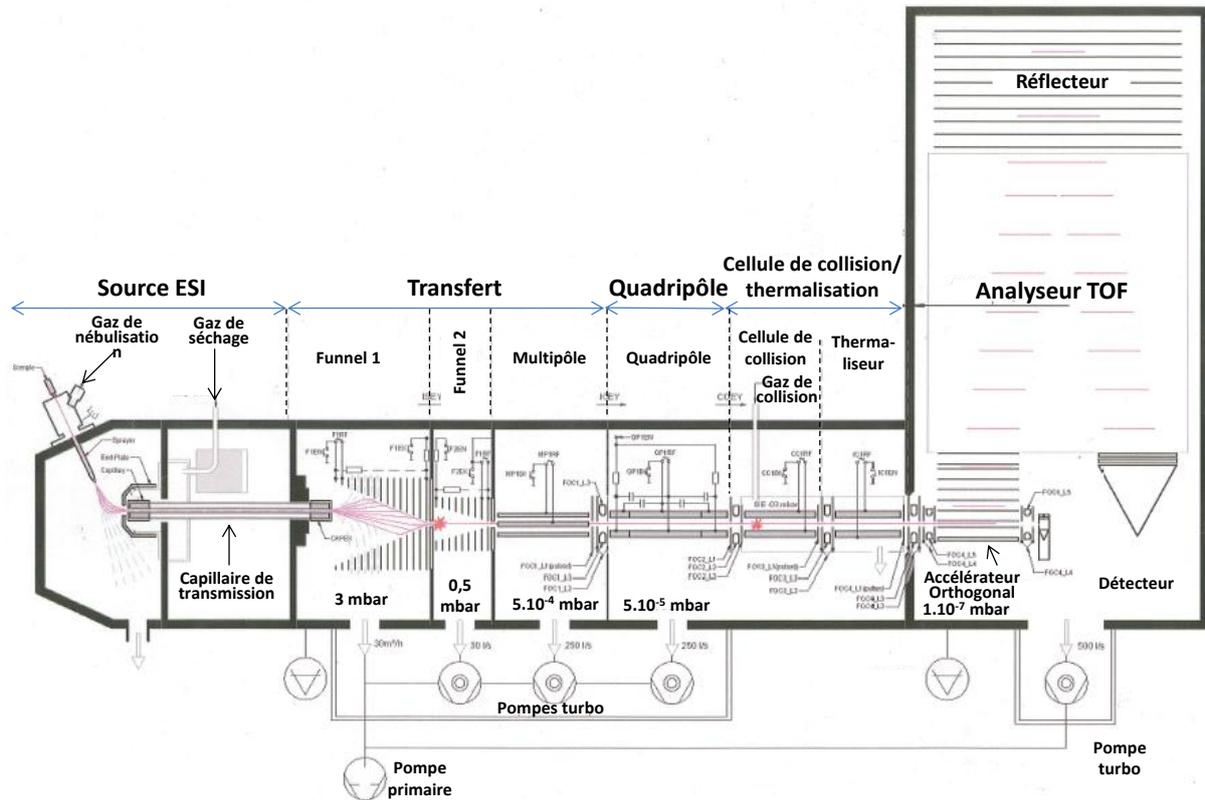


Figure 9 : Schéma du spectromètre ESI-Q-TOF MaXis de Bruker.

#### 1.2.4. Les analyseurs à transformée de Fourier

Actuellement, il existe deux types d'analyseurs à transformée de Fourier. Le premier, qui date des années 1970 est l'analyseur à résonance cyclotronique des ions à transformée de Fourier (FT-ICR) et le deuxième, beaucoup plus récent (années 2000) est l'Orbitrap™. Je décrirai brièvement ces deux analyseurs dans la suite de ce manuscrit.

##### 1.2.4.1. Analyseur à Résonance Cyclotronique des Ions à Transformée de Fourier (FT-ICR)

La technique ICR-MS a été publiée la première fois dans les années 1950 [42,43]. Cependant, c'est avec l'application des méthodes à transformée de Fourier [44] par Marshall et Comisarow dans les années 1970 que cet analyseur aux capacités résolutes impressionnantes (1 million) a connu son essor [45]. Dans une FT-ICR, les ions piègent les ions dans une cellule dont la température est proche du zéro absolu et où règne un vide poussé ( $10^{-10}$  mbar) et un champ magnétique uniforme. Ce champ, assuré par un aimant supraconducteur (5 à 20 Tesla) refroidi à l'hélium liquide, impose aux ions une trajectoire cyclotronique. L'application d'un champ électrique balayant une gamme de fréquences identique à celle de rotations des ions rend leur mouvement cohérent. Des plaques de piégeage permettent d'éviter leur sortie de la cellule et d'assurer leur piégeage (Figure 10).

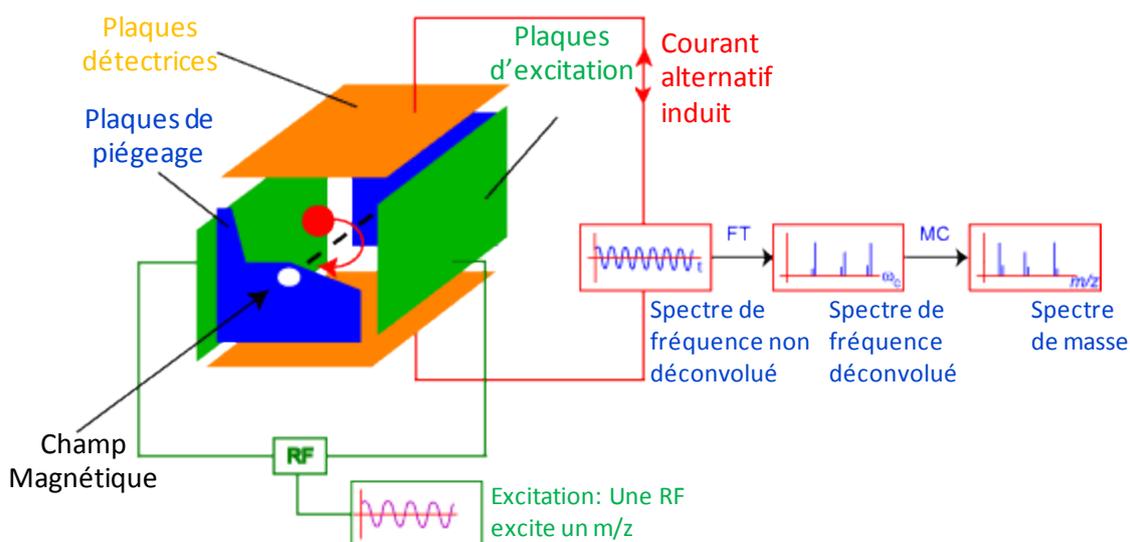


Figure 10 : Schéma d'un analyseur FT-ICR montrant le piégeage des ions, la détection et la déconvolution du signal par Transformée de Fourier (adaptée de [46])

La fréquence de rotation des ions dépend de leur  $m/z$ . Chaque ion ( $m/z$ ) est excité par un balayage d'une impulsion radiofréquence (RF) à travers les plaques d'excitation. Chaque fréquence d'excitation individuelle se couple au mouvement de l'ion et l'excite vers une orbite supérieure en le faisant entrer en résonance. Ce mouvement induira un courant électrique alternatif entre les plaques. La fréquence de ce courant est égale à celle de rotation des ions et son intensité est proportionnelle à leur quantité. Lorsque la valeur de la RF passe à celle correspondant au  $m/z$  suivant, les ions sont relaxés et reviennent à leur orbite naturelle et le paquet suivant est ensuite excité. Le résultat de ces balayages de RF est l'enregistrement d'un spectre complexe de fréquences en fonction du temps. Celui-ci est déconvolué à l'aide d'une transformée de Fourier pour obtenir un spectre de masse. Comme les ions ne sont pas détruits par ce mode de détection, ils peuvent être réanalysés et fragmentés pour faire de la MS en tandem.

#### 1.2.4.2. Analyseur Orbitrap™

L'Orbitrap™ est un analyseur de masse basé sur un nouveau principe physique et commercialisé par la société Thermo Scientific depuis le début des années 2000. Il fut inventé par Alexander Makarov [47]. L'orbitrap se compose d'une électrode externe creuse à l'intérieur de laquelle est placée d'une façon coaxiale une électrode en forme de fuseau ce qui permet l'application d'un champ électrostatique quadro-logarithmique (Figure 11). Les ions tournent autour de l'électrode interne et oscillent d'une manière harmonique selon l'axe longitudinal avec une fréquence caractéristique de leur  $m/z$  [48]. La détection et la déconvolution du signal se font de la même façon qu'avec un FT-ICR. Dans la pratique et pour la spectrométrie de masse, l'orbitrap est associé à une trappe linéaire en amont qui permet d'introduire dans l'orbitrap les paquets d'ions à analyser et aussi de sélectionner et de fragmenter les ions pour les analyses en MS/MS. Cet analyseur permet d'obtenir une résolution très élevée (de l'ordre de 100000) et une gamme dynamique de l'ordre de trois

décades. Ceci lui permet de rivaliser avec le FT-ICR avec l'avantage de ne pas nécessiter l'utilisation d'un aimant supraconducteur difficile à entretenir.

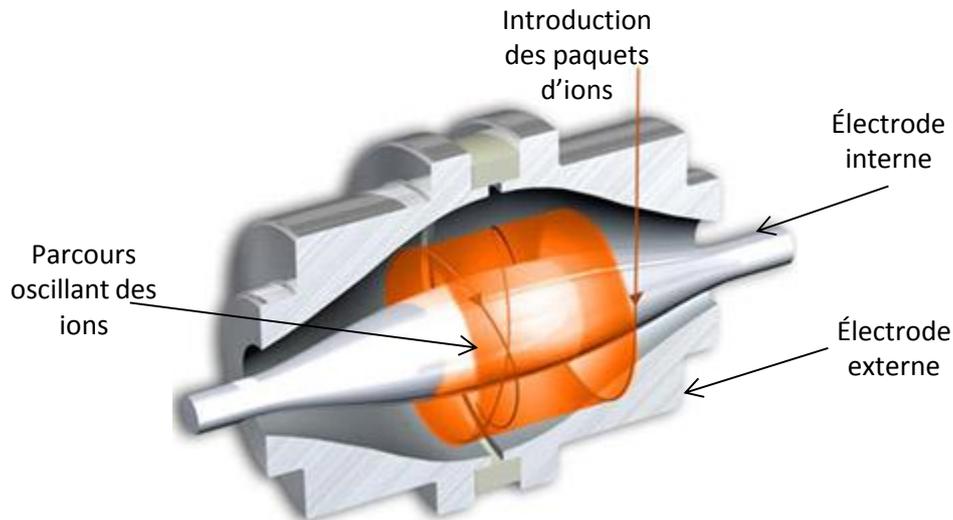


Figure 11 : Mouvement oscillant des ions autour de l'électrode interne de l'Orbitrap (adaptée de [49])

### 1.3. Les détecteurs

Il existe plusieurs types de détecteurs utilisés en spectrométrie de masse. Je me limiterai à décrire trois types de détecteurs : le multiplicateur électrostatique souvent utilisé avec les trappes d'ions, le détecteur MCP (MultiChannel Plate) et le détecteur MagneTOF généralement utilisés avec les spectromètres de masse à temps de vol.

#### 1.3.1. Le détecteur multiplicateur électrostatique

Un multiplicateur électrostatique est un détecteur formé de plusieurs dynodes de conversion. Une dynode consiste en une électrode servant de substrat à un film de matériau sélectionné pour ses capacités d'émission secondaire. Les ions énergétiques tapent sur la surface de la cathode de conversion et provoquent l'émission d'électrons secondaires. Ces derniers sont accélérés et dirigés sur une deuxième dynode. L'impact électron-dynode provoque une nouvelle émission électronique. Les électrons issus de cette dynode sont dirigés vers la dynode suivante et ainsi de suite. Le processus d'amplification continue ainsi jusqu'à la dernière dynode. Les courants atteints sont alors exploitables et seront interprétés par un traitement informatique.

#### 1.3.2. Le détecteur MCP (MicroChannel Plate)

Le détecteur MCP est un disque d'une épaisseur de l'ordre du millimètre traversé d'un côté à l'autre par des canaux parallèles ayant un diamètre de l'ordre du micron et dont la surface interne est une couche semi-conductrice [50,51]. Les ions pénètrent dans ces canaux et s'écrasent sur leurs surfaces. L'impact d'un ion provoque l'émission d'électrons secondaires qui par collisions successives avec la surface interne du canal vont permettre d'amplifier le signal qui sera finalement collecté par une anode pour être ensuite digitalisé.

### **1.3.3. Le détecteur magneTOF**

Ce type de détecteur est de plus en plus utilisé dans les spectromètres de masse à temps de vol. Il consiste en une surface plane capable elle aussi d'émettre des électrons secondaires suite à l'impact des ions. Ces électrons sont ensuite conduits et focalisés grâce à des potentiels électrostatiques déformés par un champ magnétique vers la surface d'une dynode de conversion. Un multiplicateur électrostatique permet alors d'amplifier le signal électronique qui sera ensuite digitalisé.

## **1.4. La fragmentation peptidique**

L'analyse en spectrométrie de masse en tandem MS/MS des peptides permet d'accéder à la séquence d'acides aminés de ces peptides et par la suite d'identifier les protéines correspondantes par recherche dans des banques de données. Les spectres MS/MS sont en effet obtenus par la mesure de masse des ions issus de la fragmentation des peptides. Il existe différents modes d'acquisition de données MS/MS tel que le mode d'acquisition de données dépendant DDA (le plus utilisé) et le mode d'acquisition de données indépendant (DIA). Ces modes d'acquisitions seront évoqués dans le chapitre suivant de ce manuscrit.

Classiquement, en couplage chromatographie liquide – spectrométrie de masse (LC-MS), et indépendamment de l'analyseur utilisé, l'obtention d'un spectre MS/MS informatif en mode DDA passe par deux étapes. La première consiste en l'analyse MS des peptides élués de la LC à l'instant  $t$  et la sélection d'un nombre d'ions à fragmenter. La deuxième étape consiste à isoler le premier ion précurseur sélectionné, puis à le fragmenter et enregistrer son spectre MS/MS. Cette deuxième étape est répétée autant de fois que le nombre d'ions précurseurs sélectionnés lors de la première étape avant de recommencer le cycle.

Même si l'interprétation des spectres MS/MS est effectuée à l'aide d'algorithmes spécialisés, elle est souvent complexe et demande une bonne connaissance des modes et des mécanismes de fragmentation des peptides en phase gazeuse.

### **1.4.1. La nomenclature des fragmentations peptidiques**

La nomenclature des ions fragments des peptides servant toujours de référence est celle proposée par Biemann [52] au début des années 1990. Cette nomenclature sépare les ions fragments en trois familles, les deux premières sont issues de fragmentations à basse énergie tandis que la troisième est issue de fragmentations à haute énergie :

- Ceux qui contiennent la partie N-terminale du peptide précurseur : la série a, b et c ;
- Ceux qui contiennent la partie C-terminale du peptide précurseur : la série x, y et z ;
- Ceux issus de la fragmentation des chaînes latérales des acides aminés : la série d, v et w.

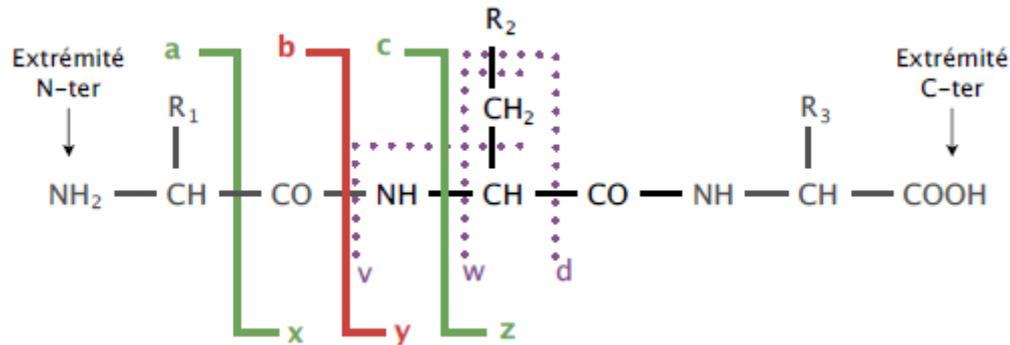


Figure 12 : Nomenclature des fragmentations peptidiques selon Biemann. R<sub>i</sub> représente les chaînes latérales des acides aminés.

#### 1.4.2. Le mode de fragmentation CID

La fragmentation induite par collision CID aussi appelée CAD (Collision Activated Dissociation) est le mode de fragmentation le plus utilisé en analyse protéomique. Le transfert des ions en phase gazeuse par ESI n'étant pas un processus à haute énergie, l'énergie interne de ces ions est faible [53]. Leur fragmentation est induite en favorisant leurs collisions avec un gaz inerte (N<sub>2</sub>, Ar). On peut distinguer une fragmentation basse énergie (de l'ordre de quelque eV obtenue avec les instruments de type Q-TOF et Trappe Ionique) et une fragmentation haute énergie (quelques KeV obtenus avec les analyseurs TOF/TOF). En général, la fragmentation CID des peptides génère majoritairement des fragments b et y observables à basse et à haute énergies. Ces fragments représentent la rupture des liaisons peptidiques qui sont plus labiles que les autres liaisons. Dans ce mode de fragmentation, l'énergie cinétique des ions est transformée lors des collisions en énergie interne. Celle-ci est répartie d'une manière uniforme sur tout le squelette peptidique, ce qui favorise la rupture des liaisons les plus labiles.

##### 1.4.2.1. Mécanisme de fragmentation

Les peptides protonés qui sont activés avec une faible énergie de collision fragmentent généralement par des réactions dirigées par la charge (« charge-directed fragmentation ») [54]. Le modèle le plus admis actuellement et servant à décrire comment les peptides protonés sont fragmentés est le modèle du proton mobile qui fut établi suite aux études effectuées par les groupes de Wysocki [55-58], Harrison [59] et Gaskell [54]. Les peptides étant formés d'acides aminés portant différentes fonctions chimiques, ils peuvent être protonés sur différents sites. Ces sites peuvent être :

- Très favorables à la protonation tels que l'amine N-terminale et les fonctions basiques des chaînes latérales ;
- Moyennement favorables tels que les atomes d'oxygène des liaisons amides ;
- Peu favorables tels que les atomes d'azote des liaisons amides.

Cependant, les collisions avec les molécules de gaz vont activer les peptides et provoquer leur excitation faisant migrer les protons des sites favorables vers des sites moins favorables tels que les atomes d'azote des liaisons amides. La liaison amide est une liaison

forte du fait de la stabilisation par délocalisation de charge entre les doublets de la fonction carbonyle et ceux de l'azote. Cependant, la protonation de cette liaison abolit cette délocalisation de charge [60] et conduit à l'affaiblissement de la liaison amide. Elle fait ainsi du carbone du groupement adjacent une cible possible pour une attaque nucléophile ce qui initie la fragmentation au niveau de la liaison peptidique. Cette fragmentation entraîne la formation d'ions fragment b et y [61] (Figure 13). Les fragments y retiennent la charge sur la chaîne latérale de l'acide aminé C-terminal ; ils sont souvent majoritaires car ils sont plus résistants que les ions b aux décompositions qui peuvent suivre.

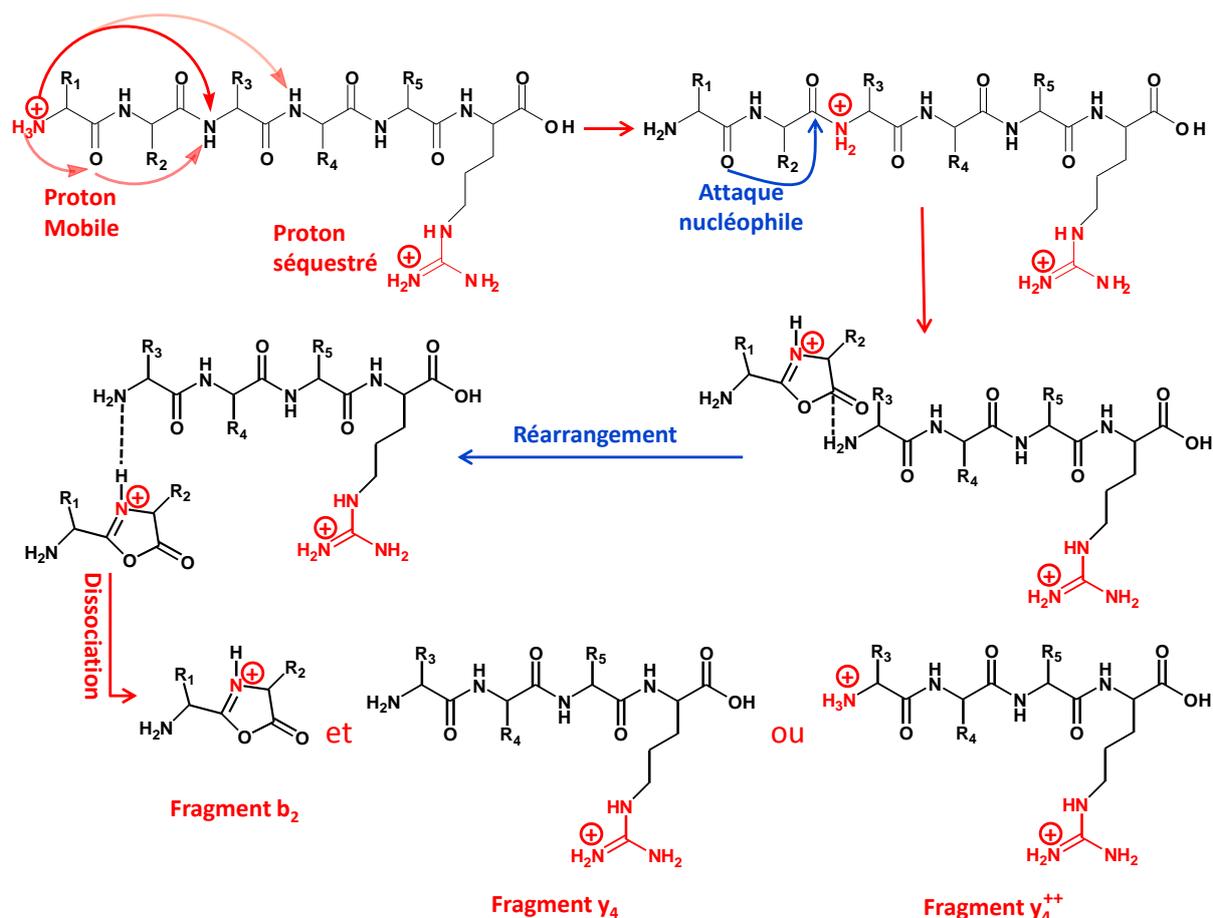


Figure 13 : Fragmentation d'un peptide tryptique bi-protoné selon le modèle du proton mobile.

L'énergie nécessaire pour mobiliser le(s) proton(s) est fonction de la composition en acides aminés (surtout les résidus basiques) et de la séquence du peptide [58]. Un proton séquestré sur un groupement basique est plus difficile à délocaliser. L'énergie nécessaire à sa délocalisation augmente de manière proportionnelle avec la basicité du peptide en phase gazeuse (Figure 14). Si les protons sont très fortement séquestrés, les fragmentations distantes de la charge deviennent très compétitives [62].

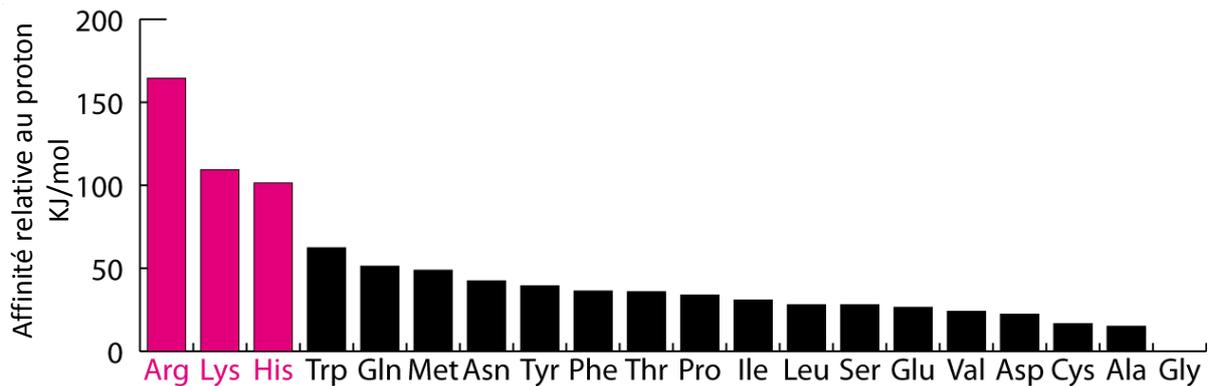


Figure 14 : Affinités au proton en phase gazeuse des vingt acides aminés par rapport à la glycine déterminées expérimentalement [63].

La majorité des fragments peptidiques obtenus par fragmentation CID peuvent être expliqués par le modèle du proton mobile. Les exceptions les plus importantes sont les coupures préférentielles aux résidus acides [64] et aux résidus prolines [65]. Les résidus acides présentent une fragmentation préférentielle en C-terminal quand le peptide n'a pas de proton mobile (fragmentation distante de la charge). Les résidus proline présentent une fragmentation préférentielle en N-terminal due à la basicité de l'azote de la proline qui favorise la séquestration du proton mobile. La fragmentation est difficile en C-terminal quand le peptide présente un proton mobile due à l'encombrement stérique.

Enfin, de nombreuses équipes travaillent sur l'établissement de règles de fragmentation précises. Ces règles sont implémentées dans les algorithmes de recherche et d'interprétation des spectres MS/MS afin d'améliorer l'efficacité et la spécificité de ces algorithmes [66,67].

#### 1.4.3. Le mode de fragmentation par transfert d'électron (ETD)

La fragmentation ETD (Electron Transfer Dissociation) est une technique de fragmentation comparable à l'ECD (Electron Capture Dissociation) et introduite en 2004 [68]. Le transfert de l'électron est assuré par l'intermédiaire d'un anion radicalaire (fluoranthène, anthracène ...) qui joue le rôle de transporteur d'électrons vers le peptide protoné, contrairement à l'ECD qui emploie des électrons libres à basse énergie.

Due à sa nature non ergodique, la fragmentation engendrée n'a pas de coupure préférentielle. Le site de coupure est aléatoire et les liaisons faibles peuvent être préservées ce qui rend ce mode de fragmentation intéressant pour l'étude de modifications post-traductionnelles telles que la phosphorylation qui, en mode CID, n'est souvent pas conservée. On observe souvent des fragments c et z avec l'ETD lors de la fragmentation de peptides tryptiques. Lorsqu'une N-protéase qui clive les peptides du côté N-terminal d'un résidu basique (telle que la Lys-N) est employée, on n'observe généralement que des fragments c, ce qui facilite l'interprétation des spectres surtout pour le séquençage *de novo* [69,70]. Ceci est dû au fait que les peptides générés seront préférentiellement bi-protonés du côté N-terminal (un proton sur l'amine N-terminale et un proton sur l'amine de la chaîne latérale de la lysine). L'ETD permet aussi, par rapport au mode CID, de fragmenter des

peptides de taille plus élevée voir des protéines entières. C'est un mode de fragmentation complémentaire à la CID et peut être alterné avec celle-ci lors d'une analyse LC-MS/MS sur les spectromètres les plus récents ouvrant ainsi l'accès à plus d'informations de séquence.

## 2. Les techniques de séparation des protéines et des peptides

L'analyse protéomique d'échantillons complexes n'est pas possible sans une étape de dé-complexification de l'échantillon en amont de la spectrométrie de masse. La raison principale est qu'un échantillon biologique de protéines (par exemple un lysat cellulaire) peut contenir généralement plusieurs centaines à plusieurs milliers de protéines différentes. S'ajoutent à cette complexité celles induites par les différentes isoformes, les modifications post-traductionnelles et les concentrations des protéines qui peuvent s'étendre sur de larges gammes dynamiques. Le développement des techniques séparatives a constitué une étape clé dans l'essor de l'analyse protéomique par spectrométrie de masse.

La dé-complexification de l'échantillon peut être réalisée soit au niveau des protéines, soit au niveau des peptides générés par digestion enzymatique des protéines soit aux deux niveaux combinés. Ici, je décrirai les principales techniques de séparation utilisées en analyse protéomique.

### 2.1. La séparation au niveau des protéines

Les techniques de séparation des protéines peuvent être classées en deux grandes catégories : Les techniques de séparation sur gels d'électrophorèse et les techniques de séparation sans gel souvent appelées « Gel-free ». Dans ce qui suit, je décris brièvement ces techniques.

#### 2.1.1. L'électrophorèse sur gel monodimensionnelle 1D SDS PAGE (Sodium DodecylSulfate PolyAcrylamide Gel Electrophoresis)

La technique de l'électrophorèse est fondée sur la migration des protéines sous l'effet d'un champ électrique dans un milieu adapté. L'électrophorèse monodimensionnelle des protéines est généralement effectuée sur des gels de polyacrylamide en utilisant le dodécylsulfate de sodium (SDS) en tant que détergent. Cette technique a été développée par Ulrich Laemmli dans les années 1970 [71] et permet de séparer les protéines en fonction de leur masse moléculaire. Le détergent SDS se fixe sur les résidus hydrophobes des protéines et favorise leur solubilisation. En milieu légèrement basique, il confère une charge négative aux protéines. La densité de charge étant sensiblement constante pour toutes les protéines, la migration se fait essentiellement en fonction de leurs masses moléculaires. La taille des protéines varie fortement dans un échantillon, il est donc souvent nécessaire d'optimiser la résolution en fonction des besoins expérimentaux [72] et ceci en optimisant le maillage du gel qui dépend du pourcentage d'acrylamide/bis-acrylamide. Plus ce pourcentage est élevé, plus les mailles du gel sont petites et plus la résistance à la migration des protéines est accentuée. Des gels en gradient d'acrylamide peuvent aussi être utilisés afin d'optimiser la résolution de séparation de protéines de toutes tailles. Un gel de concentration de faible maillage est souvent mis en amont du gel de séparation afin de créer un front de migration

dans lequel les protéines sont concentrées avant leur pénétration dans le gel de séparation de maillage plus serré.

### *2.1.2. L'électrophorèse bidimensionnelle sur gel (gel 2 D)*

L'électrophorèse bidimensionnelle sur gel [73] consiste à utiliser deux dimensions de séparation orthogonales. La première dimension est l'électrofocalisation qui permet de séparer les protéines en fonction de leur point isoélectrique pI. Elle consiste à utiliser une bandelette de gel présentant un gradient de pH et soumise à un champ électrique orienté de sa borne acide vers sa borne basique [74]. Les protéines migrent alors sur la bandelette et s'immobilisent au niveau du pH qui correspond à leur pI. La bandelette est ensuite saturée en SDS puis chargée en haut d'un Gel SDS PAGE. C'est la réalisation de la deuxième dimension qui permet de séparer les protéines en fonction de la masse moléculaire comme dans un gel 1D. Cette technique possède de bonnes capacités résolutive et jusqu'à 5000 protéines peuvent être séparées. La coloration du gel permet de visualiser sous forme de spots les protéines qui ont migré. L'intensité de la coloration du spot permet d'avoir une indication quantitative de l'abondance de la protéine. Jusqu'à 1 ng de protéines peuvent être détectées [75]. Les gels 2D permettent aussi la séparation d'isoformes de protéines (formes plus longues ou tronquées ayant des masses moléculaires différentes) et la visualisation de modifications post-traductionnelles telles que la glycosylation, la phosphorylation, etc (souvent visibles sous forme de chapelets aux mêmes poids moléculaires) [72]. Les spots d'intérêt sont découpés puis les protéines sont digérées avec une protéase et les peptides ainsi générés sont analysés par MS ou MS/MS (MALDI-TOF ou TOF/TOF ou LC-ESI-MS/MS). Toutefois cette technique présente certaines limitations, les expériences de Gel 2D sont difficiles à automatiser et la réalisation de répliquats techniques est indispensable pour vérifier la reproductibilité et la cohérence des résultats. De plus, la gamme dynamique visible sur le gel est restreinte et certaines catégories de protéines telles que les protéines membranaires et les protéines aux pI extrêmes ne sont pas compatibles avec cette technique [76,77]. Cependant de nouveaux systèmes utilisant des gels pré-coulés permettent maintenant d'automatiser les expériences de gel 2D.

Plusieurs types de coloration des protéines dans un gel existent. La sensibilité de détection en dépend considérablement. Trois types de coloration sont couramment utilisés en analyse protéomique :

- La coloration au bleu de Coomassie : C'est un colorant bleu qui se lie aux acides aminés basiques et aux aromatiques. Il a l'avantage de présenter une linéarité et une réversibilité qui le rend parfaitement adapté aux études de protéomique comparative. Elle présente une limite de détection de l'ordre du ng [78].
- La coloration à l'argent : Celle-ci utilise le nitrate d'argent pour colorer les protéines dans le gel et elle présente une limite de détection de 0,5 à 0,1 ng de protéines [79]. Cependant, la coloration n'est pas linéaire pour les fortes concentrations et des incompatibilités avec l'analyse par spectrométrie de masse ont été décrites [80],

mais des protocoles améliorés ont pu être mis au point [81] et il existe néanmoins des techniques de coloration à l'argent compatible avec la spectrométrie de masse.

- Les colorations fluorescentes : Celles-ci consistent à marquer les protéines avec des fluorochromes avant leur séparation sur gel [82]. Elles permettent d'atteindre des sensibilités de l'ordre de l'attomole. Dans le cadre d'une étude comparative, on parle de Gel 2D DiGE (Differential Gel Electrophoresis). Celle-ci consiste à mélanger deux échantillons de protéines marqués par des fluorochromes différents avant de faire migrer le mélange sur le même gel. Elle sera détaillée dans le chapitre 3 de cette partie.

Une autre technique d'électrophorèse sur gel permet d'analyser les complexes multiprotéiques. Cette technique appelée BN-PAGE (Blue Native PolyAcrylamide Gel Electrophoresis) consiste à séparer les complexes protéiques dans des conditions non-dénaturantes dans une première dimension puis de les séparer sur une deuxième dimension SDS PAGE. Les protéines sont solubilisées à l'aide de détergents non-ioniques. L'addition du bleu de Coomassie leur confère une charge négative et permet de les faire migrer sur une première dimension de gel de polyacrylamide en utilisant des tampons adéquats qui conservent l'intégrité des complexes. Les complexes ne migrent plus en fonction de leur densité de charge mais plutôt en fonction de leur taille dans le gel de polyacrylamide. Les bandes peuvent ensuite être excisées et les complexes sont électroélus et analysés (1D-BN PAGE) ou, après la première dimension, la piste de gel est déposée sur un gel SDS PAGE classique pour une deuxième dimension dénaturante (2D BN-PAGE). Dans ce cas, les complexes sont dissociés et les sous unités sont séparées selon leur masse moléculaire dans cette deuxième dimension [83,84] (Figure 15).

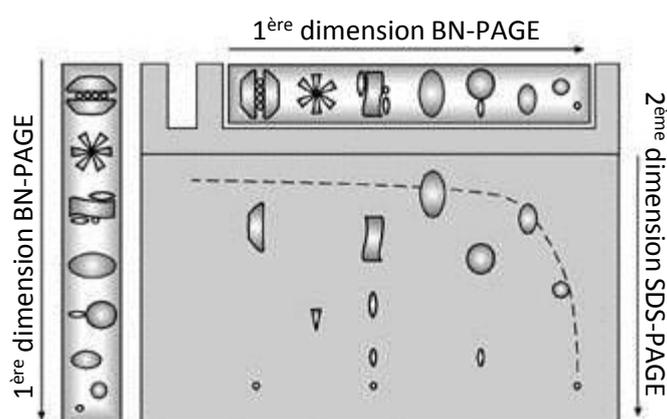


Figure 15 : Illustration du principe du 2D BN-PAGE d'après Fiala et coll. [84].

Cette technique est cependant limitée par l'insolubilité de certains complexes protéiques qui tendent à précipiter aux alentours de leur point isoélectrique, et par la taille des mailles des gels qui ne permettent pas la migration de très gros complexes.

### 2.1.3. L'électrophorèse « off gel »

Cette nouvelle technique, apparue au début des années 2000, permet de séparer les protéines selon leur point isoélectrique directement en solution, c'est « l'électrofocalisation

off gel » [85,86]. Les extraits protéiques sont déposés dans des puits qui sont sur une bande de gel à gradient de pH immobilisé. Les protéines migrent sous l'effet d'un champ électrique jusqu'à ce qu'elles atteignent le puits qui correspond à leur pI [87]. Cette technique peut être couplée à une autre dimension de séparation par gel ou par chromatographie liquide avant digestion et analyse par LC-MS/MS. Elle présente l'avantage de conserver les protéines en milieu liquide ce qui évite les pertes d'échantillon. Cependant, elle présente les mêmes limitations que la première dimension de gel 2D, à savoir les pI extrême et les protéines membranaires. Plus tard, cette technique a commencée à être utiliser aussi pour séparer les peptides [88] avant leur analyse par LC-MS/MS.

#### ***2.1.4. La séparation des protéines par chromatographie liquide***

Les techniques de chromatographie liquide peuvent être aussi utilisées pour séparer les protéines. L'utilisation de plusieurs types de chromatographie pour la séparation des protéines est décrite dans la littérature. On peut ainsi trouver la chromatographie en phase inverse [89], la chromatographie d'exclusion stérique [90] ou la chromatographie d'échange d'ions [91]. Cette dernière a aussi été utilisée en couplage avec d'autres chromatographies de sélectivités différentes [92]. Cependant, les méthodes chromatographiques pour la séparation des protéines en amont de l'analyse protéomique restent peu utilisées, ceci est lié à l'utilisation de détergents et agents chaotropes non compatibles avec la chromatographie pour la solubilisation des protéines. De plus la séparation de mélanges très complexes de protéines (Lysats cellulaires total) est peu compatible avec la chromatographie liquide.

### **2.2. La séparation au niveau peptidique**

Après la séparation ou non des protéines, une étape de digestion vient cliver ces protéines et génère une grande quantité de peptides. Ceux-ci sont souvent séparés avant leur analyse par spectrométrie de masse. L'électrofocalisation off gel peut être utilisée comme décrit dans la partie 2.1.3 de ce chapitre. Cependant, vue la diversité de leurs propriétés physico-chimiques, presque toutes les méthodes chromatographiques peuvent être utilisées pour leur séparation [93]. Les peptides sont séparés généralement sur une colonne de phase inverse C<sub>18</sub> en mode gradient avec l'acétonitrile (le plus souvent) ou le méthanol comme modifieur de la phase mobile et l'acide formique comme anions pour la formation de paire d'ions. La chromatographie liquide en phase inverse est souvent couplée directement à la source électrospray du spectromètre de masse. Ce couplage RP-HPLC-MS/MS constitue la principale technique utilisée pour l'analyse protéomique

#### ***2.2.1. Les faibles débits et les systèmes nanoHPLC (ou nanoLC) et nanoUHPLC***

Le couplage de la chromatographie liquide avec la spectrométrie de masse via une source électrospray est donc la technique de choix pour l'analyse des peptides de digestion enzymatique. Les échantillons biologiques pour l'analyse protéomique sont souvent disponibles en faible quantité. Etant donné que la sensibilité des spectromètres de masse en mode ESI dépend de la concentration des analytes et non du débit, des développements visant à miniaturiser les systèmes de chromatographie ont été réalisés. Ainsi des systèmes

de chromatographie sur des colonnes de faible diamètre (75 – 100  $\mu\text{m}$ ) permettant de travailler avec des débits très faibles (de l'ordre de 100 à 400  $\text{nl}/\text{min}$ ) ont été développés. Ces débits faibles nécessitent de travailler avec des sources nanoESI de géométrie adaptée. Pour ceci des développements au niveau des sources ont été nécessaires pour pouvoir coupler ces systèmes de chromatographie aux spectromètres de masse. On parle alors de nanoLC-MS/MS. Ces avancées permettent à ce couplage d'avoir une sensibilité de l'ordre de la basse femtomole.

Avec ces faibles débits, l'injection de volumes de quelques  $\mu\text{L}$  directement sur la colonne nécessite un temps considérable. Les peptides sont alors injectés sur une précolonne ou colonne de chargement à un débit plus élevé (de 4 à 15  $\mu\text{L}/\text{min}$ ). Ensuite, la colonne de chargement est mise en ligne avec la colonne analytique à l'aide d'une vanne à faible volume mort et l'élution est effectuée à nanodébit. Cette colonne de chargement permet non seulement de concentrer les échantillons avant leur séparation sur la colonne analytique mais aussi de les dessaler. L'efficacité et la résolution de la séparation dépendent directement du choix de la phase stationnaire. De manière générale, des colonnes de phases  $\text{C}_{18}$  avec des particules de 3 à 5  $\mu\text{m}$  et une porosité de 300  $\text{\AA}$  sont utilisées. Cependant, plusieurs développements technologiques ont visé l'augmentation de l'efficacité, la résolution et la capacité de pic des systèmes nanoLC, parmi ceux-ci:

- Des phases avec des particules de diamètre plus faible (1,7  $\mu\text{m}$ ) proposées depuis 2004 [94]: Celles-ci permettent une séparation plus efficace et une meilleure résolution mais imposent une augmentation considérable de la pression sur les systèmes (600 – 800 bar) [95,96]. Des systèmes pouvant aller à des pressions aussi élevées ont donc été développés [97] (par exemple le système nanoAcquity UPLC<sup>TM</sup> de Waters). On parle alors de Chromatographie Liquide Ultra-Haute Performance (UHPLC).
- Des phases monolithiques permettant des séparations plus efficaces des peptides [98,99].
- Des systèmes de colonnes sur puce microfluidique intégrant colonne de chargement, colonne analytique et pointe nanoESI qui permettent la réduction des volumes morts [100] (par exemple le système nanoLC-Chip-Cube<sup>TM</sup> d'Agilent Technologies [101,102] et le système TRIZAIC<sup>TM</sup> UPLC de Waters).

### ***2.2.2. La chromatographie multidimensionnelle***

Afin de s'affranchir des étapes et limitations liées à la séparation des protéines, des stratégies de couplage de plusieurs dimensions de chromatographie pour la séparation des peptides ont été développées [103]. Cette approche est appelée chromatographie multidimensionnelle [104] ou encore MudPit (Multi-Dimensional Protein Identification Technology) [105,106]. Cette stratégie consiste à fractionner les peptides issus d'une digestion enzymatique par plusieurs séparations chromatographiques (différentes phases) avant leur analyse par spectrométrie de masse en tandem. Le principe consiste à utiliser des

techniques de chromatographie basées sur des sélectivités de séparation différentes afin d'obtenir des séparations orthogonales des peptides par les différentes dimensions utilisées. Plusieurs combinaisons de dimensions chromatographiques ont été décrites. La dernière étape de fractionnement, étant toujours couplée directement à un spectromètre de masse, reste la RP-HPLC. Pour les chromatographies en amont on peut citer la chromatographie d'échange d'ion (de cations (SCX) [105-107] ou d'anions [108]), la chromatographie d'exclusion stérique [109,110], la chromatographie d'interactions hydrophiles (HILIC) [111] ou encore la chromatographie en phase inverse à pH basique [112]. Le plus souvent on utilise deux dimensions chromatographiques, celles-ci peuvent être en configuration dite « offline » ou « online » (en ligne).

Dans la configuration « offline », des fractions sont collectées lors de la première dimension chromatographique. Généralement, la collecte est automatisée à l'aide d'un robot pour permettre une meilleure reproductibilité. Les fractions collectées sont ensuite analysées en deuxième dimension. Cette deuxième dimension est couplée directement au spectromètre de masse.

Dans la configuration « online », les deux colonnes des deux dimensions sont couplées à l'aide d'un système de valves et une colonne d'enrichissement. Les peptides sont élués de la première dimension vers une colonne d'enrichissement puis transférés de façon séquentielle vers la deuxième séparation qui est couplée directement au spectromètre de masse. Le passage sur la colonne d'enrichissement permet d'éliminer les éluants de la première dimension généralement incompatibles avec la spectrométrie de masse [113].

## Chapitre 2 Les stratégies d'identification en analyse protéomique

Dans ce chapitre je présenterai les principales stratégies d'identification des protéines à l'aide des outils décrits dans le chapitre précédent. L'analyse protéomique par spectrométrie de masse fournit un ensemble de spectres MS et MS/MS qui serviront à identifier les protéines de l'échantillon étudié. L'interprétation de ces spectres passe par leur mise en relation avec les séquences protéiques correspondantes à l'aide d'outils bioinformatiques. Généralement, l'identification est réalisée à partir de données générées lors de l'analyse de peptides produits par digestion enzymatique. L'ensemble des stratégies basées sur cette approche est désigné par le terme « bottom-up proteomics ». Une autre approche impliquant l'analyse directe des protéines en MS/MS est désignée par le terme « top-down proteomics » [114]. L'approche « bottom up » reste de loin la plus utilisée surtout pour les échantillons complexes et pour les études à grande échelle [115]. En fait, l'approche « top down » reste limitée par la séparation des protéines et est essentiellement utilisée pour des échantillons peu complexes ou pour la caractérisation de protéines purifiées.

L'apparition des banques de données protéiques et leur enrichissement en séquences protéiques ainsi que les développements instrumentaux ont conduit à l'apparition de différentes stratégies pour l'identification des protéines. Ainsi, en 1993, cinq groupes ont simultanément décrit un concept appelé identification par empreinte peptidique massique PMF (« Peptide Mass Fingerprint ») [116-120]. Cette méthode consiste à identifier les protéines en mesurant la masse des peptides qu'elles génèrent suite à une digestion enzymatique et à comparer la liste des masses mesurées avec celles issues de la digestion *in silico* des protéines dans les banques de séquences. Puis, en 1994, une nouvelle technique a été décrite, permettant de corréliser les spectres MS/MS des peptides avec leurs séquences issues des banques de données [121]. La même année, Mann et coll. ont proposé une approche qui consiste à déduire des séquences partielles en acides aminés à partir des différences de masses observées entre les ions des spectres de fragmentation et d'utiliser cette séquence pour effectuer des requêtes tolérantes aux erreurs dans les banques de séquences protéiques [122]. Depuis, de nombreux algorithmes de recherche dans les banques de séquences protéiques à partir de données de spectrométrie de masse ont vu le jour.

## 1. L’empreinte peptidique massique PMF

Cette stratégie consiste à comparer la liste des masses expérimentales des peptides à l’ensemble des listes de masses calculées à partir des séquences des peptides issues de la digestion *in silico* des protéines présentes dans la banque de données (Figure 1). De nombreux moteurs de recherche PMF sont développés. On peut citer entre autres Mascot [123], MOWSE [117], MS-Fit [124], ProFound [125], etc ...

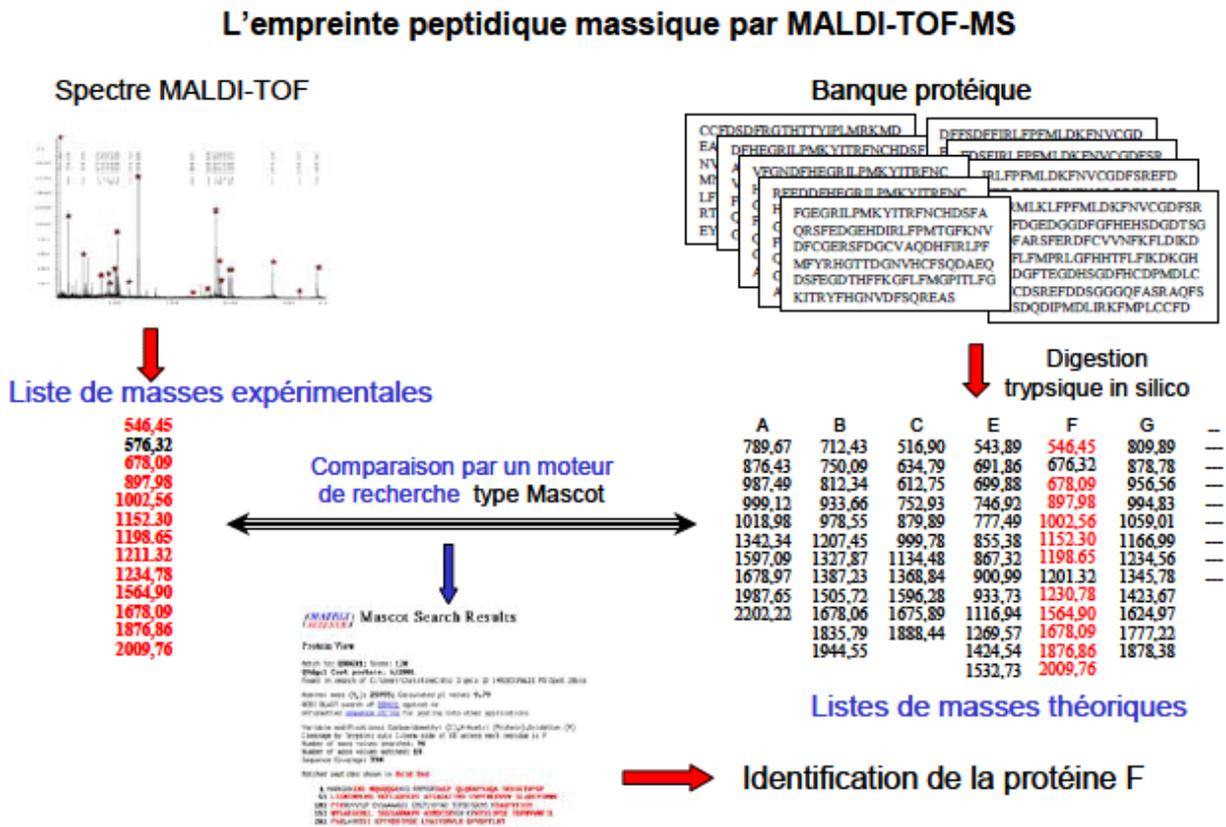


Figure 1 : Illustration de la stratégie d’identification des protéines par empreinte peptidique massique (d’après [126]).

En général, cette stratégie est utilisée pour identifier des protéines séparées par gel 2D puis digérées et analysées par MALDI-TOF-MS. Elle requière un instrument présentant une bonne précision de mesure de masse (< 10 ppm) et une bonne résolution (> 10000) [127]. L’utilisation d’instruments de haute résolution telle que l’ESI-TOF et le FT-ICR ont aussi été décrites. Les critères d’acceptation des identifications par PMF sont décrites dans les directives de publication des données de protéomique [128-130]. Elles dépendent essentiellement du nombre de peptides identifiés par protéine, des erreurs sur la mesure de masse et du nombre de pics non assignés sur les spectres. L’acceptation de résultats issus uniquement de PMF est toutefois de plus en plus restreinte. En effet, avec la croissance considérable du nombre d’entrées dans les banques de séquences protéiques, la spécificité du PMF est devenu très limitée [131] par rapport au séquençage par MS/MS des peptides [24].

## 2. Les stratégies par analyses MS/MS

Ces stratégies consistent à utiliser les données des spectres MS/MS pour réaliser l'identification des protéines. Les peptides générés après digestion enzymatique des protéines sont analysés sur un système LC-MS/MS. Trois modes d'acquisition existent :

- Mode données-dépendant (DDA): les ions à fragmenter sont sélectionnés automatiquement puis isolés et fragmentés. C'est le mode d'acquisition le plus utilisé.
- Mode données-indépendant (DIA) ou MS<sup>E</sup> : introduit récemment, aucune sélection n'est effectuée et tous les ions sont fragmentés en même temps.
- Sélection manuel : les ions à fragmenter sont sélectionnés manuellement, isolés puis fragmentés et le spectre MS/MS est enregistré. Ce mode est peu utilisé en analyse protéomique.

Les spectres MS et MS/MS obtenus sont ensuite interprétés afin d'obtenir des informations sur les séquences des peptides permettant de remonter aux protéines correspondantes.

### 2.1. Identification des protéines en mode d'acquisition données-dépendants (DDA)

Ce mode d'acquisition consiste à paramétrer le spectromètre de masse de façon à ce qu'il effectue automatiquement la sélection, l'isolation et la fragmentation des ions. Ces actions sont effectuées par cycles alternant MS et MS/MS. Un cycle commence par l'acquisition d'un spectre MS suivie par la sélection d'un certain nombre d'ions à fragmenter. Ces ions sont ensuite isolés et fragmentés un par un et le spectre MS/MS de chacun est enregistré avant le début d'un nouveau cycle. La sélection des ions à fragmenter est réalisée sans intervention directe de l'utilisateur. Les paramètres de cette sélection sont préalablement définis par l'utilisateur et sont entre autres:

- Le nombre d'ions à sélectionner sur chaque spectre MS (normalement les n ions les plus intenses) ;
- L'état de charge préférentiel des ions à sélectionner ;
- Le seuil d'intensité qu'un ion doit avoir pour être sélectionné ;
- Un temps d'exclusion pour les ions qui ont déjà été sélectionnés (afin d'éviter la redondance dans la sélection des ions).

Il est également possible de paramétrer des listes d'inclusion d'ions recherchés spécifiquement afin de forcer leur sélection pour la MS/MS ou des listes d'exclusion afin d'éviter la sélection de certains ions. Ces listes d'exclusion peuvent être intéressantes dans le cas de répétition de l'analyse LC-MS/MS du même digest. Ainsi, les ions qui ont conduit à des identifications de peptides lors de la première analyse sont exclus dans la deuxième. Cette méthode permet donc de favoriser la sélection d'autres ions et d'augmenter le nombre d'identification.

Le mode DDA reste à ce jour le mode le plus utilisé en analyse protéomique par LC-MS/MS. Après la fin de l'analyse, des fichiers contenant l'ensemble des informations concernant les spectres, les chromatogrammes et les temps de rétention sont générés ; ils seront utilisés lors de l'étape de l'identification. Plusieurs approches pour l'interprétation des données LC-MS/MS en mode DDA existent. On peut citer :

- Approche par recherche dans les banques de données protéiques ;
- Approche par séquençage *de novo* ;
- Approche par recherche dans les banques de spectres ;
- Approche hybride combinant le séquençage *de novo* et la recherche dans les banques de données protéiques.

### **2.1.1. Identification par recherche dans les banques de données protéiques**

Cette approche aussi appelée, par analogie à celle de l'empreinte peptidique massique PMF, « Peptide Fragment Fingerprinting » ou PFF [132], est la méthode la plus fréquemment utilisée dans les études protéomiques à grande échelle.

Après l'analyse LC-MS/MS en mode DDA, un fichier contenant pour chaque spectre MS/MS enregistré, la masse de l'ion précurseur et la liste de ses fragments générés ainsi que son temps de rétention. L'approche PFF consiste à comparer des listes de masses des ions précurseurs et de leurs fragments associés avec des listes de masses des peptides issues de la digestion théorique des protéines présentes dans des banques de séquences et de la fragmentation théorique associée à ces peptides (Figure 2).

Plusieurs algorithmes permettant de réaliser ce type de recherche existent [121,123,124,133-135]. Même s'ils présentent beaucoup de similarités dans le principe, chaque logiciel présente des particularités de paramétrages et des divergences sur certains points du processus d'identification et du traitement des résultats. Pour tous ces moteurs de recherche, plusieurs paramètres doivent être réglés par l'utilisateur, les plus importants sont :

- La tolérance sur la masse de l'ion précurseur et celle des ions fragments ;
- L'enzyme de digestion et ses sites de coupure spécifique ainsi que le nombre de coupures non réalisées autorisé ;
- Les modifications des acides aminés autorisés ;
- Les types de fragments recherchés.

Au cours de cette thèse, j'ai utilisé deux logiciels moteurs de recherche différents, Mascot [123] et le logiciel libre (« open source ») OMSSA [135]. Par la suite je détaillerai le fonctionnement de ces deux moteurs de recherche.

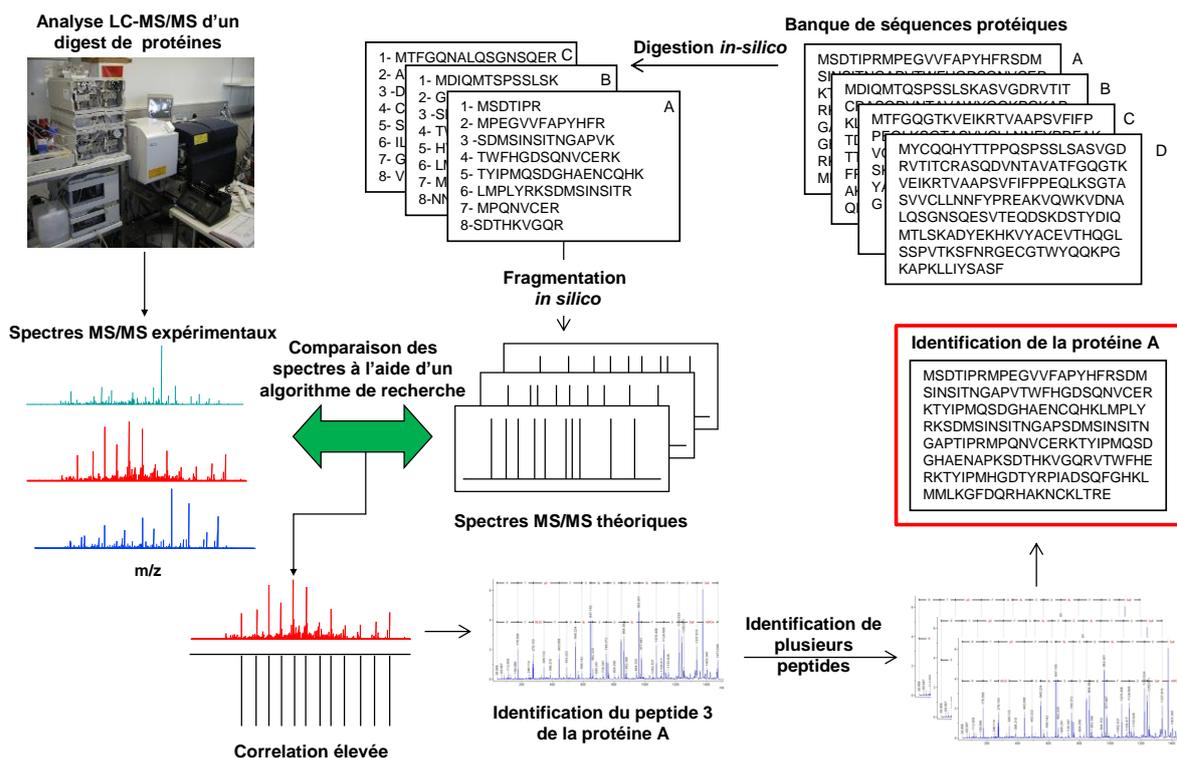


Figure 2 : Illustration de l'identification des peptides et des protéines par recherche dans les banques avec l'approche PFF.

Par ailleurs, et en général, seule une fraction des spectres MS/MS enregistrés conduit à l'identification de peptides par cette approche. Ceci peut être lié à plusieurs raisons : la qualité des spectres qui n'est pas suffisante, des spectres qui ne correspondent pas à des peptides mais à d'autres molécules présentes dans l'échantillon, ou encore des peptides modifiés ou qui ne sont pas dans la banque de données interrogées. Cependant, cette fraction de spectres qui ne conduit pas à des identifications peut s'expliquer en partie par le moteur de recherche utilisé. En effet les résultats d'identification peuvent être différents d'un moteur de recherche à l'autre. Ceci est dû à l'utilisation d'algorithmes de recherche différents pour la comparaison des données théoriques et expérimentales et le calcul du score de corrélation (voir paragraphe suivant) entre ces données [136]. Ceci a conduit certains chercheurs à recommander l'utilisation plusieurs moteurs de recherche afin de profiter de la complémentarité que peut présenter chacun des algorithmes pour renforcer la validation des résultats.

### 2.1.1.1. Attribution de scores par les moteurs de recherche aux identifications

Généralement, les moteurs de recherche attribuent un score aux identifications. Les modes de calcul de ces scores divergent d'un algorithme de recherche à l'autre. Ici, je décrirai brièvement les approches utilisées par les moteurs Mascot (Matrix Sciences) et OMSSA (NCBI).

### 2.1.1.1.1. *Le score de corrélation*

Lors du traitement des données MS/MS, un score de corrélation qui est fonction du degré de similarité entre le spectre expérimental et le spectre théorique du peptide permet de classer les séquences des peptides attribués à un spectre donné. Généralement, seul le peptide ayant le meilleur score de corrélation est conservé pour la suite de l'évaluation de l'identification. Ce score est calculé de manière propre à chaque moteur de recherche et n'est pas forcément visible dans les résultats. Pour l'algorithme de recherche Mascot, ce score de corrélation est appelé « ion score », mais son mode de calcul n'est pas publié par l'éditeur. De manière générale, le score de corrélation dépend du nombre de fragments attribués ou non attribués, de l'intensité des fragments attribués ou non attribués, de la présence de séries de fragments du même type et de l'erreur sur la masse des fragments.

### 2.1.1.1.2. *Le score probabiliste*

Le score de corrélation est ensuite converti en une mesure statistique appelée « E-value » ou « expectation » value. Cette valeur correspond au nombre de valeurs attendu avec un score supérieur ou égal au score de corrélation sous l'hypothèse que ces peptides soient attribués par hasard. Pour OMSSA, cette valeur est calculée en considérant que la distribution des scores de corrélation suit une distribution de Poisson. Plus cette E-value est élevée, moins l'identification est probable. Vue l'étendue de la gamme de E-value, cette valeur est convertie en score probabiliste «  $-\log(\text{E-value})$  ».

Mascot affiche le score de corrélation (« ion score ») et une valeur score d'identité (« Identity score »). Ce score d'identité correspond à la valeur de score pour laquelle l'identification a une E-value de 0,05 par défaut (mais qui peut être modifiée), c'est-à-dire que la confiance est de 0,95 (95 %) pour un score de corrélation égal à ce score d'identité. On a donc généralement tendance à présenter le score probabiliste comme étant la différence entre le score d'identité et le score de corrélation (« ion score – Identity score »). Plus cette différence est grande, plus la valeur de E-value est petite. La valeur de E-value pour l'identification d'un peptide donné, correspond à la valeur fixe de E-value utilisée dans le calcul de l'« identity score » (0,05 par défaut) divisée par cette différence de score.

Les paramètres dont dépend le score probabiliste sont liés à l'espace de recherche dans lequel les identifications sont effectuées, à savoir, la taille de la banque de données protéiques, les enzymes définies et leur spécificité, la tolérance sur la masse de l'ion précurseur et les modifications des acides aminés. Plus l'espace de recherche est grand, plus le score probabiliste sera petit et le risque de fausses identifications (faux positifs) est élevé.

Récemment, un nouveau type de score fut introduit, le H-score [137] qui est un score calculé pour les résultats générés par les algorithmes de recherche existant. Il s'appuie sur le fait que l'identification d'un peptide complètement séquencé est plus fiable qu'un homologue partiellement séquencé comme décrit dans les études sur le séquençage *de novo* [138,139]. Il utilise la précision de mesure des ions fragments et est indépendant de leur intensité. Il ajoute donc un point pour chaque site de dissociation identifié indépendamment

de son intensité et rajoute trois points si tous les fragments possibles sont identifiés sur le spectre. Ce score s'est montré, d'après l'étude de Savitsky et coll., capable d'identifier 24 % de spectres en plus avec 1 % de taux de faux positifs par rapport aux méthodes classiques utilisant les scores Mascot. Il requiert cependant l'utilisation de spectromètre de masse qui permettent d'obtenir une haute précision de mesure de masse sur les ions fragments (< 20 ppm) et qui génèrent des spectres MS/MS avec très peu de bruit [137].

#### 2.1.1.2. La combinaison de plusieurs moteurs de recherche

Certaines études ont montré que les moteurs de recherche présentent des performances différentes et dépendantes du type de données MS utilisées, de l'échantillon analysé et du spectromètre de masse qui a servi à générer les données [140,141]. Plusieurs études ont reporté que la combinaison de plusieurs moteurs de recherche permet d'augmenter le nombre de peptides identifiés. Ceci conduit à plus de fiabilité dans l'identification des protéines (avec plus de peptides) ou à augmenter le nombre de protéines identifiées (avec l'identification de nouveaux peptides) [142-145]. De plus, cette approche permet d'effectuer des validations croisées des résultats et ainsi de renforcer la fiabilité des identifications identiques issues de plusieurs moteurs de recherche [144].

Toutefois, il est nécessaire de pouvoir combiner les scores des identifications des différents moteurs de recherche pour obtenir un nouveau score probabiliste afin de pouvoir tirer pleinement profit de cette approche. Plusieurs études ont proposé des procédures pour calculer ce score [136,146,147] et certaines ont été intégrées dans des logiciels de traitements de résultats de recherche protéomique.

#### 2.1.1.3. Evaluation des résultats d'identifications et du taux de fausses découvertes

Les résultats obtenus après les recherches dans les banques de données avec les différents algorithmes qui existent ne sont jamais sûrs et doivent être inspectés et validés. Il était habituel dans les années 1990 et le début des années 2000 de générer une liste d'identifications en appliquant un seuil minimal que devait dépasser le score d'identification calculé par l'algorithme de recherche pour qu'elle soit considérée sûre. Ceci était souvent accompagné d'une inspection manuelle des spectres MS/MS par un expert. Or les scores générés par un moteur de recherche dépendent de plusieurs paramètres qui changent d'une expérience à l'autre comme la performance du spectromètre de masse (précision, résolution), la qualité des spectres MS/MS et l'espace de recherche. Il n'est pas donc judicieux d'utiliser le même seuil de score pour des expériences différentes et de plus, pour les expériences à grande échelle, l'inspection manuelle est très couteuse en temps et reste une solution subjective et dépendante du niveau d'expertise de l'expérimentateur.

Plusieurs articles indiquent que les données protéomiques publiées contiennent un grand nombre de faux positifs [148-155] et que la pertinence statistique associée à l'assignation de spectres MS/MS à des séquences de peptides n'est pas en corrélation avec la vraisemblance de l'identification de ces peptides. Ceci est attribué à des vrais

correspondances spectre-peptides assignées à des protéines individuelles, tandis que des correspondances spectre-peptides se produisent avec une distribution identique à travers la base de données [156]. Ceci a conduit à désigner l'analyse bioinformatique comme étant le maillon faible ou le talon d'Achille de la protéomique [157]. Le taux alarmant de faux positifs dans les publications de données protéomiques a conduit la communauté scientifique à établir des directives pour l'analyse et la publication de données par les éditoriaux des journaux [158]. Ici, je décris les stratégies les plus utilisées pour la validation des identifications en analyse protéomique.

### *2.1.1.3.1. La stratégie Target-Decoy*

La stratégie « Target-Decoy » est apparue en 2002 [159] et a été développée au cours des années 2000 [143,160,161]. C'est une approche facile à implémenter et applicable à plusieurs situations en analyse protéomique [158]. Elle consiste à effectuer la recherche dans une banque protéique « Target » (cible) concaténée avec une banque « Decoy » (leurre) de même taille. Une autre alternative est d'effectuer la recherche dans les deux banques séparément. La banque Decoy ou leurre est une banque qui contient les mêmes séquences que dans la banque Target mais inversées, mélangées ou randomisées. Quelques études ont évalué les différentes façons de générer les séquences leurres et ont généralement trouvé peu de différences [161-163]. L'hypothèse de base de cette approche est que l'identification de peptides Decoy et l'identification de faux positifs dans la banque Target suivent la même distribution. Des seuils de score sont ensuite établis de façon à avoir un taux de fausses identifications acceptable. Ce taux est calculé comme étant le rapport du nombre d'identifications Decoy et du nombre d'identifications Target (Figure 3).

Ils existent deux façons pour faire du Target-Decoy, la première consiste à faire deux recherches indépendantes dans la banque Target puis dans la banque Decoy, tandis que dans la deuxième la recherche est effectuée dans une banque composite contenant les séquences Target et les séquences Decoy [164-166]. Dans le cas de la recherche séparée, le taux de fausses découvertes FDR (« False Discovery Rate ») calculé est surestimé. Ceci est due au fait que les spectres MS/MS ayant conduit à des identifications correctes dans la banque Target sont aussi utilisés pour la recherche dans la banque Decoy et donc peuvent correspondre à des séquences leurres. Une correction doit donc être effectuée en prenant ceci en compte [167]. La recherche dans une seule banque composite est de loin la plus utilisée [158]. Ceci permet de s'affranchir du problème décrit ci-dessus mais le FDR est aussi surestimé dû au fait de la compétition entre les peptides. Un spectre peut avoir une meilleure correspondance avec une séquence leurre qu'avec une vraie séquence et donc il sera assigné à la séquence leurre avec un meilleur score ce qui fait augmenter l'estimation du taux de FDR et fait diminuer le nombre de vrais positifs [168].

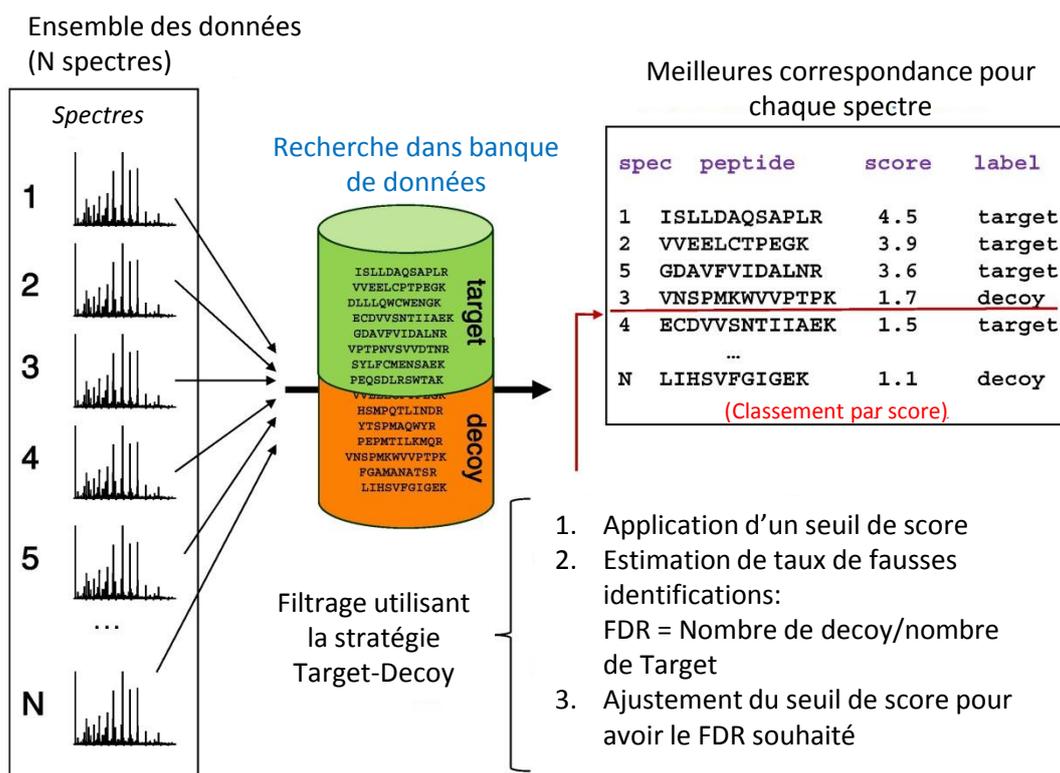


Figure 3 : Illustration de la stratégie Target-Decoy pour l'estimation du taux de fausses découvertes (adaptée de [158]).

La stratégie Target-Decoy permet d'avoir une estimation globale du taux de fausses identifications, et reste la stratégie la plus utilisée, cependant, elle ne fournit pas un score statistique de confiance pour une identification individuelle.

### 2.1.1.3.2. Les approches empiriques de Bayes

Celles-ci permettent de calculer une probabilité à posteriori (« posterior probability ») ou un score statistique individuel pour une identification donnée. Elle consiste à déterminer une distribution à *posteriori* de la probabilité  $p$  d'une loi binomiale [169]. Dans le contexte d'identification des peptides, ce modèle de taux de fausses identifications fut introduit par Keller et Coll. [170] et est implémenté dans l'outil bioinformatique PeptideProphet. Il s'agit d'une approche statistique qui explique la distribution d'intérêt (ici les scores probabilistes des correspondances spectre-peptides observés dans l'ensemble des données) comme étant une distribution bimodal (correspondances correctes ou incorrectes). La figure 3 illustre cette approche [158]. Dans cette approche, la fraction ( $\pi$ , dans la figure 3) de toutes les correspondances spectre-peptides correctes et les autres paramètres régissant la distribution des scores pour les correspondances correctes et incorrectes ( $f_1(S)$  et  $f_0(S)$  respectivement) sont calculés en utilisant l'algorithme Espérance-Maximisation [171]. Le résultat est une probabilité qui est calculée pour chaque correspondance à posteriori de la recherche dans la banque de données et qui peut être utilisée pour filtrer les résultats et estimer le FDR. Ainsi, l'ensemble des données d'identifications peut être filtré en utilisant le seuil de probabilité qui correspond au FDR visé.

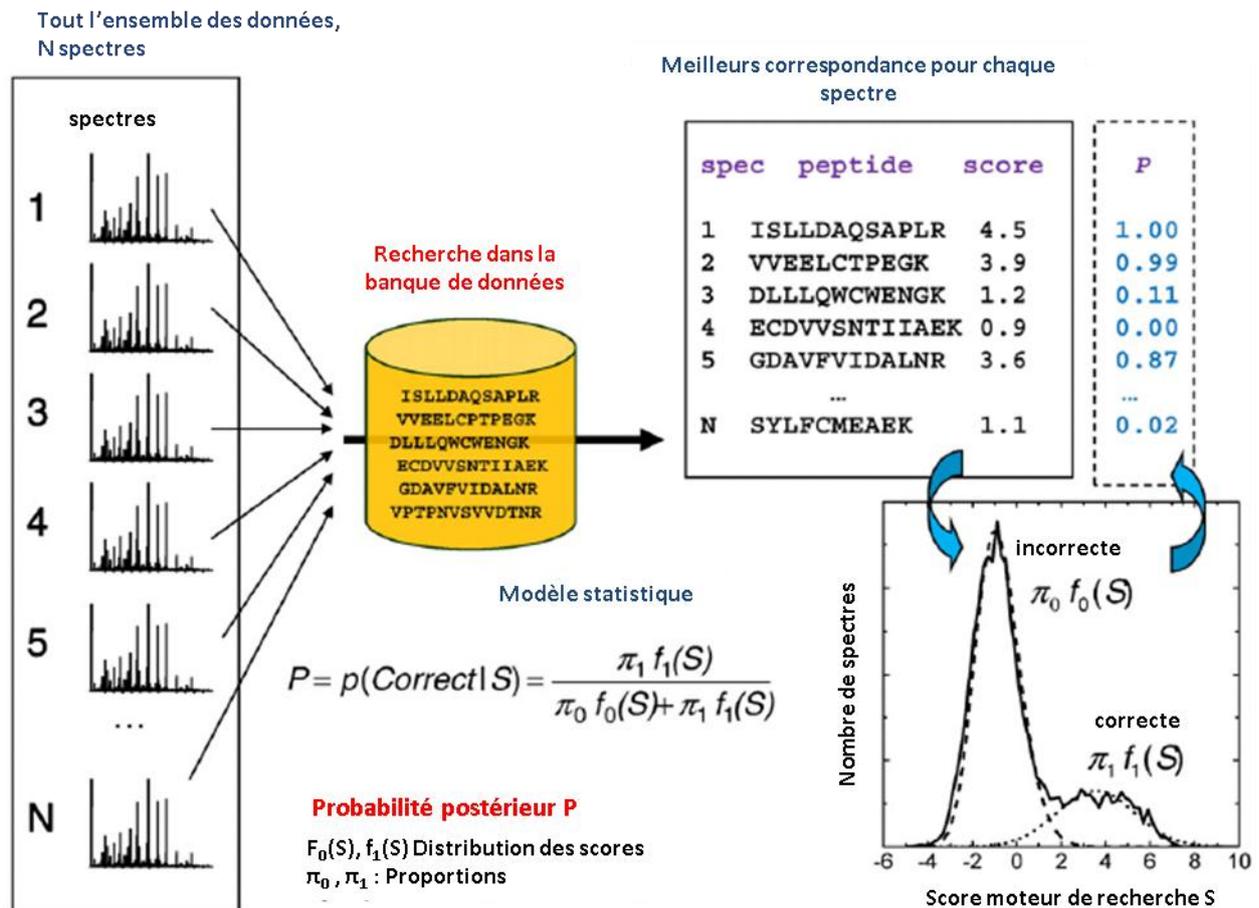


Figure 4: Approche empirique de Bayes pour le calcul de probabilité postérieur. La distribution la plus probable parmi les correspondance spectres-peptides correctes et incorrectes sont modélisées à partir de l'ensemble des données en utilisant l'algorithme de l'Espérance-Maximisation (adaptée de [158]).

Une amélioration de l'approche est l'utilisation de séquences Decoy qui permet d'intégrer leurs scores automatiquement dans la distribution des correspondances incorrectes. Ceci permet d'avoir une référence stable pour la modélisation de la distribution des correspondances incorrectes ce qui permet une estimation plus précise et robuste du FDR.

L'évaluation du FDR au niveau protéique, après avoir évalué la confiance accordée aux identifications peptidiques, peut être aussi réalisée par les approches empiriques de Bayes. PeptideProphet combine par exemple les probabilités postérieures d'identification des différents peptides participant à l'identification d'une protéine donnée, appelés « sibling peptides », afin de calculer la probabilité pour que l'identification de la protéine soit correcte. De plus, PeptideProphet recalcule la confiance accordée aux identifications des peptides en tenant compte de l'existence de « sibling peptides » qui sont les peptides appartenant à une même protéine [156].

### 2.1.2. Identification par séquençage de novo

Cette approche consiste à déduire la séquence du peptide ou une partie de cette séquence directement du spectre MS/MS. Ceci peut être réalisé manuellement ou par un logiciel approprié. La détermination des fragments de séquence sur les spectres est réalisée

à l'aide des différences de masses entre les différents ions fragments en prenant en compte les règles de fragmentations (voir chapitre 1 paragraphe 1.4). Les séquences ou fragments de séquences ainsi déterminés sont ensuite soumis à des algorithmes de recherche de similarités de séquences du type MS-BLAST (« Basic Local Alignment Search Tool ») [172,173]. Après l'interprétation des spectres MS/MS, les fragments de séquences obtenus sont comparés aux séquences de protéines présentes dans les banques. MS-BLAST permet de soumettre l'ensemble des fragments de séquences d'une analyse donnée en une seule requête [174]. Lorsqu'un peptide montre une similarité avec un peptide présent dans la banque, MS-BLAST lui attribue un score individuel. Si plusieurs peptides identifiés de cette manière correspondent à une seule protéine un score global est calculé et il tiendra compte de la somme des scores individuels. Habermaann et al. ont réalisé une évaluation de MS-BLAST en 2004 et ont montré que les identifications effectuées sur plus de 20000 requêtes présentaient un taux de faux positifs inférieur à 3% [175].

Cette approche présente l'avantage de pouvoir identifier des protéines absentes des banques de données ou plutôt des orthologues de protéines par homologie de séquences. Ceci est particulièrement utile lors de l'étude de protéomes provenant d'organismes dont le génome n'est pas séquencé ou n'est que partiellement séquencé. De plus, cette approche est tolérante aux erreurs et prend donc en compte des différences dans les séquences ou les modifications d'acides aminés qui peuvent être présentes (mutation, inversion, délétion, insertion, ...) et qui ne sont généralement pas prises en compte dans une approche par recherche dans les banques de données. Il est évident que l'utilisation de cette approche nécessite des spectres MS/MS de bonne qualité et donc informatifs.

### ***2.1.3. Identification par recherche dans les banques de spectres***

Au fil des années les études protéomiques appliquées à des organismes modèles tel que l'Homme, le rat et la levure ont permis d'accumuler de grandes quantités de données d'identification protéomique avec un grand degré de redondance. Ces données ont été rassemblées dans des banques spécialisées comme PRIDE [176] et PeptideAtlas [177]. Ceci a permis le développement d'une nouvelle approche qui consiste à comparer les spectres MS/MS expérimentaux à une librairie de spectres intégrant un grand nombre de spectres de fragmentation de peptides issus d'identifications correctes antérieures. Des nouveaux outils permettant la comparaison et la corrélation entre les spectres expérimentaux et les spectres de librairies ont donc été développés. On peut citer parmi ceux-ci X! P3 [178], SpectraST [179] et Biblispec [180].

Si cette approche est performante en termes de vitesse, de spécificité et de sensibilité [179], elle devra souvent être complétée par des recherches dans des banques de données protéiques et ceci parce que les librairies de spectres restent limitées à quelques organismes dont les spectres ont déjà été plusieurs fois identifiés.

#### **2.1.4. Les approches de recherches hybrides**

Des approches combinant à la fois des éléments du séquençage *de novo* et de la recherche dans les banques de données protéiques ont été aussi développées. La plus commune consiste à extraire pour chaque spectre MS/MS des bouts de séquences appelés « tags » [122]. Un tag est une séquence courte en acides aminés précédée et suivie d'une masse qui désigne sa position dans la séquence peptidique. C'est seulement avec les peptides contenant ces tags de séquence que la recherche dans les banques de données est réalisée. Ceci réduit considérablement le nombre de comparaison à effectuer et le temps de recherche [158]. Ces approches hybrides sont particulièrement intéressantes pour l'identification de peptides modifiés, soit chimiquement, soit par modifications post-traductionnelles [181-183].

#### **2.2. Identification en mode d'acquisition données-indépendant**

Récemment, une nouvelle approche a été proposée par la compagnie Waters sur ses nouvelles générations de TOF. Dans cette nouvelle approche LC-MS/MS appelée « MS-Expression », les ions ne sont pas sélectionnés et isolés avant leur fragmentation. En effet, après l'enregistrement d'un spectre MS, tous les ions sont fragmentés en même temps dans la cellule de collision. Les spectres de fragmentation ainsi générés et appelés spectres MS<sup>E</sup> présentent les fragments de tous les ions visualisés dans le balayage MS. Puisque chaque fragment suit le profil d'élution de son précurseur, l'alignement précis des temps d'élution permet d'associer chaque fragment à un ion précurseur dont la masse exacte est déterminée sur le spectre MS. Des spectres de chaque précurseur sont ainsi reconstitués et des recherches sont effectuées dans les banques de données protéiques [184]. Théoriquement, la perte d'information est très limitée étant donné que tous les ions seront fragmentés [185]. Cependant, ce mode d'acquisition ne permet pas une vérification manuelle des résultats vue la complexité des spectres obtenues et la difficulté à reconstruire les spectres individuels des peptides.

## Chapitre 3 Les stratégies de quantification en analyse protéomique

Un des plus grands défis de l'analyse protéomique est de pouvoir, après l'analyse qualitative, fournir des informations quantitatives sur l'abondance des protéines et leur niveau d'expression [186]. Ceci était traditionnellement effectué à l'aide de gels d'électrophorèse 2D [187]. Après que la spectrométrie de masse se soit illustrée comme étant la méthode de choix pour l'analyse des protéomes complexes, de nouvelles méthodes de quantification de protéines basées sur la spectrométrie de masse ont été développées. Cependant, l'analyse des protéines par LC-MS/MS n'est pas une technique quantitative en soit car elle dépend d'une multitude de facteurs qui sont difficilement contrôlables. Ces facteurs se trouvent à tous les niveaux de l'analyse protéomique allant de la préparation de l'échantillon (accessibilité de l'enzyme lors de la digestion, solubilité des peptides et des protéines...) à la séparation en chromatographique et la détection en MS liées aux propriétés intrinsèques des peptides (taille, composition en acides aminés, polarité, ...) [187]. Il est donc généralement nécessaire de comparer chaque peptide individuellement à travers les expériences [188].

En plus des analyses protéomiques comparatives et quantitatives par des approches basées sur les gels 2D, on trouve deux catégories de méthodes de quantification de protéines en spectrométrie de masse. Les approches « label free » qui sont effectuées sans aucun marquage des protéines et des peptides et les approches basées sur les marquages aux isotopes stables appelées « SILE quantitation » (« Stable Isotope Labelling Experiments »). Domon et Aebersold ont décrit dans un article en 2010 [189] les trois stratégies de protéomique quantitative, à savoir la stratégie de quantification en « shotgun » ou « découverte », la stratégie de quantification dirigée (« directed ») et la stratégie de quantification ciblée (« targeted »). Dans ce chapitre, nous décrirons les principales méthodes de quantification en analyse protéomique et nous commenterons les trois stratégies décrites dans cet article.

### 1. Quantification basées sur les gels d'électrophorèse 2D

Les gels d'électrophorèse bidimensionnelle permettent de visualiser des différences d'abondance entre les protéines et donc d'effectuer une quantification relative des protéines dans différentes conditions [9]. Ceci est réalisé à l'aide de logiciels permettant la

comparaison d'image et des intensités entre les différents spots. Traditionnellement, les gels sont colorés au Bleu de Coomassie, mais récemment, une nouvelle technique appelée gel 2D DiGE (Differential Gel Electrophoresis) a été développée et permet une meilleure reproductibilité et une meilleure sensibilité.

### **1.1. Protéomique quantitative à l'aide de l'électrophorèse bidimensionnelle classique**

Dans un gel d'électrophorèse 2D, l'intensité d'un spot, sous forme d'intégration de la densité optique représentée dans ce spot, est directement proportionnelle à l'abondance de cette protéine. Deux échantillons peuvent ainsi être comparés grâce à une quantification relative basée sur les rapports d'intensités des spots d'une même protéine et ceci à l'aide de l'analyse d'image. Pour les études différentielles, la coloration au bleu colloïdal est préférée à celle au nitrate d'argent en raison de sa gamme linéaire plus étendue. En effet, si la coloration au nitrate d'argent permet une meilleure sensibilité, l'intensité de la coloration varie d'une protéine à une autre et sa gamme linéaire est très étroite [190]. Enfin, l'électrophorèse bidimensionnelle classique, en plus de ces limitations citées dans le chapitre précédent, souffre de problèmes de reproductibilité et de variabilité liés au traitement individuel des échantillons. L'électrophorèse différentielle 2D DiGE et les systèmes d'automatisation récents représentent des améliorations de cette approche.

### **1.2. L'électrophorèse différentielle 2D DiGE**

Cette technique introduite par Unlü et al. en 1997 [191] permet la quantification relative de deux échantillons protéiques différents sur le même gel d'électrophorèse. Les extraits de protéines sont tout d'abord marqués avec des fluorochromes différents, ayant la même masse moléculaire mais des longueurs d'onde d'excitation et d'émission différentes. Ces fluorochromes sont des cyanines appelées Cy3 et Cy5. Les protéines ainsi marquées sont mélangées et séparées simultanément sur le même gel d'acrylamide bidimensionnel. Les protéines sur le gel sont ensuite visualisées à l'aide d'un scanner approprié permettant l'acquisition d'images aux longueurs d'ondes spécifiques de chaque fluorochrome. Ces images sont ensuite comparées à l'aide d'un logiciel d'analyse d'image ce qui permet de réaliser la quantification relative des protéines. Alban et al. ont introduit une amélioration de cette technique par l'utilisation d'un standard interne qui lui est marqué par un fluorochrome différent, le Cy2 [192]. L'utilisation de ce standard interne permet d'effectuer des comparaisons entre plusieurs échantillons séparés sur des gels différents. Ce standard interne est constitué d'un mélange en quantité égale de tous les extraits protéiques à comparer et marqués donc au Cy2. Toutes les protéines de tous les extraits seront représentées dans ce standard qui est donc déposé sur tous les gels avec les échantillons à comparer. Ceci permet donc de normaliser par rapport à ce standard interne tous les échantillons. L'abondance d'une protéine est donc déterminée en calculant le ratio de l'intensité du spot dans lequel elle se trouve par rapport à l'intensité du spot correspondant à la même protéine dans le standard. Les variations techniques sont ainsi dissociées des

variations biologiques ce qui constituait une limitation principale de l'approche par gel 2D classique. La figure 1 illustre la stratégie d'analyse par gel 2D DiGE.

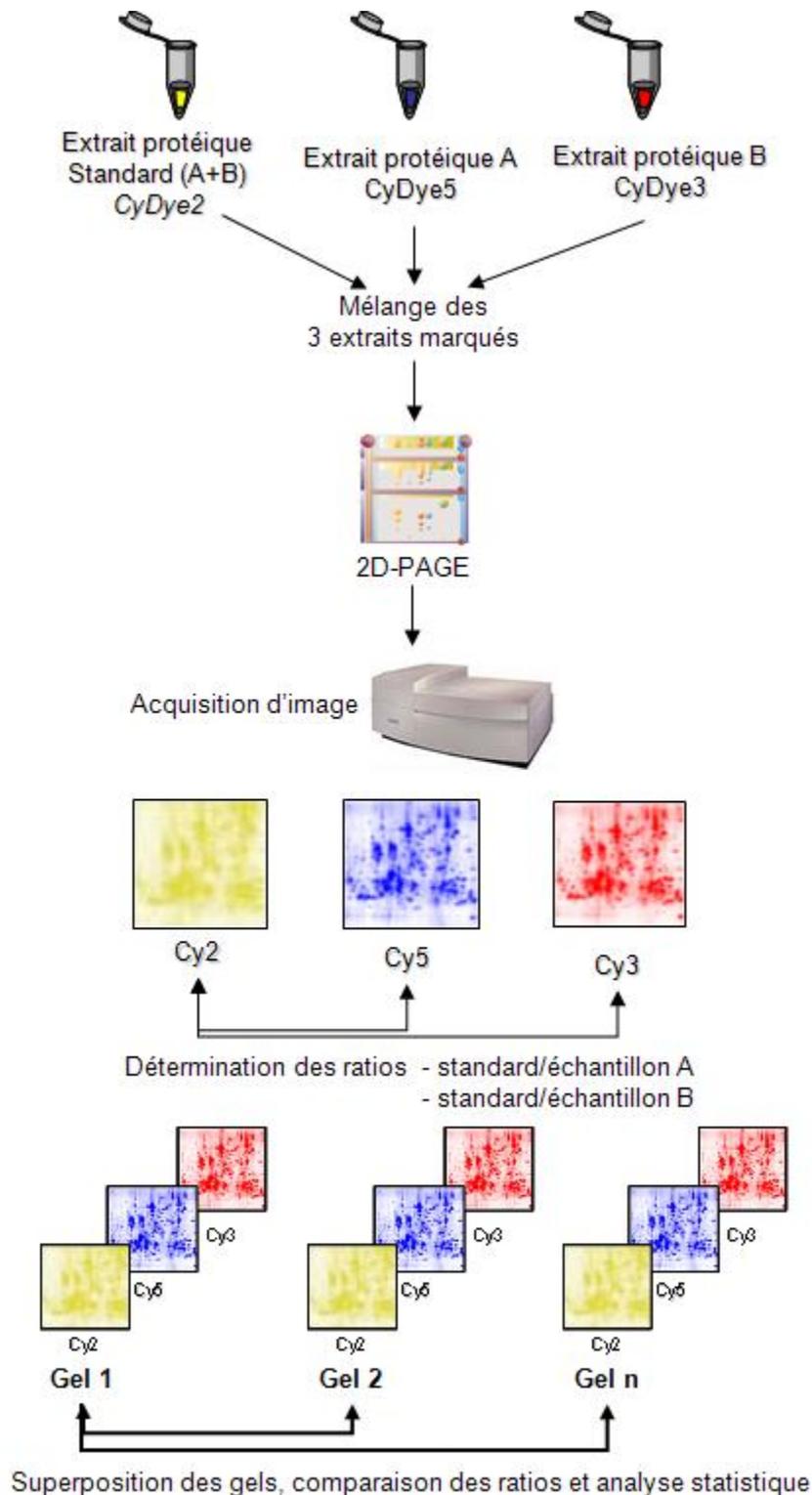


Figure 1 : Illustration de la stratégie d'analyse en Gel d'électrophorèse 2D DiGE avec l'utilisation d'un standard interne (d'après [193]).

L'avantage majeur de cette technique est de s'affranchir, grâce à l'utilisation du standard interne, des problèmes de reproductibilité observés dans l'analyse par gel 2D

classique. Les variations dans les profils de spot d'un gel à l'autre sont ainsi réduites. Bien que la réalisation de répliques biologiques reste nécessaire pour avoir des données statistiques, la comparaison des échantillons sur un même gel et l'utilisation du standard interne pour la normalisation ne nécessitent pas d'effectuer des répliques techniques [194] ce qui réduit considérablement le temps et la difficulté de la réalisation des analyses. De plus, l'utilisation des fluorochromes permet de gagner en linéarité, en gamme dynamique (3,6 ordres de grandeur) et en sensibilité (~0.025 ng) par rapport à la coloration au Bleu de Coomassie (100 ng) et même par rapport à l'argent (~0,5 ng) [195]. Ces avantages ont permis à cette technique de supplanter l'approche gel 2D classique. Cependant elle présente toujours les limitations intrinsèques au gel et à l'électrofocalisation à savoir l'analyse des protéines membranaires et les pI extrêmes.

## 2. Les méthodes sans marquage (« Label Free »)

Les méthodes dites « Label Free » sont des méthodes de quantification relative reposant sur des données spectrales et n'impliquant aucun marquage. Deux principales stratégies peuvent être distinguées, la première est basée sur la comparaison des intensités des signaux MS des peptides tandis que la deuxième est basée sur le comptage des spectres MS/MS (« *Spectral counting* »).

### 2.1. La quantification basée sur la comparaison des intensités des signaux MS

Dans cette stratégie, les échantillons à comparer sont analysés séparément. Après identification des peptides, les courants d'ions de chaque peptide sont extraits et les aires des pics chromatographiques correspondants sont intégrées. Les valeurs obtenues pour l'ensemble des peptides issus d'une même protéine sont comparées à celles obtenues dans une ou plusieurs analyses. Ainsi les rapports des aires des pics de peptides correspondant à une protéine donnée dans différentes conditions permettent d'avoir des informations sur son abondance relative [196-200]. Ceci peut être réalisé à partir d'analyses en LC-MS ou en LC-MS/MS. Dans ce dernier cas, un ajustement entre la fréquence d'acquisition des spectres MS et MS/MS doit être réalisé. En effet, si le séquençage d'un maximum de peptides par MS/MS est requis pour obtenir un plus grand nombre de protéines identifiées, une quantification robuste par intégration des aires des pics chromatographiques nécessite un nombre suffisant de spectre MS pour bien décrire ces pics chromatographiques (idéalement au moins 9 points par pic) [188]. Ainsi, même avec les instruments présentant la meilleure performance en termes de vitesse d'acquisition, la quantification la plus juste est obtenue au détriment du nombre d'indentifications des protéines présentes et inversement.

Une autre approche appelée « *Accurate Mass and Time tag* » (AMT) [201] consiste à effectuer la LC-MS et la LC-MS/MS séparément. Ainsi l'analyse en LC-MS/MS sert à identifier le maximum de peptides et par suite des protéines. Elle permet de constituer une large banque d'identification des peptides définis par leur séquence, leur rapport m/z et leur temps de rétention. La quantification est ensuite réalisée lors de l'analyse LC-MS. L'échantillonnage des signaux des peptides intacts est optimisé. Les couples m/z et temps de

rétenion dans la base AMT servent de coordonnées pour l'identification des peptides lors de la quantification. Une autre approche consiste à d'abord réaliser l'analyse LC-MS pour quantifier les espèces différentielles. Une analyse LC-MS/MS est ensuite réalisée afin d'identifier les peptides ainsi quantifiés en établissant des listes d'inclusion pour favoriser l'identification des espèces d'intérêt.

Il est évident que dans cette stratégie, la reproductibilité de la séparation chromatographique est critique et doit être optimisée afin que chaque peptide soit élué à un temps de rétention identique d'une analyse à l'autre. Des logiciels spécialisés sont souvent utilisés pour ajuster l'alignement des profils chromatographiques [188]. Il est aussi nécessaire d'utiliser des spectromètres de masse de haute résolution de type Q-TOF ou IT-Orbitrap afin d'éviter les problèmes de signaux parasites et de pouvoir extraire les courants d'ion avec une tolérance restreinte sur la fenêtre  $m/z$ . Un des principaux inconvénients de cette technique est les erreurs systématiques introduites par les variabilités dues à la préparation et à l'analyse séparée des échantillons à comparer. Cette variabilité peut être diminuée en introduisant des quantités fixes de protéines qui serviront de standards internes et permettront ainsi de normaliser les signaux.

## 2.2. La quantification « *Spectral Counting* » basée sur les données d'identification

Cette approche est basée sur l'observation empirique que, plus une protéine est concentrée dans un mélange donné, plus ces peptides vont être sélectionnés et fragmentés et donc un plus grand nombre de spectres MS/MS seront associés à cette protéine. Par conséquent, la comparaison du nombre de spectres MS/MS associés à une protéine dans deux échantillons différents permet une quantification relative de cette protéine [202-205]. Cette approche utilise donc les données MS/MS à la fois pour identifier et pour quantifier contrairement à l'approche précédente qui était basée sur les données MS. L'approche « spectral counting » implique donc une réalisation d'un maximum de spectres MS/MS informatifs et de qualité afin de permettre une quantification robuste. En effet, il a été observé que la précision de la quantification est directement dépendante du nombre de spectres MS/MS considéré. Ainsi, il est possible de détecter une variation d'un facteur 3 dans l'abondance d'une protéine avec 4 ou 5 spectres, mais mettre en évidence une variation d'un facteur 2 requiert pas moins de 15 spectres [206]. Les spectromètres de masse de type trappe ionique (linéaire ou tridimensionnelle) sont privilégiés puisqu'ils présentent l'avantage de pouvoir enchaîner les cycles de MS et de MS/MS à grande vitesse par rapport aux autres instruments tels que les Q-TOF. D'un autre côté, les paramètres d'exclusion dynamique couramment utilisés lors des analyses LC-MS/MS en mode données dépendant (DDA) doivent donc être adaptés.

Dans cette approche, le nombre de spectres des peptides est considéré comme un index de l'abondance des protéines [203,207,208]. Or un des biais envisagés est directement lié à la taille de la protéine. Les protéines de grande taille génèrent, par digestion

enzymatique, un nombre plus important de peptides et donc auront un plus grand nombre de spectres MS/MS. De plus, un faible nombre de spectres MS/MS peut avoir de multiples raisons telles qu'une protéine de séquence courte ou une faible détection par MS des peptides liée à leurs propriétés physico-chimiques et à la complexité de l'échantillon [209]. Le PAI (Protéine Abundance Index) [210] et plus tard le emPAI (Exponentially modified PAI) [211] ont été introduits pour permettre de corriger certains de ces biais. Ces index permettent de normaliser le nombre de peptides observés par rapport aux nombres de peptides théoriquement observables de la protéine d'intérêt. La normalisation peut être aussi réalisée en divisant le nombre de spectres MS/MS par la masse ou le nombre d'acides aminés de la protéine [212-214]. Ces normalisations permettent une correction assez satisfaisante de l'estimation des abondances relatives des protéines [215]. Cette méthode « Spectral Counting » a aussi été utilisée pour analyser des échantillons marqués aux isotopes stables [216].

Certains chercheurs, se basant sur la corrélation entre la quantité de protéines et le nombre de spectre ont proposé d'étendre l'utilisation de l'approche « Spectral Counting » à l'estimation des niveaux d'expression absolus des protéines. Ainsi, la connaissance de la quantité d'une ou de plusieurs protéines présentes dans les mélanges ou ajoutées en tant que standards internes permet d'estimer l'abondance absolue de l'ensemble des protéines du mélange étudié. Dans ce cadre, la technique APEX (Absolute Protein EXpression measurement) [217] a été proposée par Lu et coll. en 2007. Celle-ci prend en compte le nombre de peptides observés et la probabilité que les peptides soient détectés par le spectromètre de masse. Le facteur de correction  $O_i$  est ainsi une mesure de la probabilité d'observer un peptide tryptique au cours de l'analyse et est calculé en fonction de sa séquence et de ses propriétés physico-chimique. Ceci a été réalisé à l'aide d'algorithmes spécialisés qui sont éduqués à l'aide de données déjà existantes. Une combinaison de APEX à la LC-MS<sup>E</sup> et la SRM (selected reaction monitoring) a été récemment utilisée dans l'étude du protéome du pathogène *Leptospira interrogans* et a permis de réaliser la quantification absolue de 50% du protéome théorique de cette espèce [218].

En général, les stratégies de quantification « Label Free » présentent l'avantage, par rapport aux techniques impliquant des marquages, de ne pas être limitées en terme de nombre d'expériences, d'échantillons et de conditions biologiques à comparer. Elles fournissent aussi une gamme dynamique plus large pour la quantification. Toutefois, ces stratégies sont considérées comme étant moins précises car l'ensemble des variations accumulées dans les différentes étapes de la préparation d'échantillon et de l'analyse se répercute sur les données finales ; les échantillons étant traités individuellement contrairement aux expériences impliquant des marquages. Les stratégies « Label Free » ont été utilisées dans divers études pour explorer la dynamique de protéomes entiers dans le but de découvrir des biomarqueurs potentiels de pathologies [219,220] ou des voies de signalisation induites par un traitement [221].

### 3. Les méthodes de marquage par isotopes stables pour la quantification des protéines par spectrométrie de masse

Les approches de marquage aux isotopes stables sont basées sur la dilution isotopique. Elle s'appuie donc sur le fait qu'un peptide marqué avec des isotopes stables présente les mêmes propriétés physico-chimiques que son homologue non marqué et ne diffère que par sa masse moléculaire. Il présentera donc les mêmes comportements en chromatographie et en spectrométrie de masse [9]. Il existe différentes techniques de marquages aux isotopes stables. Une d'entre elle est une approche ciblée utilisant des peptides de synthèse marqués qui sont ajoutés à l'échantillon. Elle permet la quantification absolue par une technique de spectrométrie de masse appelée SRM (« Selective Reaction Monitoring »). Les autres techniques permettent la quantification relative globale entre différentes conditions biologiques [222-224] et diffèrent par le mode d'introduction du marquage [188]. On distingue alors :

- Le marquage métabolique ou *in situ* [225] qui consiste à effectuer le marquage au niveau des cellules en les cultivant dans un milieu enrichi en isotope stable.
- Le marquage par réaction enzymatique réalisé au niveau des peptides. L'incorporation des isotopes stables s'effectue par digestion enzymatique dans de l'eau lourde ( $H_2^{18}O$ ).
- Le marquage chimique réalisé par dérivation chimique des protéines ou des peptides.

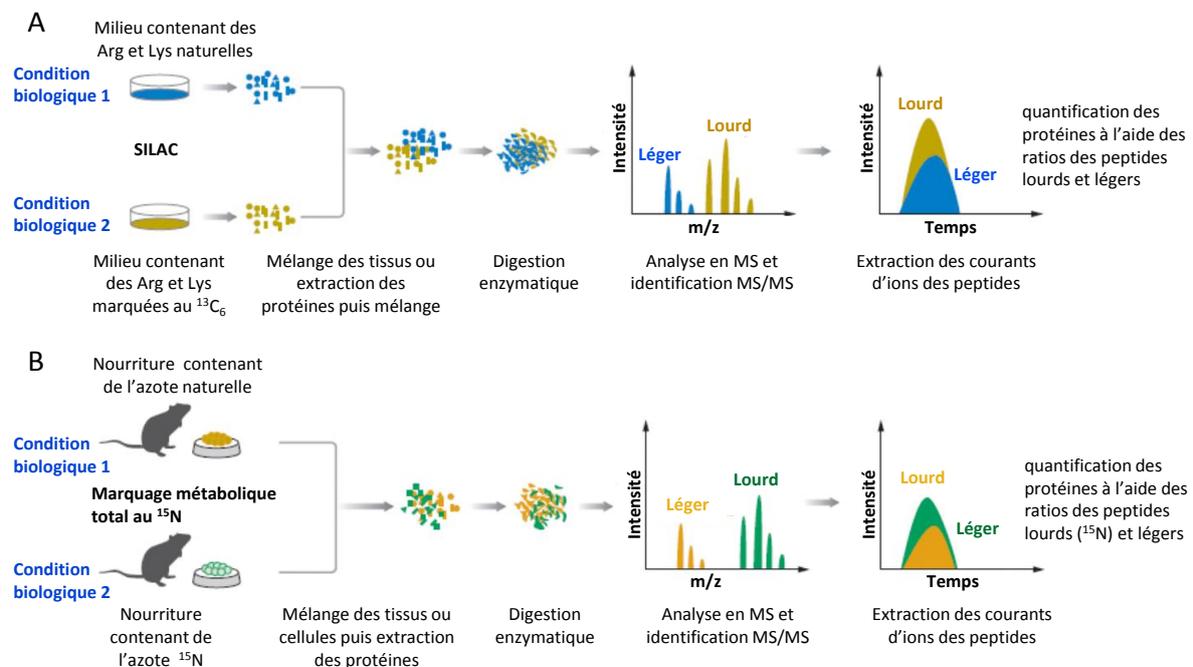
#### 3.1. Le marquage métabolique

Dans cette approche, les isotopes stables sont incorporés dans les protéines des organismes à l'aide de milieu de culture ou de la nourriture marqué aux isotopes stables. C'est donc à l'étape la plus en amont possible que le marquage est effectué. Ceci permet à toutes les protéines produites dans l'organisme étudié d'incorporer le marquage et donc de fournir un standard interne pour chaque protéine et peptide dans l'échantillon biologique. Lorsque les tissus ou cellules des organismes cultivés dans des milieux lourds et légers sont mélangés en quantité égale puis traités ensemble de la même façon, toutes les causes techniques de variabilité et de non-reproductibilité dues aux traitements tels que la lyse des cellules, l'extraction des protéines et la digestion enzymatique sont éliminées [225].

La technique de marquage métabolique la plus connue a été introduite par le groupe de Matthias Mann en 2002 [226] et consiste à utiliser un milieu de culture enrichi en acides aminés marqués aux isotopes stables. Cette approche est connue sous le nom SILAC (« Stable Isotope Labelling by Amino-acids in Cell culture »). L'acide aminé qui sert pour le marquage peut être choisi parmi les acides aminés essentiels, qui ne peuvent être synthétisés *de novo*, et ceci pour une incorporation plus rapide et plus complète dans les protéines. Les acides aminés essentiels sont la lysine, la leucine, l'isoleucine, la phénylalanine, le tryptophane, la méthionine, la valine et la thréonine. Normalement c'est la leucine et la lysine sont les acides aminés les plus utilisés étant donné leur abondance naturelle favorable dans les protéines. Cependant, l'utilisation simultanée de la lysine et

l'arginine marquées permet d'obtenir un marquage de tous les peptides tryptiques excepté le peptide C-terminal. L'arginine n'est pas un acide aminé essentiel mais plusieurs équipes ont montré son incorporation complète [227-232]. Après mélange et digestion enzymatique, les peptides sont analysés en LC-MS/MS. Les peptides marqués coéluent en chromatographie avec leurs homologues non marqués. L'identification d'un des peptides lourds ou légers permet l'identification d'une protéine. La quantification est effectuée en extrayant les courants d'ions de chaque paire de peptides et en calculant le ratio d'intensité (Figure 2–A).

Une autre approche applicable aux organismes capables de croître avec une alimentation contenant exclusivement des nutriments marqués aux isotopes stables a aussi été décrite (Figure 2–B). Généralement c'est l'azote  $^{15}\text{N}$  [233] qui est utilisé puisque c'est un atome moins abondant dans les peptides que le carbone, mais le  $^{13}\text{C}$  et  $^{18}\text{O}$  ont aussi été utilisés [225]. Cette stratégie a été appliquée avec succès sur plusieurs organismes telle que la levure [202], les plantes [234], la drosophile [235] ou encore le rat [236]. Bien qu'une substitution presque totale des atomes naturels par leurs isotopes a pu être obtenue [225], cette approche est très difficile à mettre en œuvre et demande beaucoup de temps et de ressources. De plus, le nombre d'isotopes incorporés est dépendant de la séquence du peptide (par exemple du nombre d'azote dans un peptide donné) et donc l'analyse des données générées par cette approche s'avère plus compliquée [233,237,238] que dans l'approche SILAC où le nombre de label isotopique est défini (1 acide aminé  $^{13}\text{C}_6$ -Arg ou  $^{13}\text{C}_6$ -Lys). Par conséquent, SILAC reste la technique la plus utilisée du marquage métaboliques chez les eucaryotes et les organismes supérieurs en culture cellulaire [188].



**Figure 2 : Marquage métabolique des protéines *in situ* et quantification : A- par la méthode SILAC utilisant des milieux de culture contenant des acides aminés marqués aux isotopes stable, B- par marquage métabolique total au  $^{15}\text{N}$  en utilisant des nutriments enrichi en  $^{15}\text{N}$  (adaptée de [225]).**

Le marquage métabolique, et plus particulièrement l'approche SILAC, est considéré comme étant la méthode la plus précise de quantification relative et globale par spectrométrie de masse. Ceci est due à l'introduction du marquage dans l'étape la plus en amont possible permettant de s'affranchir des variabilités techniques [188,238]. Elle est bien adaptée à l'étude de variations de l'abondance des protéines, des modifications post-traductionnelles [229,230,239,240] et des interactions protéine-protéine [241]. Néanmoins, pour les modifications post-traductionnelles, la quantification est effectuée seulement au niveau des peptides et donc s'appuie sur une seule observation (par opposition à la quantification des protéines qui utilise plusieurs peptides pour quantifier une protéine) ce qui rend la validation statistique des résultats difficile [188]. De plus, le marquage métabolique présente quelques limitations. La première étant le nombre restreint de marquage possible ce qui limite donc le nombre de conditions biologiques comparables dans une expérience (maximum 3 pour SILAC [242]). D'autre part, si certaines lignées cellulaires incorporent relativement facilement les acides aminés lourds, d'autres sont très sensibles au milieu de culture et ne permettent pas le marquage métabolique ou bien requièrent une attention particulière [188]. Un des exemples cités dans les revues est celui de la conversion de l'excès d'arginine en proline ce qui complique l'interprétation des résultats. Dans ce cas-là, un contrôle de la quantité d'arginine dans le milieu de culture est nécessaire [227]. Néanmoins, une solution à ce problème a été proposée. Elle consiste à utiliser de l'arginine  $^{15}\text{N}_4$  à la place de l'arginine légère (naturelle) et de la [ $^{15}\text{C}_6$ ,  $^{15}\text{N}_1$ ]-arginine en tant qu'arginine lourde [243]. De cette façon, la conversion de l'arginine conduira à la formation de  $^{15}\text{N}_1$ -proline et [ $^{15}\text{C}_5$ ,  $^{15}\text{N}$ ]-proline respectivement. Ainsi le taux de conversion peut être déterminé d'une façon précise et un facteur de correction peut être calculé.

### 3.2. Le marquage par réaction enzymatique

Dans cette méthode, le marquage est effectué par l'incorporation d'un ou deux atomes d' $^{18}\text{O}$  par la réaction enzymatique d'une protéase dans de l'eau lourde  $\text{H}_2^{18}\text{O}$  [244] permettant ainsi le déplacement de la masse du peptide marqué de 2 ou 4 Da. Le nombre d'atomes d'oxygène lourds incorporés dépend de la protéase utilisée. Un déplacement de 2 Da est généralement insuffisant pour différencier les enveloppes isotopiques des isotopomères. La trypsine et la Glu C permettent d'introduire 2 atomes  $^{18}\text{O}$  et sont donc préférées à la LysN et aux autres enzymes qui n'introduisent qu'un seul atome d'oxygène lourd [245]. En effet, dans de l'eau lourde et lors du clivage de la liaison peptidique par la trypsine ou la GluC, un  $^{18}\text{O}$  est échangé sur le groupement hydroxyl et un autre sur le carbonyle de l'extrémité C-terminale du peptide [225]. Vu la nature enzymatique de la réaction de marquage, les réactions secondaires non spécifiques sont évitées [188]. A des pHs extrêmes, l'isotope lourd est rechangé [246] mais dans les conditions classiques utilisées en spectrométrie de masse, le marquage est relativement stable. Le marquage enzymatique peut être effectué avec de faibles quantités de matériel et est compatible avec pratiquement tous les types de préparation d'échantillon. Cependant, étant réalisé au niveau des peptides, des variabilités techniques peuvent être observés puisque les échantillons à comparer sont traités séparément dans les étapes précédant le marquage. En

plus du déplacement relativement faible de la masse par le marquage, un autre inconvénient de cette technique est que la réaction de marquage est très souvent incomplète et des incorporations différentielles entre peptides de  $^{18}\text{O}$  ont été observées compliquant ainsi l'interprétation des données [247,248]. Les limitations de cette technique sont détaillées dans une revue de Ye et Coll. [249].

### 3.3. Le marquage chimique

Le marquage chimique consiste à modifier chimiquement les protéines ou les peptides à l'aide d'un réactif qui est marqué ou non aux isotopes stables. Tous les groupements réactifs des chaînes latérales des acides aminés (thiols, amines, carboxyles) peuvent servir comme site de dérivation. Ainsi, il existe un grand nombre de réactifs chimiques permettant d'effectuer ce genre de marquage. En pratique, la dérivation est souvent effectuée sur les groupements amines des lysines (et des amines N-terminales) ou sur les groupements thiols des cystéines. Les techniques de marquage chimique peuvent être classées en deux catégories : le marquage isotopique et le marquage isobarique.

#### 3.3.1. Le marquage isotopique

Cette technique consiste à utiliser une forme lourde et une forme légère du réactif de marquage afin de marquer *in vitro* séparément les échantillons. Les échantillons modifiés chimiquement sont ensuite mélangés et analysés simultanément en LC-MS/MS comme dans le cas du marquage métabolique. Les rapports d'intensité entre peptides marqués lourd et léger sont utilisés pour la quantification (Figure 3). Il est préférable que la différence de masse entre la forme lourde et la forme légère soit supérieure à 4 Da pour permettre de bien distinguer les massifs isotopiques [9].

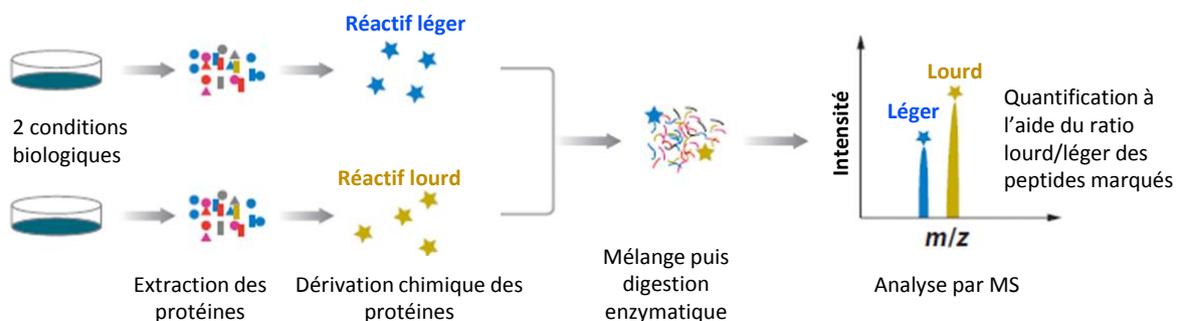


Figure 3 : Illustration de la quantification relative à l'aide du marquage chimique isotopique *in vitro* (adaptée de [225]).

Une des premières stratégies de marquage isotopique fut ICAT (« Isotope-Coded Affinity Tag ») introduite par Gygi et al. [222]. Dans cette stratégie les groupements thiols des cystéines sont dérivés en milieu réducteur à l'aide du réactif ICAT lourd (avec 8 deutériums) ou léger. Après cette modification chimique, les protéines sont mélangées et digérées. Les peptides marqués sont ensuite enrichis sur une colonne avidine à l'aide du groupement biotine contenu dans le réactif de dérivation [250] puis analysés par spectrométrie de masse. L'enrichissement des peptides marqués à l'aide de ce réactif représente un des avantages majeurs de cette technique. Cependant, elle est limitée aux protéines contenant des cystéines qui sont des acides aminés relativement rares (1,82 %).

De plus, elle est restreinte à la comparaison de seulement deux conditions biologiques [225]. Plusieurs variantes du réactif ICAT ont été développées à partir du réactif d'origine visant à améliorer l'enrichissement des peptides marqués et leur co-élution [251-253]. En effet, des petites différences de temps de rétention sont souvent observées entre les peptides deutérés et leurs homologues non deutérés en HPLC phase inverse [197]. Ceci complique l'analyse des données puisque la quantification relative des peptides ne peut plus être effectuée sur un spectre et la fenêtre de temps de rétention doit être élargie en conséquence ce qui peut induire des problèmes d'interférence avec d'autres espèces [188]. Ceci a conduit à l'utilisation du  $^{13}\text{C}$  et du  $^{15}\text{N}$  qui eux n'altèrent pas la coélution. De plus, le tag biotinylé utilisé pour l'enrichissement est devenu clivable en milieu acide ce qui simplifie l'analyse des spectres MS/MS, cette amélioration fut baptisée cICAT (« cleavable ICAT »)[254].

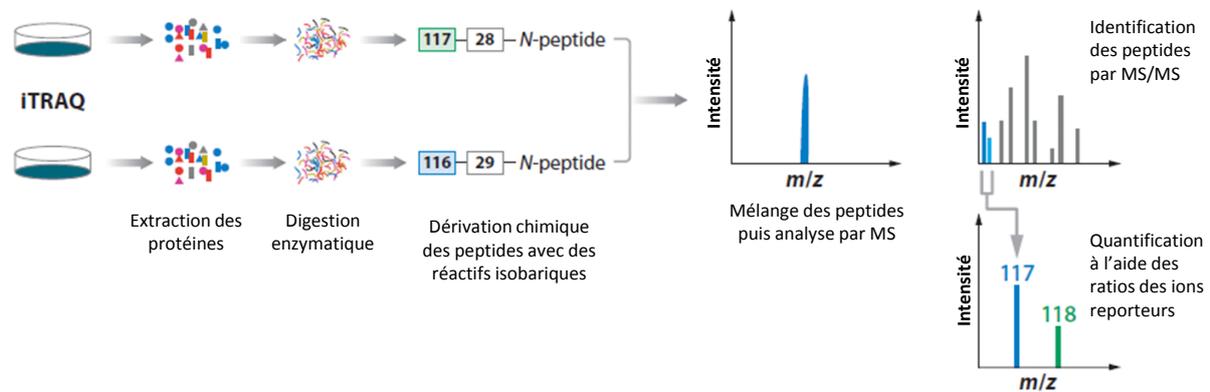
Une autre approche appelée ICPL (« Isotope-Coded Protéine Label ») [255] consiste à modifier les fonctions amines des lysines et des extrémités N-terminales. Cette technique permet d'obtenir plus de peptides modifiés (la lysine étant un acide aminé abondant), et élargit le nombre de peptides utilisés pour la quantification. Cependant, lorsque le marquage est effectué au niveau protéique, la trypsine ne coupe plus les lysines modifiées ce qui conduit à l'obtention de peptides de grande taille difficiles à analyser en MS.

En générale, les approches par marquage isotopique sont bien adaptées à la quantification relative globale des protéines et sont assez versatiles, cependant elles présentent certaines limitations. Le rendement du marquage est un paramètre important pour l'obtention d'une quantification fiable. De plus, des réactions parasites peuvent être observées conduisant à des marquages non spécifiques ayant des effets indésirables sur la fiabilité de la quantification et sur l'interprétation des données [188].

### ***3.3.2. Le marquage isobarique***

Cette technique consiste à marquer les amines N-terminales des peptides avec des réactifs isobariques qui ont la même masse mais dont la fragmentation génère des fragments spécifiques appelés ions reporteurs. Les peptides homologues ainsi marqués auront la même masse moléculaire et sont mélangés après marquage ; ils coélueront parfaitement en chromatographie. Un seul ion sera présent sur le spectre MS et un seul m/z sera donc isolé pour la fragmentation. La quantification se fait donc à l'aide des rapports d'intensité MS/MS des ions reporteurs. Ces ions reporteurs ont des faibles m/z (de 114 à 118) évitant ainsi l'interférence avec les fragments du squelette peptidique [9]. L'approche la plus connue est l'approche « iTRAQ » (« isobaric Tag for Relative and Absolute Quantification ») [256] (Figure 4). Dans sa version commerciale, elle permet de réaliser simultanément la quantification relative de 8 échantillons grâce à 8 réactifs isobariques [257]. Ceci est particulièrement intéressant pour l'étude de l'évolution d'un système biologique à différents points dans le temps ou plus généralement pour comparer différents traitements ou conditions biologiques dans une même expérience. Une autre

approche similaire appelée TMT (« Tandem Mass Tag ») a été décrite [258] et est aussi commercialisée.



**Figure 4 : Stratégie de marquage isobarique, après digestion enzymatique, les amines N-terminales des peptides sont marquées avec des réactifs isobariques, puis les échantillons sont mélangés et analysés par LC-MS/MS en mode DDA. Les spectres MS/MS ainsi générés permettent l'identification des peptides et la quantification relative se fait à l'aide des ions rapporteurs (adaptée de [225]).**

Dans cette approche, la précision de la quantification dépend largement de la fenêtre d'isolation des ions à fragmenter puisque tous les ions dans cette fenêtre contribueront à la génération d'ions rapporteurs. Un des inconvénients de cette technique est que la dérivation se fait au niveau peptidique, et donc toutes les variabilités en amont de la digestion ne sont pas prises en compte pour la quantification.

### 3.4. Quantification absolue ciblée avec les stratégies de dilution isotopique

L'utilisation de peptides de synthèse marqués aux isotopes stables en tant qu'étalons pour la quantification par spectrométrie de masse remonte à l'année 1983 et a été publiée par Desiderio et al. [259]. Récemment, et avec les avancées en analyse protéomique, cette approche est utilisée pour la quantification absolue et est connue sous la dénomination AQUA pour « Absolute QUAntification » [260]. Contrairement aux approches de quantification globale décrites ci-dessus, il est indispensable de déterminer les protéines cibles avant d'aborder la quantification. Les peptides synthétique de référence sont choisis en fonction de la ou des protéine(s) d'intérêt à quantifier. La stratégie la plus simple consiste à ajouter des quantités connues de peptides synthétiques marqués à un digest de protéines et de comparer les intensités de ces peptides marqués avec celles de leurs homologues endogènes. Bien qu'ils soient onéreux, la disponibilité des peptides AQUA dans le commerce avec des puretés élevées a rendu cette approche attractive dans le cadre de l'analyse et de la validation de biomarqueurs potentiels d'intérêt clinique [261]. Le principal inconvénient de cette approche est que, dans le cadre d'une analyse protéomique, les échantillons sont très complexes et il y a un risque de coélution de peptides isobariques ou de masses très proches qui viennent fausser la quantification. Le choix des peptides étalons est aussi très importants, ces peptides doivent être protéotypiques de la protéine à quantifier et donc non partagés avec d'autres protéines issues du même protéome. De plus, il est difficile de déterminer la quantité d'étalon à ajouter lorsque la gamme dynamique des protéines est très étendue (10 ordres de grandeur pour les tissus sanguins). Les peptides de référence

étant ajoutés après la digestion enzymatique, des variabilités dues à la préparation d'échantillon peuvent être observées. Une stratégie qui consiste à exprimer une protéine chimérique marquée aux isotopes stables et contenant dans sa séquences des peptides concatémères des peptides à quantifier permet de surmonter ce dernier problème. Cette protéine chimérique peut être introduite plus tôt dans l'expérience (en amont de l'extraction des protéines) et donc limiter les variabilités dues aux étapes de fractionnement et de digestion. Cette stratégie fut introduite par Pratt et al. sous le nom « QconCAT » (« Quantification conCATemer ») [262]. Toutefois, Rivers et al. ont observé que les rendements de digestion de ces protéines chimériques QconCAT sont différents de ceux des protéines cibles endogènes [263]. Une autre alternative est la stratégie « PSAQ » (« Protein Standard Absolute Quantification ») qui consiste à utiliser une protéine recombinante, ayant des peptides marqués aux isotopes stables, et dont la séquence en acides aminés est identique à celle de la protéine d'intérêt [264]. La mise en œuvre de ces stratégies reste néanmoins difficile et onéreuse.

### 3.5. La quantification absolue combinant la dilution isotopique à la SRM

La méthode de quantification absolue alliant spécificité, gamme dynamique et sensibilité reste la combinaison de la stratégie de dilution isotopique avec la SRM (« Selected Réaction Monitoring ») à l'aide d'un spectromètre de masse de type triple quadripôle. Cette méthode est depuis longtemps utilisée pour l'analyse et la quantification de petites molécules notamment dans des applications cliniques. Cependant, l'analyse ciblée des protéines par SRM se distingue de celles des petites molécules par plusieurs aspects dont le plus important est que l'analyse ne se fait pas directement sur les protéines et nécessite de passer par une digestion enzymatique. La quantification des protéines est donc effectuée au niveau des peptides. De ce fait, le passage par plusieurs étapes de préparation est nécessaire afin de produire et sélectionner ces peptides indispensables à l'analyse. Les autres aspects qui distinguent la quantification des protéines par SRM sont la complexité et la gamme dynamique du protéome. A ceci s'ajoute le fait que, dans le cas de la protéomique quantitative, la matrice et les analytes sont de même nature, à savoir des peptides ce qui entraîne des risques en termes d'interférences, de suppression d'ions et de limite de détection [265].

Typiquement, dans une analyse SRM, une liste des peptides protéotypiques est définie. Les peptides protéotypiques d'une protéine sont des peptides dont la séquence n'est partagée avec aucune autre protéine susceptible d'être présente dans le mélange à analyser. Ce sont les ions correspondants à ces peptides qui seront sélectionnés par le premier quadripôle pour être analysés par la suite. L'ion précurseur ainsi sélectionné sera fragmenté en mode CID dans la cellule de collision puis un ou plusieurs fragments seront sélectionnés par le troisième quadripôle (Figure 5). Plusieurs couples précurseur/fragment peuvent être suivis dans un cycle de mesure [265]. Ces couples précurseur/fragment appelés transitions sont très spécifiques du peptide protéotypique et par suite de la protéine à doser. L'ajout de versions marquées aux isotopes stables de ces peptides protéotypiques

tels que les peptides AQUA [260], QconCAT [262] ou PSAQ [264] en tant que standard interne permet la quantification absolue des protéines [266].

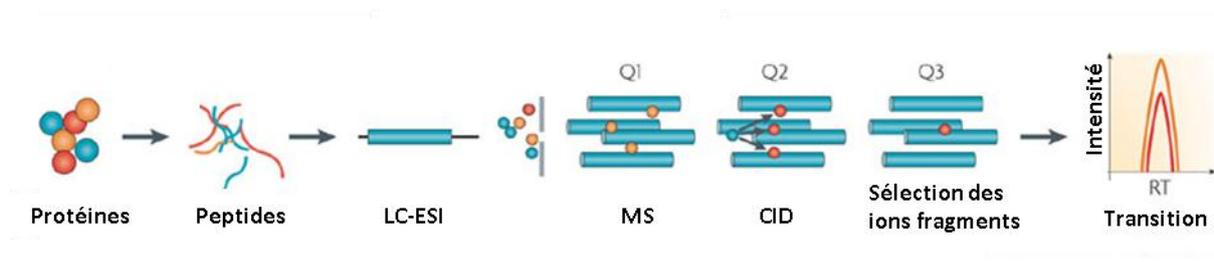


Figure 5 : Illustration du principe de la SRM sur un triple quadrupôle (adaptée de [266]).

Le choix des protéines d'intérêt à quantifier dans un protéome donné est basé sur des études protéomiques préalables. D'un autre côté, le choix des peptides et des transitions à suivre pour chaque protéine est assez compliqué et requiert une attention particulière. Ces peptides doivent satisfaire certains critères, le premier étant leur protéotypicité, le deuxième étant qu'ils soient facilement détectés en LC-MS. Des bases de données de peptides provenant de données protéomiques existent et facilitent ce choix. Ces bases de données telles que PeptideAtlas [267] et PRIDE [268] sont en croissance constante et constituent une mine d'or pour la conception d'analyse en SRM et le choix des peptides. En cas d'absence de données préalables, des algorithmes permettant la prédiction des propriétés physico-chimiques des peptides tels que ESP predictor [269], PeptideSieve [270], PepFly [271] et STEPP [272] orientent la sélection des peptides. Un des problèmes souvent rencontrés est lorsque la séquence du peptide choisi contient un acide aminé susceptible d'être modifié chimiquement ou de subir des modifications post-traductionnelles qui viennent fausser la quantification puisque le peptide sera ainsi présent sous différentes formes dans l'échantillon. C'est pour cela qu'il est préférable d'utiliser plusieurs peptides par protéine si une quantification absolue précise est souhaitée [265].

Après le choix des peptides, c'est au tour des transitions d'être sélectionnées. Cette étape est critique dans la mesure où les transitions sélectionnées doivent assurer une sensibilité et une spécificité optimale. Là aussi, les données protéomiques contenues dans les bases mentionnées dans le paragraphe précédent sont très utiles. Une base récente et particulièrement intéressante appelée SRMAtlas [273] contient des transitions optimisées. Des peptides de synthèse de type AQUA sont utilisés pour optimiser ces transitions et paramétrer le spectromètre de masse afin d'améliorer leurs intensités et leurs détections et par suite la sensibilité [265].

Le temps de rétention peut aussi être utilisé en tant que paramètre supplémentaire pour suivre les transitions. Ainsi, l'analyse est divisée en plusieurs segments de temps et dans un segment donné, seulement les transitions des peptides correspondant à ce temps d'élution sont suivies. Ce mode d'acquisition appelé « Time-scheduled SRM » permet d'augmenter le nombre de peptides suivis à quelques centaines dans une analyse LC-MS sans sacrifier l'échantillonnage ni la sensibilité. Afin d'augmenter davantage le nombre de

peptides qui peuvent être quantifiés en SRM, la i-SRM (« intelligent SRM ») a été récemment introduite par Kiyomani et coll. [274]. Celle-ci consiste à utiliser deux modes d'acquisition, le premier étant « composé spécifique » et dans lequel seulement deux transitions primaires sont suivies et le second étant « données dépendant » et dans lequel les 2 transitions primaires et six transitions secondaires supplémentaires sont mesurées. Seules les deux transitions primaires sont utilisées pour la quantification et sont continuellement suivies dans la fenêtre de temps de rétention prédéfinie ; tandis que l'ensemble des transitions servent à constituer un spectre composite qui permet de valider l'identité du peptide quantifié par les 2 transitions primaires. Ceci est plus avantageux que l'utilisation d'un spectre « full scan » ou d'un nombre supérieur de transition primaire qui nécessite un temps d'acquisition plus important pour la validation de l'identité du peptide [265].

#### **4. Les trois grandes stratégies d'analyse protéomique quantitative par spectrométrie de masse**

Nous avons vu dans ce chapitre les principales techniques de quantification en analyse protéomique. Ces techniques peuvent être utilisées dans différentes stratégies d'analyse protéomique en fonction de la problématique biologique posée. Les stratégies de quantification en analyse protéomique par spectrométrie de masse peuvent être réparties en trois classes comme proposé par Domon et Aebersold [189]: Les stratégies « shotgun » ou découverte, les stratégies basées sur la MS dirigée ou pilotée (« directed MS ») et les stratégies basées sur la MS ciblée (« targeted MS »).

##### **4.1. Les stratégies « shotgun » ou découverte**

Les stratégies dites de découverte consistent à analyser par LC-MS/MS le digest d'un protéome d'intérêt en mode données dépendant (DDA). Le spectromètre de masse effectue des cycles de balayage MS et MS/MS en continu générant un grand nombre de spectres MS/MS qui servent par la suite à séquencer les peptides afin de les identifier. La sélection des peptides pour la MS/MS se fait de manière heuristique lors des balayages MS. La quantification va de pair avec l'identification dans ce type de stratégies, elle est effectuée dans « la foule » à l'aide des données MS (méthodes SILE) ou MS/MS (par exemple : iTRAQ, spectral counting). La précision des données de quantification de ces stratégies est limitée. Ces stratégies sont souvent utilisées de façon qualitative dans le but d'identifier un grand nombre de protéines dans un échantillon complexe. Etant donné qu'aucune connaissance préalable des protéines contenues dans l'échantillon n'est nécessaire, ces stratégies sont bien adaptées pour la réalisation d'expériences de découverte générale. La limitation principale de cette méthode est dans le biais induit par la nature heuristique du procédé de sélection des ions précurseurs et la sélection préférentielle des ions les plus abondants. Ceci est d'autant plus accentué lorsqu'il s'agit d'échantillons complexes dans lesquels le nombre des analytes est beaucoup plus important que la capacité de pic du système LC-MS induisant des résultats d'identification peu reproductibles [189].

#### 4.2. Les stratégies basées sur la MS dirigée

Les stratégies basées sur la MS dirigée consistent à effectuer des analyses en deux temps. Dans un premier temps, une première analyse LC-MS ou LC-MS/MS est effectuée afin de sélectionner les éléments pour constituer une liste d'inclusion. Les informations collectées dans cette première analyse sont utilisées pour construire l'analyse du même échantillon dans un second temps. Ces stratégies sont donc moins biaisées par la sélection heuristique et préférentielle des composés les plus abondants. L'utilisation de la liste d'inclusion permet de cibler un ensemble d'ions d'intérêt. De plus, la dissociation des étapes d'identification et de quantification permet d'obtenir des données de quantification plus fiables [189].

#### 4.3. Les stratégies de MS ciblée

Dans les stratégies dites de MS ciblées, la construction de l'expérience est basée sur des informations et des données issues d'expériences préalables. La méthode la plus souvent utilisée dans ce type de stratégie est la SRM. Ces stratégies ciblent donc un ensemble de protéines ou de peptides d'intérêt particulier. Les détails du déroulement des expériences de SRM sont décrits dans le paragraphe 3.5 de ce chapitre. Parmi toutes les stratégies de quantification en analyse protéomique, les stratégies de MS ciblée permettent d'obtenir les quantifications les plus précises et les plus reproductibles. De plus ces stratégies sont les plus sensibles et possèdent les limites de détection et les limites de quantification les plus faibles [189].

#### 4.4. Les facteurs clés à prendre en compte dans la construction d'une expérience de protéomique quantitative

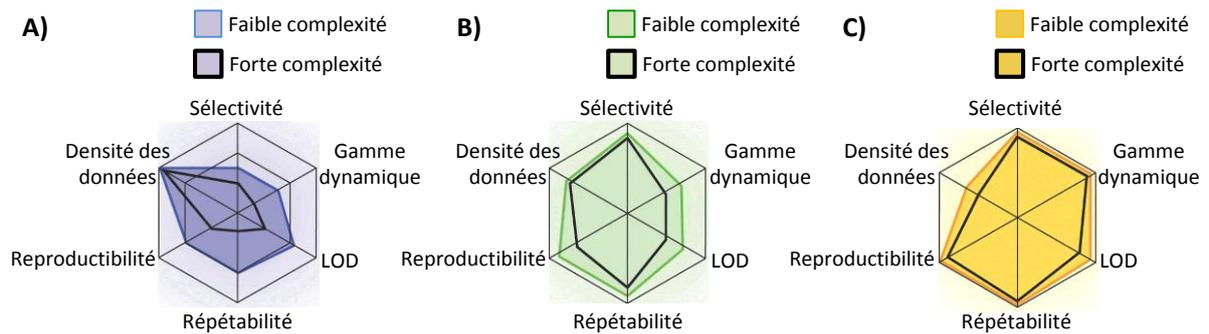
Lors de la réalisation d'une étude protéomique quantitative, plusieurs facteurs clés doivent être considérés. L'importance à accorder à chacun de ces facteurs dépend du type de résultats recherchés [189]. Ces facteurs sont au nombre de six:

- La sélectivité : c'est la capacité d'une méthode à identifier et quantifier un analyte particulier dans un échantillon complexe sans interférence avec les autres composés présents dans l'échantillon.
- La limite de détection (LOD): la LOD d'une méthode est la quantité minimale d'un analyte donné capable d'être détectée avec fiabilité. La limite de quantification étroitement liée à la LOD est définie comme étant la quantité minimale d'un analyte permettant sa quantification avec fiabilité. LOD et LOQ dépendent non seulement de la méthode et des propriétés physico-chimiques de l'analyte mais aussi de la complexité de la matrice dans laquelle il se trouve.
- La gamme dynamique : la gamme dynamique d'un instrument est la gamme d'intensité entre le signal le plus intense et le signal le moins intense d'un analyte dans une seule et même analyse.
- La densité des données : c'est le nombre de mesures acquises dans une expérience. Dans une analyse « shotgun » c'est le nombre de spectres MS/MS qui ont conduit à

des identifications. Dans une expérience MS dirigée ou MS ciblée, c'est le nombre de peptides analysés et quantifiés.

- La répétabilité : C'est l'aptitude d'une méthode à générer des résultats identiques pour des échantillons identiques analysés dans les mêmes conditions.
- La reproductibilité : c'est l'aptitude à obtenir des résultats identiques pour des échantillons identiques analysés par la même méthode mais dans des conditions différentes (expérimentateurs différents, appareils différents, ...).

La figure 6 représente ses six facteurs sur six axes sur lesquels les caractéristiques des trois stratégies décrites ci-dessus sont reportées.



**Figure 6 : Profils des performances classées selon 6 axes représentant les 6 facteurs clés de la quantification en analyse protéomique. (A) pour la stratégie « shotgun » ou découverte, (B) pour la stratégie MS dirigée et (C) pour la stratégie MS ciblée (adaptée de [189]).**



## Chapitre 4 L'analyse des sites de clivage protéolytique des protéines : Etat actuel de l'art

La protéolyse est l'hydrolyse irréversible de la liaison peptidique ou isopeptidique par l'action d'une protéase. La protéolyse des protéines est une modification post-traductionnelle qui affecte toute protéine à un moment ou un autre de son cycle de vie. Elle consiste en la dégradation totale des protéines, mais aussi en des clivages précis et limités à des sites spécifiques. Ces clivages protéolytiques spécifiques modulent la structure, la fonction et la localisation des protéines [275]. En effet, les protéases interviennent dans un grand nombre de processus biologiques, tels que la réplication et la réparation de l'ADN, la mitose, la prolifération cellulaire, la différenciation, la migration, l'hémostase, l'apoptose, etc. [276] Dans le génome humain, 570 gènes sont prédits codant pour des protéases, néanmoins, seulement une minorité de celles-ci a été caractérisée [277]. Près de 700 protéases/peptidases humaines sont répertoriées dans les banques de données [278] et représentent environ 1,3 % du protéome humain [275]. Le dérèglement de l'activité et de la distribution des protéases ou de leurs inhibiteurs sont associés à plusieurs pathologies, tel que le cancer [279-282], les maladies inflammatoires [283], l'Alzheimer [284,285] et les interactions hôte/pathogène [286-288]. Par conséquent, de même pour les modifications post-traductionnelles, si la protéolyse n'est pas caractérisée dans les analyses protéomiques, une quantité considérable d'informations est alors perdue et les relations structure/activité sont imprécises voir erronées [289].

En 2002, Lopez-Otin et Overall ont introduit le terme dégradomique [290] qualifiant l'étude du dégradome d'un nouveau challenge pour l'analyse protéomique. La dégradomique est l'étude des protéases, de leurs inhibiteurs, de leurs substrats et leurs sites de clivages ainsi que la manière dont l'activité protéolytique influence l'organisme, l'organe, le tissu ou la cellule. Dans ce contexte, la caractérisation des sites de clivages des protéines par les protéases relève d'une importance majeure. En effet, le site de clivage détermine la structure primaire du substrat clivé et par suite sa fonction biologique. La position exacte du site de clivage d'une protéine peut avoir une influence majeure sur son activité biologique, en particulier dans le cas où plusieurs sites se trouvent l'un à proximité de l'autre. Un bon exemple est le cas du clivage protéolytique de la protéine APP (« Amyloid Precursor Protein ») qui peut conduire à deux protéines différentes de seulement deux acides aminés, l'une est soluble tandis que l'autre est l'espèce prédominante des dépôts observés dans les

cerveaux atteints d'Alzheimer [284]. De plus, la détermination des sites de clivage permet de mieux prédire l'activité des protéases et d'identifier des séquences consensus de clivage pour certaines d'entre elles [275]. Au cours des dernières années, plusieurs méthodes et stratégies ont été développées en analyses protéomiques dans le but d'étudier le « dégradome ». Ces méthodes visent essentiellement à identifier, caractériser et quantifier la nouvelle extrémité N-terminale des protéines clivées. Dans ce chapitre, nous décrivons les principales méthodes développées pour l'étude des sites de clivage protéolytique. Il est nécessaire de distinguer les méthodes dites de sélection positive, qui visent à cibler directement les peptides N-terminaux dans l'étape de purification, tandis que les méthodes dites de sélection négative visent à éliminer les peptides internes issus de la digestion enzymatique (trypsine) lors de la préparation d'échantillon.

## 1. Méthode basée sur une sélection positive des peptides N-terminaux

### 1.1. Stratégie d'enrichissement des peptides N-terminaux par biotinylation des $\alpha$ -amino-groupes

Afin d'identifier les substrats des protéases et les sites de clivage protéolytique, l'équipe de Salvesen a proposé une méthode qui consiste à effectuer une réaction de guanidination des amines  $\epsilon$  (amines des chaînes latérales des lysines) suivie d'une dérivation des amines- $\alpha$  libres par un groupement N-HydroxySuccinimide (NHS)-biotine clivable [291]. Après une étape de digestion enzymatique, les peptides N-terminaux marqués par la biotine sont retenus sur une colonne streptavidine. Ils sont ensuite libérés par clivage du groupement biotine à l'aide du dithiothreitol et analysés par LC-MS/MS (Figure 1).

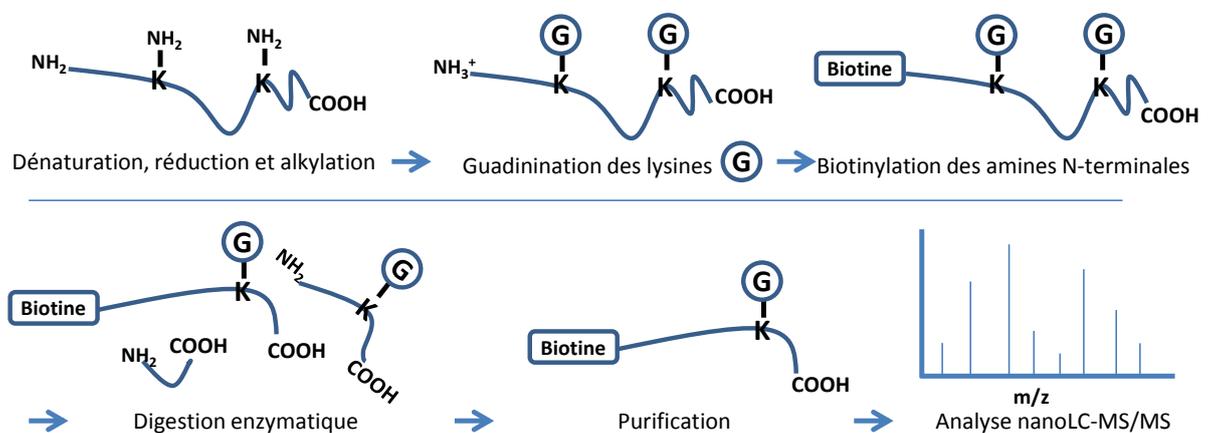


Figure 1 : Enrichissement positif des peptides N-terminaux par dérivation avec un groupement biotine

Cette méthode ne permet néanmoins pas d'analyser les protéines acétylées en position N-terminales et est limitée à la caractérisation des extrémités N-terminales libres. De plus, l'absence de marquage isotopique ne permet pas de déterminer le taux de protéolyse et limite la méthode à la description de l'état protéolytique de l'échantillon. Cette technique souffre aussi d'un problème de contamination par des peptides internes non-issus de clivages protéolytiques et qui sont biotinylés dû à une guanidination incomplète des

lysines ou à des réactions parasites sur les chaînes latérales des sérines, thréonines et histidines. Ceci conduit à l'identification de faux positifs [289].

### **1.2. Stratégie de dérivation par le réactif du séquençage d'Edman**

Cette stratégie appelée N-CLAP (« N-terminalomics by Chemical Labeling of the  $\alpha$ -amine of Proteins ») consiste à bloquer toutes les amines primaires ( $\alpha$  et  $\epsilon$ ) par dérivation, à l'aide du réactif d'Edman, le phényl-isothiocyanate (PITC) [292]. L'échantillon est ensuite traité avec de l'acide trifluoroacétique qui induit une cyclisation intramoléculaire des amines  $\alpha$  dérivées au PITC. Ceci conduit au clivage l'acide aminé N-terminal générant ainsi une nouvelle amine  $\alpha$  libre. Les amines  $\epsilon$  des lysines dérivées au PITC restent intactes et ne sont pas affectées. Les nouvelles amines libres sont ensuite biotinylées à l'aide d'un réactif contenant un linker clivable et les peptides biotinylés sont capturés sur des billes de streptavidine. Après élimination des peptides internes, les peptides N-terminaux sont libérés par clivage du linker et analysés par LC-MS/MS. Comme pour la méthode précédente, cette technique ne permet pas de déterminer les taux de protéolyse et d'effectuer des études comparatives ce qui la limite à la description de l'état protéolytique de l'échantillon.

### **1.3. La stratégie PICS (« Proteomic Identification of Cleavage Sites ») d'identification des sites de clivage à partir d'une librairie de peptides naturels**

Cette stratégie développée dans le laboratoire d'Overall se distingue des autres techniques décrites dans ce chapitre par l'utilisation d'une librairie de peptides naturels issus du protéome étudié au lieu de travailler directement sur les protéines substrats. Les protéines sont tout d'abord digérées à la trypsine ou à l'endoprotéinase GluC générant ainsi la librairie de peptides. Après réduction et alkylation, les amines ( $\alpha$  et  $\epsilon$ ) libres sont bloquées par méthylation en milieu réducteur. La librairie de peptides est ensuite incubée avec la protéase d'intérêt et les amines néo-N-terminales ainsi générées par clivage protéolytiques sont biotinylées. Les peptides néo-N-terminaux sont ensuite purifiés à l'aide de billes d'avidine et analysés par LC-MS/MS afin de déterminer les sites de clivage par la protéase (Figure 2) [293].

Cette technique a été appliquée avec succès dans une étude des sites de clivage de 4 métalloprotéinases contenues dans le venin de vipère [294] et a permis de caractériser entre 23 et 57 substrats peptidiques et des séquences préférentielles de clivage. Cependant, la modification chimique (alkylation et méthylation) des peptides avant l'action de la protéase étudiée peut conduire à des interférences avec l'activité de la protéase qui ne reconnaît plus son site de clivage et induit donc de vrais négatifs. Étant donné que cette stratégie utilise des peptides substrats et non des protéines dans leurs formes natives, les effets des structures secondaires et tertiaires des protéines dans leur milieu sur l'activité de la protéase et l'accessibilité des sites de clivage ne sont pas pris en compte. Cette stratégie permet néanmoins de réaliser un profilage rapide des sites de clivage préférentiels de la protéase.

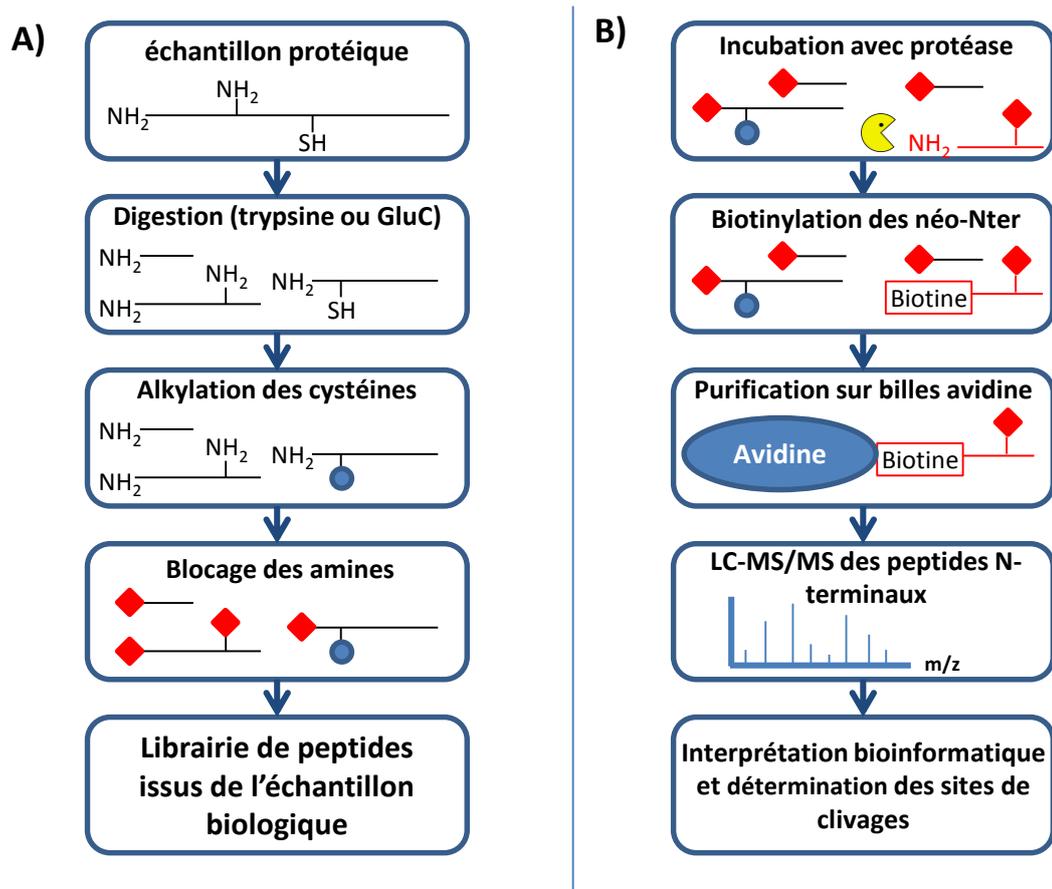


Figure 2 : La stratégie PICS, A) génération de la librairie de peptides naturels issus de la digestion enzymatique du protéome d'intérêt. B) identification des sites de clivage par LC-MS/MS après incubation de la librairie de peptides avec la protéase d'intérêt et biotinylation des amines  $\alpha$  néo-N-terminales.

Une autre technique développée dans le même laboratoire et utilisant des peptides substrats a été récemment décrite [295]. Cette approche appelée ATOMS (« Amino-Terminal Oriented Mass Spectrometry of Substrates ») utilise, contrairement à PICS, un digest d'une protéine substrat individuelle sur lequel on fait réagir la protéase d'intérêt et non des librairies complexes de peptides. Le digest de la protéine substrat d'intérêt est incubé avec la protéase ou avec le tampon (contrôle). Ensuite les amines libres sont dérivées avec du formaldéhyde marqué aux isotopes stables afin de pouvoir comparer l'échantillon traité avec le contrôle. Cette étape de dérivation permet de marquer toutes les amines ( $\alpha$  et  $\epsilon$ ) y compris les amines néo-terminales générées par la protéase. Tous les peptides seront présents lors de l'analyse LC-MS/MS sous forme de doublet (lourd et léger), sauf les peptides néo-N-terminaux qui seront sous forme de singlet, ce qui permet de les identifier. Cette technique, comparée par les auteurs au séquençage d'Edman, n'a cependant pas été appliquée sur des échantillons complexes.

## 2. Méthodes basées sur la sélection négative des peptides N-terminaux

### 2.1. La méthode COFRADIC (« COmbined FRActional DIagonal Chromatography »)

Cette méthode développée par l'équipe de J. Vandekerckhove est sans doute la méthode de référence dans le domaine de l'analyse des sites de clivage protéolytique. Elle consiste tout d'abord à acétyler toutes les amines primaires libres ( $\alpha$  et  $\epsilon$ ) des protéines à l'aide de l'anhydride acétique. Après digestion à la trypsine, les peptides sont séparés par RP-HPLC et les fractions sont collectées. Les amines libres des peptides contenus dans ces fractions et issus de la digestion trypsique sont ensuite bloquées à l'aide de l'acide 2,4,6-trinitrobenzenesulfonique (TNBS) qui induit une augmentation de l'hydrophobicité des peptides. Une deuxième étape de RP-HPLC est réalisée pour chacune des fractions collectées lors de la première étape de chromatographie. Les peptides internes dérivés au TNBS éluent tardivement et sont éliminés, tandis que les peptides N-terminaux acétylés sont élués au même temps de rétention que dans la première étape et sont ainsi collectés puis analysés par nanoLC-MS/MS [296] (Figure 3).

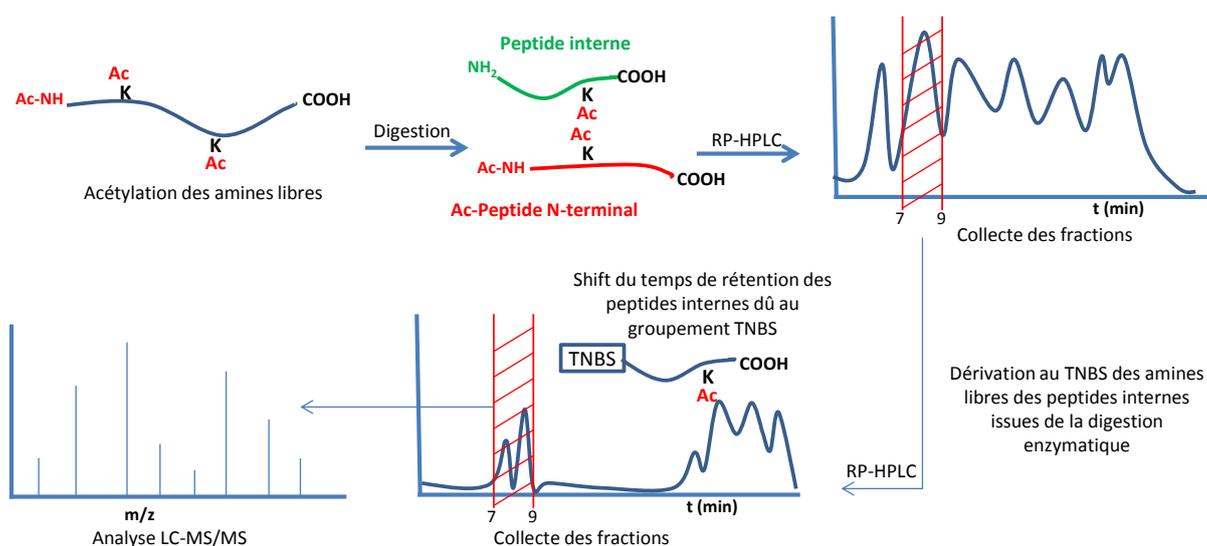


Figure 3 : Illustration de la stratégie COFRADIC pour l'enrichissement en peptides N-terminaux et caractérisation des sites de clivage protéolytique.

Plusieurs améliorations ont ensuite été apportées à cette stratégie. L'incorporation d'isotopes stables à l'aide d'une digestion différentielle à la trypsine dans de l'eau et de l'eau lourde ( $H_2^{18}O$ ) permet une quantification relative des taux de protéolyse que subit chaque protéine [297]. Un marquage chimique aux isotopes stables grâce à une trideutéro-acétylation [298] permet de différencier l'acétylation naturelle par les N-acétyl-transférases des acétylations induites chimiquement dans le protocole expérimental [299]. L'utilisation du marquage métabolique SILAC à l'aide d'arginine lourde a aussi été décrit avec cette technique et permet d'effectuer des études de quantification différentielle des peptides N-terminaux [300]. Une autre amélioration consiste à introduire une étape de chromatographie d'échange de cations (SCX) à pH = 3 après l'étape de digestion trypsique.

Les peptides dont les amines sont bloquées sont moins retenus en SCX que les peptides ayant des amines libres. Cette étape de SCX a permis d'augmenter significativement le nombre de peptides N-terminaux identifiés. Une étape permettant l'élimination des peptides de digestion trypsique avec une extrémité N-terminal pyroglutamyl qui ne réagit pas avec le TNBS a aussi été ajoutée [298].

La stratégie COFRADIC est donc une approche puissante pour la caractérisation des extrémités N-terminales des protéines dans des échantillons biologiques complexes [300-303]. Cependant, la procédure expérimentale est assez fastidieuse et compliquée avec deux étapes de dérivation chimique et deux étapes de séparation avec collecte de fractions conduisant à plus de 96 analyses LC-MS/MS par échantillon. Elle nécessite un temps machine important et souffre probablement d'importantes pertes de matériels dues au nombre important d'étapes de préparation.

### **2.2. Stratégie de sélection négative des peptides N-terminaux par biotinylation des peptides internes**

Cette stratégie consiste à acétyler les amines primaires libres ( $\alpha$  et  $\epsilon$ ) des protéines et, après digestion trypsique, à biotinyler les amines  $\alpha$  des peptides internes de digestion. Le mélange de peptides est ensuite incubé avec des billes de streptavidine immobilisée. Ceci permet d'éliminer les peptides internes diminuant ainsi la complexité de l'échantillon [304]. Une amélioration de cette technique fut introduite en éliminant l'étape de biotinylation et en remplaçant les billes de streptavidine par un groupement aminoréactif immobilisé sur sépharose (Sépharose NHS-activée) [305]. Les peptides internes sont ainsi directement capturés sur les billes de sépharose par leur amine libre. Par rapport à la stratégie COFRADIC, cette technique est simple et rapide tout en étant efficace, néanmoins elle n'est pas adaptée aux études différentielles de processus protéolytique ou à la quantification des taux de protéolyse.

### **2.3. Stratégie d'enrichissement *in silico* en peptides N-terminaux à l'aide du marquage iTRAQ**

Cette stratégie fut introduite par le groupe de Salvesen en 2007. Elle consiste à traiter un échantillon de protéines avec la protéase d'intérêt active et à le comparer avec le même échantillon traité avec la même protéase inactivée. Les amines N-terminales libres des protéines sont dérivées à l'aide de réactifs iTRAQ après guanidination des lysines. Les deux échantillons sont ensuite mélangés et les protéines sont digérées, fractionnées par chromatographie et analysées en deux temps par MALDI-TOF/TOF [306]. Dans un premier temps, un balayage MS/MS à faible énergie laser est effectué afin de détecter les peptides qui présentent des ions reporteurs issus de la fragmentation du tag iTRAQ. Une liste d'inclusion de ces peptides est ensuite générée afin de réaliser, dans un second temps, une analyse TOF/TOF-MS/MS à haute énergie laser permettant de fragmenter ces peptides N-terminaux et par la suite de les identifier. Le calcul des ratios des ions reporteurs permet d'identifier les extrémités N-terminales issus d'un clivage protéolytique (ratio  $\neq 1$ ) par rapport à celles des protéines sans clivage (ratio = 1) (Figure 4).

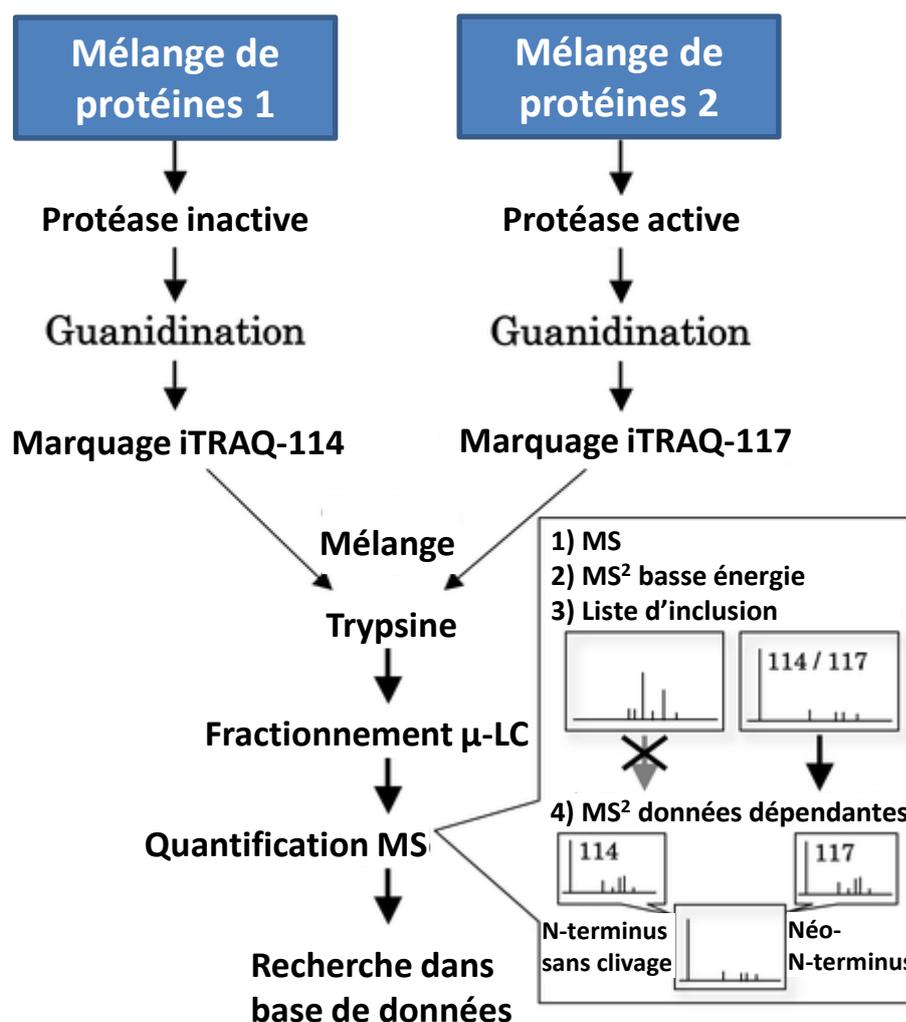


Figure 4 : Illustration de la stratégie d'enrichissement in silico à l'aide du marquage iTRAQ (Adaptée de [306]).

Si cette technique a été appliquée avec succès sur des échantillons peu complexes, son application à un échantillon complexe a permis d'identifier seulement 20 sites de clivage protéolytique [306]. Ceci est probablement dû à une forte suppression de signal étant donné le nombre important de peptides internes présents dans les dépôts MALDI.

#### 2.4. La stratégie TAILS (« Terminal Amine Isotope Labeling of Substrates »)

Dans cette stratégie, les protéomes de deux échantillons, l'un contenant la protéase active et l'autre la protéase inactive (contrôle), sont réduits, alkylés et puis les amines primaires libres ( $\alpha$  et  $\epsilon$ ) sont dérivées avec des réactifs marqués aux isotopes stables. Deux différents types de dérivation ont été décrits, l'un isotopique (formaldéhyde lourd et léger) pour une quantification en mode MS [307] et l'autre isobarique (iTRAQ) permettant une quantification en mode MS/MS [308]. Les protéines sont ensuite digérées à la trypsine, puis les peptides résultants sont mélangés. L'enrichissement en peptide N-terminaux est réalisé en faisant réagir les amines libres des peptides internes avec un polymère amino-réactif soluble qui permet de les éliminer par la suite. Les peptides N-terminaux sont analysés par LC-MS/MS et le calcul des ratios d'abondance permet de déterminer l'origine et la nature de

l'extrémité N-terminale. Un ratio de  $\sim 1$  indique une extrémité N-terminale originale non issue du clivage par la protéase, tandis qu'un singlet indique que l'extrémité N-terminale est issue d'un clivage par la protéase. L'utilisation du réactif iTRAQ (iTRAQ-TAILS) permet en plus de comparer jusqu'à huit échantillons dans une même expérience (Figure 5) [308].

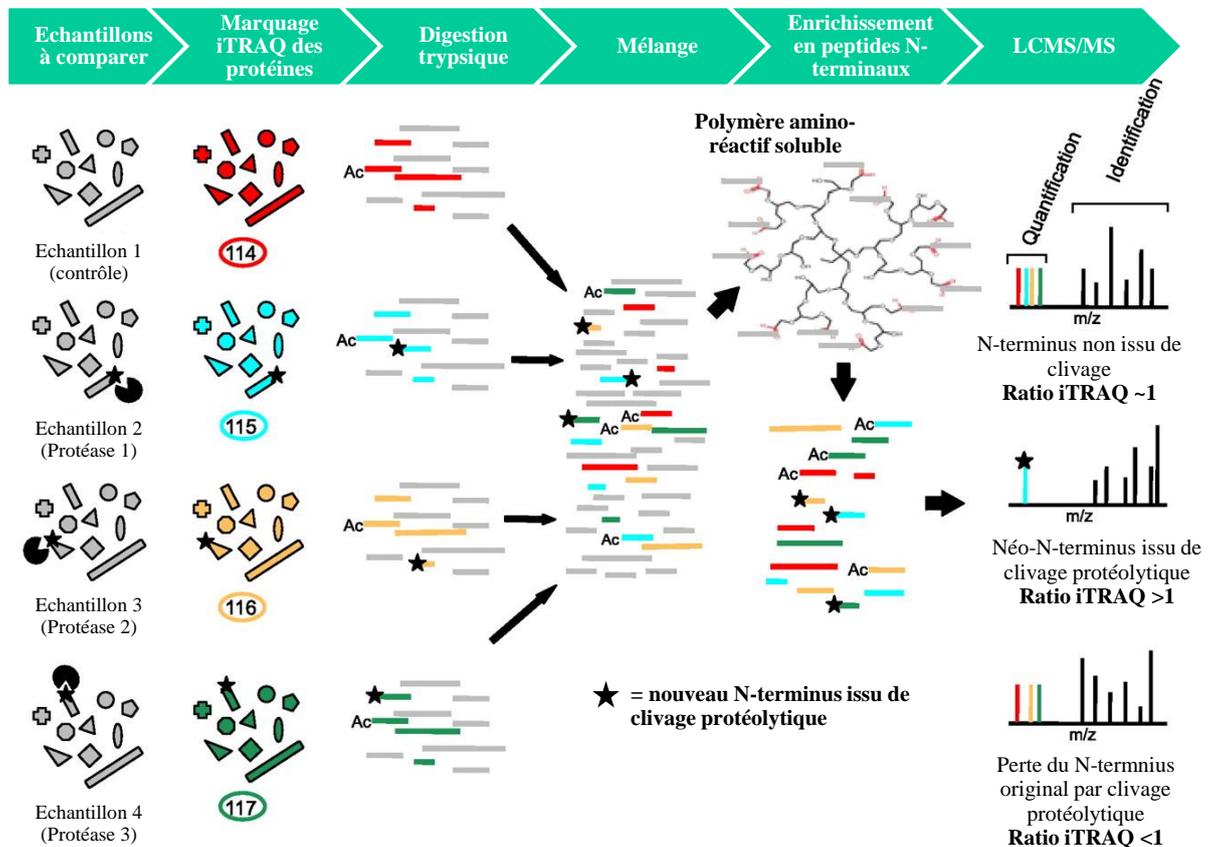


Figure 5 : Illustration de la stratégie iTRAQ-TAILS. Comparaison simultanée de quatre échantillons grâce au marquage différentielle iTRAQ et enrichissement par sélection négative des peptides N-terminaux (adaptée de [308]).

Cette approche permet, comme la stratégie COFRADIC, de caractériser et de quantifier les sites de clivage protéolytique et d'annoter les extrémités N-terminales des protéines non clivées (acétylées ou non acétylées) dans la même expérience. Elle présente l'avantage d'avoir moins d'étapes de préparation d'échantillon que COFRADIC et donc moins de perte d'échantillons et ne nécessite pas un temps machine considérable.

## 2.5. Synthèse

Généralement, les méthodes d'enrichissement en peptides N-terminaux par sélection négative sont préférées à celles par sélection positive, qui ne permettent pas de quantifier les peptides N-terminaux et donc ne permettent pas d'effectuer des études comparatives ou de déterminer les taux de protéolyses. De plus, toutes les stratégies d'enrichissement par sélection positive utilisent la biotinylation des amines  $\alpha$  pour purifier les peptides N-terminaux. Par suite, seules les extrémités N-terminales libres sont caractérisées, tandis que celles qui sont acétylées ou qui portent un pyroglutamyl sont exclues. Contrairement aux stratégies de sélection positive, les stratégies de sélection négative permettent d'analyser

tous les peptides N-terminaux. COFRADIC et TAILS restent les deux techniques qui permettent de réaliser les études les plus complètes des sites de clivages des protéines par les protéases et sont compatibles avec les échantillons complexes. COFRADIC est cependant une technique difficile à utiliser en routine étant donné le nombre important d'étapes de séparation, de fractionnements et d'analyses LC-MS/MS. Plusieurs revues des stratégies d'analyses des clivages protéolytiques sont disponibles dans la littérature [275,289,309].

Toutes les techniques que nous avons décrites dans ce chapitre modifient d'une façon considérable les protocoles d'analyses protéomiques. Elles conduisent toutes à l'élimination des peptides de digestion internes qui ne sont donc pas analysés. L'identification du substrat de la protéase est donc réalisée par le séquençage d'un seul peptide par MS/MS : le peptide N-terminal. Des interrogations sur la qualité des identifications des protéines avec le spectre MS/MS d'un seul peptide ont souvent été émises [310]. De plus, le peptide N-terminal est un peptide portant une seule extrémité spécifique de l'enzyme de digestion (l'extrémité C-terminale). L'identification des protéines à l'aide d'un seul peptide ayant une seule extrémité spécifique de l'enzyme de digestion peut conduire à un nombre important de faux positifs [311], surtout lorsqu'il s'agit d'analyses d'échantillons complexes qui génèrent un grand nombre de spectres MS/MS. L'identification de la protéine n'est donc pas renforcée par l'identification de plusieurs éléments de sa séquence. Ceci a conduit la communauté scientifique à élaborer des directives pour la publication de données d'analyses protéomiques. Ces directives précisent que le nombre de peptides participant à l'identification des protéines doit être systématiquement reporté [129,312].

Il n'existe pas aujourd'hui une méthode qui permet de caractériser avec précision les sites de protéolyses tout en conservant les peptides internes et qui soit compatible avec l'analyse protéomique à haut débit. Nous avons donc proposé et développé une nouvelle approche qui permet de conserver les peptides internes tout en permettant la caractérisation des peptides N-terminaux. Cette approche sera détaillée dans la partie II de ce manuscrit.



## Références bibliographiques

1. Edman, P., A method for the determination of amino acid sequence in peptides. *Arch Biochem* **1946**, *22* (3), 475.
2. Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., and Whitehouse, C. M., Electrospray ionization for mass spectrometry of large biomolecules. *Science* **1989**, *246* (4926), 64-71.
3. Karas, M., and Hillenkamp, F., Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem* **1988**, *60* (20), 2299-2301.
4. Tanaka, T., Slamon, D. J., Shimoda, H., Waki, C., Kawaguchi, Y., Tanaka, Y., and Ida, N., Expression of Ha-ras oncogene products in human neuroblastomas and the significant correlation with a patient's prognosis. *Cancer Res* **1988**, *48* (4), 1030-1034.
5. Domon, B., and Aebersold, R., Mass spectrometry and protein analysis. *Science* **2006**, *312* (5771), 212-217.
6. Colangelo, J., Licon, V., Benen, J., Visser, J., Bergmann, C., and Orlando, R., Characterization of the N-linked glycosylation site of recombinant pectate lyase. *Rapid Commun Mass Spectrom* **1999**, *13* (23), 2382-2387.
7. Cravatt, B. F., Simon, G. M., and Yates, J. R., 3rd, The biological impact of mass-spectrometry-based proteomics. *Nature* **2007**, *450* (7172), 991-1000.
8. Elliott, M. H., Smith, D. S., Parker, C. E., and Borchers, C., Current trends in quantitative proteomics. *Journal of Mass Spectrometry* **2009**, *44* (12), 1637-1660.
9. Schulze, W. X., and Usadel, B., Quantitation in Mass-Spectrometry-Based Proteomics. *Annual Review of Plant Biology* **2010**, *61* (1), 491-516.
10. Whitehouse, C. M., Dreyer, R. N., Yamashita, M., and Fenn, J. B., Electrospray interface for liquid chromatographs and mass spectrometers. *Anal Chem* **1985**, *57* (3), 675-679.
11. Rayleigh, L., On the equilibrium of liquid conducting masses charged with electricity. *Philosophical magazine* **1882**, *14*, 184-186.
12. Dole, M., Mack, L. L., Hines, R. L., Mobley, R. C., Ferguson, L. D., and Alice, M. B. (1968) *Molecular Beams of Macroions*, AIP
13. Thomson, B. A., and Iribarne, J. V. (1979) *Field induced ion evaporation from liquid surfaces at atmospheric pressure*, AIP
14. Kebarle, P., A brief overview of the present status of the mechanisms involved in electrospray mass spectrometry. *J Mass Spectrom* **2000**, *35* (7), 804-817.
15. Atmanène, C. (2009) Nouvelles perspective pour l'analyse et la caractérisation structurale de complexes biologiques par spectrométrie de masse supramoléculaire. in *Thèse de l'Institut Pluridisciplinaire Hubert Curien - Département des Sciences Analytiques*, Université de Strasbourg, Strasbourg
16. Kebarle, P., and Verkerk, U. H., Electrospray: from ions in solution to ions in the gas phase, what we know now. *Mass Spectrom Rev* **2009**, *28* (6), 898-917.
17. Hoffmann, P., and Stroobant, V. (2007) *Mass Spectrometry: Principles and Applications*, Wiley
18. Kebarle, P., and Peschke, M., On the mechanisms by which the charged droplets produced by electrospray lead to gas phase ions. *Analytica Chimica Acta* **2000**, *406* (1), 11-35.

19. De La Mora, J. F., aacute, ndez, and Loscertales, I. G., The current emitted by highly conducting Taylor cones. *Journal of Fluid Mechanics* **1994**, *260*, 155-184.
20. Wilm, M. S., and Mann, M., Electrospray and Taylor-Cone theory, Dole's beam of macromolecules at last? *International Journal of Mass Spectrometry and Ion Processes* **1994**, *136* (2-3), 167-180.
21. Wilm, M., and Mann, M., Analytical properties of the nanoelectrospray ion source. *Anal Chem* **1996**, *68* (1), 1-8.
22. Koster, S., and Verpoorte, E., A decade of microfluidic analysis coupled with electrospray mass spectrometry: an overview. *Lab Chip* **2007**, *7* (11), 1394-1412.
23. Hillenkamp, F., Karas, M., Beavis, R. C., and Chait, B. T., Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers. *Anal Chem* **1991**, *63* (24), 1193A-1203A.
24. Steen, H., and Mann, M., The ABC's (and XYZ's) of peptide sequencing. *Nat Rev Mol Cell Biol* **2004**, *5* (9), 699-711.
25. Zenobi, R., and Knochenmuss, R., Ion formation in MALDI mass spectrometry. *Mass Spectrometry Reviews* **1998**, *17* (5), 337-366.
26. [www.magnet.fsu.edu](http://www.magnet.fsu.edu).
27. Aebersold, R., and Mann, M., Mass spectrometry-based proteomics. *Nature* **2003**, *422* (6928), 198-207.
28. Campana, J. E., Elementary theory of the quadrupole mass filter. *International Journal of Mass Spectrometry and Ion Physics* **1980**, *33* (2), 101-117.
29. Ouyang, Z., Wu, G., Song, Y., Li, H., Plass, W. R., and Cooks, R. G., Rectilinear ion trap: concepts, calculations, and analytical performance of a new mass analyzer. *Anal Chem* **2004**, *76* (16), 4595-4605.
30. Ouyang, Z., Gao, L., Fico, M., Chappell, W. J., Noll, R. J., and Cooks, R. G., Quadrupole ion traps and trap arrays: geometry, material, scale, performance. *Eur J Mass Spectrom (Chichester, Eng)* **2007**, *13* (1), 13-18.
31. March, R. E., Quadrupole ion traps. *Mass Spectrom Rev* **2009**, *28* (6), 961-989.
32. Dehmelt, H. G. (1968) Radiofrequency Spectroscopy of Stored Ions I: Storage. in *Advances in Atomic and Molecular Physics* (Bates, D. R., and Immanuel, E. eds.), Academic Press. pp 53-72
33. Glish, G. L., and Vachet, R. W., The basics of mass spectrometry in the twenty-first century. *Nat Rev Drug Discov* **2003**, *2* (2), 140-150.
34. Béchade, G. (2009) Avancée en analyse protéomique par spectrométrie de masse: Recherche de biomarqueurs et caractérisation de complexes covalents. in *Thèse de l'Institut Pluridisciplinaire Hubert Curien - Département des Sciences Analytiques*, Université de Strasbourg, Strasbourg
35. Wiley, W. C., and McLaren, I. H. (1955) *Time-of-Flight Mass Spectrometer with Improved Resolution*, AIP
36. Krutchinsky, A. N., Chernushevich, I. V., Spicer, V. L., Ens, W., and Standing, K. G., Collisional Damping Interface for an Electrospray Ionization Time-of-Flight Mass Spectrometer. *Journal of the American Society for Mass Spectrometry* **1998**, *9* (6), 569-579.
37. Becker, C. H., and Gillen, K. T., Can nonresonant multiphoton ionization be ultrasensitive? *J. Opt. Soc. Am. B* **1985**, *2* (9), 1438-1443.
38. A.F. Dodonov, I. V. C., T.F. Dodonova, V.V. Raznikov, V. Tal'roze. (1987, February 25) USSR Patent N° 1681340A1. (Patent, U. ed.
39. Dawson, J. H. J., and Guilhaus, M., Orthogonal-acceleration time-of-flight mass spectrometer. *Rapid Communications in Mass Spectrometry* **1989**, *3* (5), 155-159.
40. Shaffer, S. A., Tang, K., Anderson, G. A., Prior, D. C., Udseth, H. R., and 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.

41. Shaffer, S. A., Tolmachev, A., Prior, D. C., Anderson, G. A., Udseth, H. R., and Smith, R. D., Characterization of an Improved Electrodynamical Ion Funnel Interface for Electrospray Ionization Mass Spectrometry. *Analytical Chemistry* **1999**, *71* (15), 2957-2964.
42. Hipple, J. A., Sommer, H., and Thomas, H. A., A Precise Method of Determining the Faraday by Magnetic Resonance. *Physical Review* **1949**, *76* (12), 1877.
43. Sommer, H., Thomas, H. A., and Hipple, J. A., The Measurement of  $e/m$  by Cyclotron Resonance. *Physical Review* **1951**, *82* (5), 697.
44. Cooley, J. W., and Tukey, J. W., An algorithm for machine calculation of complex Fourier series. *Math. Comp.* **1965**, *19*, 297-301.
45. Marshall, A. G., Hendrickson, C. L., and Jackson, G. S., Fourier transform ion cyclotron resonance mass spectrometry: a primer. *Mass Spectrom Rev* **1998**, *17* (1), 1-35.
46. Gates, P., Mass Spectrometry Resource. The Fourier-Transform Ion Cyclotron Resonance (FT-ICR). University of Bristol <http://www.chm.bris.ac.uk/ms/theory/fticr-massspec.html>.
47. Makarov, A., Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. *Anal Chem* **2000**, *72* (6), 1156-1162.
48. Hu, Q., Noll, R. J., Li, H., Makarov, A., Hardman, M., and Graham Cooks, R., The Orbitrap: a new mass spectrometer. *Journal of Mass Spectrometry* **2005**, *40* (4), 430-443.
49. [www.thermo.com](http://www.thermo.com). Orbitrap analyzer.
50. Kume, H., Koyama, K., Nakatsugawa, K., Suzuki, S., and Fatlowitz, D., Ultrafast microchannel plate photomultipliers. *Appl Opt* **1988**, *27* (6), 1170-1178.
51. Wiza, J. L., Microchannel Plate Detectors. *Nuclear Instruments and Methods* **1979**, *162*, 587-601.
52. Biemann, K., Appendix 5. Nomenclature for peptide fragment ions (positive ions). *Methods Enzymol* **1990**, *193*, 886-887.
53. Gaskell, S. J., Electrospray: Principles and Practice. *Journal of Mass Spectrometry* **1997**, *32* (7), 677-688.
54. Cox, K., Gaskell, S., Morris, M., and Whiting, A., Role of the site of protonation in the low-energy decompositions of gas-phase peptide ions. *Journal of the American Society for Mass Spectrometry* **1996**, *7* (6), 522-531.
55. Jones, J. L., Dongre, A. R., Somogyi, A., and Wysocki, V. H., Sequence Dependence of Peptide Fragmentation Efficiency Curves Determined by Electrospray Ionization/Surface-Induced Dissociation Mass Spectrometry. *Journal of the American Chemical Society* **1994**, *116* (18), 8368-8369.
56. Dongré, A. R., Jones, J. L., Somogyi, Á., and Wysocki, V. H., Influence of Peptide Composition, Gas-Phase Basicity, and Chemical Modification on Fragmentation Efficiency: Evidence for the Mobile Proton Model. *Journal of the American Chemical Society* **1996**, *118* (35), 8365-8374.
57. Tsaprailis, G., Nair, H., Somogyi, Á., Wysocki, V. H., Zhong, W., Futrell, J. H., Summerfield, S. G., and Gaskell, S. J., Influence of Secondary Structure on the Fragmentation of Protonated Peptides. *Journal of the American Chemical Society* **1999**, *121* (22), 5142-5154.
58. Wysocki, V. H., Tsaprailis, G., Smith, L. L., and Breci, L. A., Mobile and localized protons: a framework for understanding peptide dissociation. *Journal of Mass Spectrometry* **2000**, *35* (12), 1399-1406.
59. Harrison, A. G., and Yalcin, T., Proton mobility in protonated amino acids and peptides. *International Journal of Mass Spectrometry and Ion Processes* **1997**, *165-166*, 339-347.
60. Somogyi, Á., Wysocki, V. H., and Mayer, I., The effect of protonation site on bond strengths in simple peptides: Application of ab initio and modified neglect of differential overlap bond orders and modified neglect of differential overlap energy partitioning. *Journal of the American Society for Mass Spectrometry* **1994**, *5* (8), 704-717.
61. Paizs, B., and Suhai, S., Towards understanding the tandem mass spectra of protonated oligopeptides. 1: mechanism of amide bond cleavage. *J Am Soc Mass Spectrom* **2004**, *15* (1), 103-113.

62. Roth, K. D., Huang, Z. H., Sadagopan, N., and Watson, J. T., Charge derivatization of peptides for analysis by mass spectrometry. *Mass Spectrom Rev* **1998**, *17* (4), 255-274.
63. [www.webbook.nist.gov](http://www.webbook.nist.gov).
64. Gu, C., Tsaprailis, G., Brecci, L., and Wysocki, V. H., Selective gas-phase cleavage at the peptide bond C-terminal to aspartic acid in fixed-charge derivatives of Asp-containing peptides. *Anal Chem* **2000**, *72* (23), 5804-5813.
65. Vaisar, T., and Urban, J., Probing the proline effect in CID of protonated peptides. *J Mass Spectrom* **1996**, *31* (10), 1185-1187.
66. Paizs, B., and Suhai, S., Fragmentation pathways of protonated peptides. *Mass Spectrometry Reviews* **2005**, *24* (4), 508-548.
67. Barton, S. J., and Whittaker, J. C., Review of factors that influence the abundance of ions produced in a tandem mass spectrometer and statistical methods for discovering these factors. *Mass Spectrometry Reviews* **2009**, *28* (1), 177-187.
68. Syka, J. E., Coon, J. J., Schroeder, M. J., Shabanowitz, J., and Hunt, D. F., Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci U S A* **2004**, *101* (26), 9528-9533.
69. Taouatas, N., Drugan, M. M., Heck, A. J. R., and Mohammed, S., Straightforward ladder sequencing of peptides using a Lys-N metalloendopeptidase. *Nat Meth* **2008**, *5* (5), 405-407.
70. van Breukelen, B., Georgiou, A., Drugan, M. M., Taouatas, N., Mohammed, S., and Heck, A. J. R., LysNDeNovo: An algorithm enabling de novo sequencing of Lys-N generated peptides fragmented by electron transfer dissociation. *Proteomics* **2010**, *10* (6), 1196-1201.
71. Laemmli, U. K., Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **1970**, *227* (5259), 680-685.
72. Rabilloud, T., Vaezzadeh, A. R., Potier, N., Lelong, C., Leize-Wagner, E., and Chevallet, M., Power and limitations of electrophoretic separations in proteomics strategies. *Mass Spectrom Rev* **2009**, *28* (5), 816-843.
73. O'Farrell, P. H., High resolution two-dimensional electrophoresis of proteins. *Journal of Biological Chemistry* **1975**, *250* (10), 4007-4021.
74. Righetti, P. G., and Gianazza, E., Isoelectric focusing in immobilized pH gradients: theory and newer methodology. *Methods Biochem Anal* **1987**, *32*, 215-278.
75. Gorg, A., Weiss, W., and Dunn, M. J., Current two-dimensional electrophoresis technology for proteomics. *Proteomics* **2004**, *4* (12), 3665-3685.
76. Santoni, V., Molloy, M., and Rabilloud, T., Membrane proteins and proteomics: un amour impossible? *Electrophoresis* **2000**, *21* (6), 1054-1070.
77. Braun, R. J., Kinkl, N., Beer, M., and Ueffing, M., Two-dimensional electrophoresis of membrane proteins. *Anal Bioanal Chem* **2007**, *389* (4), 1033-1045.
78. Neuhoff, V., Arold, N., Taube, D., and Ehrhardt, W., Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250. *Electrophoresis* **1988**, *9* (6), 255-262.
79. Shevchenko, A., Wilm, M., Vorm, O., and Mann, M., Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal Chem* **1996**, *68* (5), 850-858.
80. Gevaert, K., and Vandekerckhove, J., Protein identification methods in proteomics. *Electrophoresis* **2000**, *21* (6), 1145-1154.
81. Richert, S., Luche, S., Chevallet, M., Van Dorsselaer, A., Leize-Wagner, E., and Rabilloud, T., About the mechanism of interference of silver staining with peptide mass spectrometry. *Proteomics* **2004**, *4* (4), 909-916.
82. Rabilloud, T., Strub, J. M., Luche, S., van Dorsselaer, A., and Lunardi, J., A comparison between Sypro Ruby and ruthenium II tris (bathophenanthroline disulfonate) as fluorescent stains for protein detection in gels. *Proteomics* **2001**, *1* (5), 699-704.
83. Wittig, I., Braun, H.-P., and Schagger, H., Blue native PAGE. *Nat. Protocols* **2006**, *1* (1), 418-428.

84. Fiala, G. J., Schamel, W. W. A., and Blumenthal, B., Blue Native Polyacrylamide Gel Electrophoresis (BN-PAGE) for Analysis of Multiprotein Complexes from Cellular Lysates. *J Vis Exp* **2011**, (48), e2164.
85. Ros, A., Faupel, M., Mees, H., Oostrum, J., Ferrigno, R., Reymond, F., Michel, P., Rossier, J. S., and Girault, H. H., Protein purification by Off-Gel electrophoresis. *Proteomics* **2002**, 2 (2), 151-156.
86. Arnaud, I. L., Josserand, J., Rossier, J. S., and Girault, H. H., Finite element simulation of Off-Gel trade mark buffering. *Electrophoresis* **2002**, 23 (19), 3253-3261.
87. Michel, P. E., Reymond, F., Arnaud, I. L., Josserand, J., Girault, H. H., and Rossier, J. S., Protein fractionation in a multicompartiment device using Off-Gel isoelectric focusing. *Electrophoresis* **2003**, 24 (1-2), 3-11.
88. Lam, H. T., Josserand, J., Lion, N., and Girault, H. H., Modeling the isoelectric focusing of peptides in an OFFGEL multicompartiment cell. *J Proteome Res* **2007**, 6 (5), 1666-1676.
89. le Coutre, J., Whitelegge, J. P., Gross, A., Turk, E., Wright, E. M., Kaback, H. R., and Faull, K. F., Proteomics on Full-Length Membrane Proteins Using Mass Spectrometry†. *Biochemistry* **2000**, 39 (15), 4237-4242.
90. Zhang, Z., Smith, D. L., and Smith, J. B., Multiple separations facilitate identification of protein variants by mass spectrometry. *Proteomics* **2001**, 1 (8), 1001-1009.
91. Schluesener, D., Fischer, F., Kruip, J., Rogner, M., and Poetsch, A., Mapping the membrane proteome of *Corynebacterium glutamicum*. *Proteomics* **2005**, 5 (5), 1317-1330.
92. McDonald, T., Sheng, S., Stanley, B., Chen, D., Ko, Y., Cole, R. N., Pedersen, P., and Van Eyk, J. E., Expanding the subproteome of the inner mitochondria using protein separation technologies: one- and two-dimensional liquid chromatography and two-dimensional gel electrophoresis. *Mol Cell Proteomics* **2006**, 5 (12), 2392-2411.
93. Sandra, K., Moshir, M., D'Hondt, F., Verleysen, K., Kas, K., and Sandra, P., Highly efficient peptide separations in proteomics Part 1. Unidimensional high performance liquid chromatography. *J Chromatogr B Analyt Technol Biomed Life Sci* **2008**, 866 (1-2), 48-63.
94. Neverova, I., and Van Eyk, J. E., Role of chromatographic techniques in proteomic analysis. *J Chromatogr B Analyt Technol Biomed Life Sci* **2005**, 815 (1-2), 51-63.
95. Shen, Y., Smith, R. D., Unger, K. K., Kumar, D., and Lubda, D., Ultrahigh-throughput proteomics using fast RPLC separations with ESI-MS/MS. *Anal Chem* **2005**, 77 (20), 6692-6701.
96. Shen, Y., Zhang, R., Moore, R. J., Kim, J., Metz, T. O., Hixson, K. K., Zhao, R., Livesay, E. A., Udseth, H. R., and Smith, R. D., Automated 20 kpsi RPLC-MS and MS/MS with chromatographic peak capacities of 1000-1500 and capabilities in proteomics and metabolomics. *Anal Chem* **2005**, 77 (10), 3090-3100.
97. Kay, R. G., Gregory, B., Grace, P. B., and Pleasance, S., The application of ultra-performance liquid chromatography/tandem mass spectrometry to the detection and quantitation of apolipoproteins in human serum. *Rapid Commun Mass Spectrom* **2007**, 21 (16), 2585-2593.
98. Premstaller, A., Oberacher, H., Walcher, W., Timperio, A. M., Zolla, L., Chervet, J. P., Cavusoglu, N., van Dorsselaer, A., and Huber, C. G., High-performance liquid chromatography-electrospray ionization mass spectrometry using monolithic capillary columns for proteomic studies. *Anal Chem* **2001**, 73 (11), 2390-2396.
99. Rieux, L., Niederlander, H., Verpoorte, E., and Bischoff, R., Silica monolithic columns: synthesis, characterisation and applications to the analysis of biological molecules. *J Sep Sci* **2005**, 28 (14), 1628-1641.
100. Fortier, M. H., Bonneil, E., Goodley, P., and Thibault, P., Integrated microfluidic device for mass spectrometry-based proteomics and its application to biomarker discovery programs. *Anal Chem* **2005**, 77 (6), 1631-1640.
101. Yin, H., and Killeen, K., The fundamental aspects and applications of Agilent HPLC-Chip. *J Sep Sci* **2007**, 30 (10), 1427-1434.

102. Yin, H., Killeen, K., Brennen, R., Sobek, D., Werlich, M., and van de Goor, T., Microfluidic chip for peptide analysis with an integrated HPLC column, sample enrichment column, and nanoelectrospray tip. *Anal Chem* **2005**, *77* (2), 527-533.
103. Link, A. J., Eng, J., Schieltz, D. M., Carmack, E., Mize, G. J., Morris, D. R., Garvik, B. M., and Yates, J. R., 3rd, Direct analysis of protein complexes using mass spectrometry. *Nat Biotechnol* **1999**, *17* (7), 676-682.
104. Giddings, J. C., Two-dimensional separations: concept and promise. *Anal Chem* **1984**, *56* (12), 1258A-1260A, 1262A, 1264A passim.
105. Washburn, M. P., Wolters, D., and Yates, J. R., 3rd, Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* **2001**, *19* (3), 242-247.
106. Wolters, D. A., Washburn, M. P., and Yates, J. R., 3rd, An automated multidimensional protein identification technology for shotgun proteomics. *Anal Chem* **2001**, *73* (23), 5683-5690.
107. Wagner, Y., Sickmann, A., Meyer, H. E., and Daum, G., Multidimensional nano-HPLC for analysis of protein complexes. *J Am Soc Mass Spectrom* **2003**, *14* (9), 1003-1011.
108. Motoyama, A., Xu, T., Ruse, C. I., Wohlschlegel, J. A., and Yates, J. R., 3rd, Anion and cation mixed-bed ion exchange for enhanced multidimensional separations of peptides and phosphopeptides. *Anal Chem* **2007**, *79* (10), 3623-3634.
109. Opiteck, G. J., and Jorgenson, J. W., Two-dimensional SEC/RPLC coupled to mass spectrometry for the analysis of peptides. *Anal Chem* **1997**, *69* (13), 2283-2291.
110. Opiteck, G. J., Lewis, K. C., Jorgenson, J. W., and Anderegg, R. J., Comprehensive on-line LC/LC/MS of proteins. *Anal Chem* **1997**, *69* (8), 1518-1524.
111. Boersema, P. J., Divecha, N., Heck, A. J., and Mohammed, S., Evaluation and optimization of ZIC-HILIC-RP as an alternative MudPIT strategy. *J Proteome Res* **2007**, *6* (3), 937-946.
112. Delmotte, N., Lasaosa, M., Tholey, A., Heinzle, E., and 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. *J Proteome Res* **2007**, *6* (11), 4363-4373.
113. Nagele, E., Vollmer, M., and Horth, P., Improved 2D nano-LC/MS for proteomics applications: a comparative analysis using yeast proteome. *J Biomol Tech* **2004**, *15* (2), 134-143.
114. McLafferty, F. W., Breuker, K., Jin, M., Han, X., Infusini, G., Jiang, H., Kong, X., and Begley, T. P., Top-down MS, a powerful complement to the high capabilities of proteolysis proteomics. *FEBS J* **2007**, *274* (24), 6256-6268.
115. Han, X., Aslanian, A., and Yates, J. R., 3rd, Mass spectrometry for proteomics. *Curr Opin Chem Biol* **2008**, *12* (5), 483-490.
116. Henzel, W. J., Billeci, T. M., Stults, J. T., Wong, S. C., Grimley, C., and Watanabe, C., Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. *Proc Natl Acad Sci U S A* **1993**, *90* (11), 5011-5015.
117. Pappin, D. J., Hojrup, P., and Bleasby, A. J., Rapid identification of proteins by peptide-mass fingerprinting. *Curr Biol* **1993**, *3* (6), 327-332.
118. Yates, J. R., 3rd, Speicher, S., Griffin, P. R., and Hunkapiller, T., Peptide mass maps: a highly informative approach to protein identification. *Anal Biochem* **1993**, *214* (2), 397-408.
119. Mann, M., Hojrup, P., and Roepstorff, P., Use of mass spectrometric molecular weight information to identify proteins in sequence databases. *Biol Mass Spectrom* **1993**, *22* (6), 338-345.
120. James, P., Quadroni, M., Carafoli, E., and Gonnet, G., Protein identification by mass profile fingerprinting. *Biochem Biophys Res Commun* **1993**, *195* (1), 58-64.
121. Eng, J. K., McCormack, A. L., and Yates Iii, J. R., 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.
122. Mann, M., and Wilm, M., Error-tolerant identification of peptides in sequence databases by peptide sequence tags. *Anal Chem* **1994**, *66* (24), 4390-4399.

123. Perkins, D. N., Pappin, D. J., Creasy, D. M., and Cottrell, J. S., Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **1999**, *20* (18), 3551-3567.
124. Clauser, K. R., Baker, P., and Burlingame, A. L., Role of accurate mass measurement (+/- 10 ppm) in protein identification strategies employing MS or MS/MS and database searching. *Anal Chem* **1999**, *71* (14), 2871-2882.
125. Zhang, W., and Chait, B. T., ProFound: an expert system for protein identification using mass spectrometric peptide mapping information. *Anal Chem* **2000**, *72* (11), 2482-2489.
126. Carapito, C. (2006) Vers une meilleure utilisation des données de spectrométrie de masse en analyse protéomique. in *Thèse de l'Institut Pluridisciplinaire Hubert Curien - Département des Sciences Analytiques*, Université Louis Pasteur, Strasbourg
127. Horn, D. M., Peters, E. C., Klock, H., Meyers, A., and 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.
128. Carr, S., Aebersold, R., Baldwin, M., Burlingame, A., Clauser, K., and Nesvizhskii, A., The need for guidelines in publication of peptide and protein identification data: Working Group on Publication Guidelines for Peptide and Protein Identification Data. *Mol Cell Proteomics* **2004**, *3* (6), 531-533.
129. Bradshaw, R. A., Revised draft guidelines for proteomic data publication. *Mol Cell Proteomics* **2005**, *4* (9), 1223-1225.
130. Wilkins, M. R., Appel, R. D., Van Eyk, J. E., Chung, M. C., Gorg, A., Hecker, M., Huber, L. A., Langen, H., Link, A. J., Paik, Y. K., Patterson, S. D., Pennington, S. R., Rabilloud, T., Simpson, R. J., Weiss, W., and Dunn, M. J., Guidelines for the next 10 years of proteomics. *Proteomics* **2006**, *6* (1), 4-8.
131. Gattiker, A., Bienvenut, W. V., Bairoch, A., and Gasteiger, E., FindPept, a tool to identify unmatched masses in peptide mass fingerprinting protein identification. *Proteomics* **2002**, *2* (10), 1435-1444.
132. Blueggel, M., Chamrad, D., and Meyer, H. E., Bioinformatics in proteomics. *Curr Pharm Biotechnol* **2004**, *5* (1), 79-88.
133. Craig, R., and Beavis, R. C., TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* **2004**, *20* (9), 1466-1467.
134. Colinge, J., Masselot, A., Giron, M., Dessingy, T., and Magnin, J., OLAV: towards high-throughput tandem mass spectrometry data identification. *Proteomics* **2003**, *3* (8), 1454-1463.
135. Geer, L. Y., Markey, S. P., Kowalak, J. A., Wagner, L., Xu, M., Maynard, D. M., Yang, X., Shi, W., and Bryant, S. H., Open Mass Spectrometry Search Algorithm. *Journal of Proteome Research* **2004**, *3* (5), 958-964.
136. Searle, B. C., Turner, M., and Nesvizhskii, A. I., Improving sensitivity by probabilistically combining results from multiple MS/MS search methodologies. *J Proteome Res* **2008**, *7* (1), 245-253.
137. Savitski, M. M., Mathieson, T., Becher, I., and Bantscheff, M., H-score, a mass accuracy driven rescoring approach for improved peptide identification in modification rich samples. *J Proteome Res* **2010**, *9* (11), 5511-5516.
138. Savitski, M. M., Nielsen, M. L., Kjeldsen, F., and Zubarev, R. A., Proteomics-grade de novo sequencing approach. *J Proteome Res* **2005**, *4* (6), 2348-2354.
139. Shechter, D., Dormann, H. L., Allis, C. D., and Hake, S. B., Extraction, purification and analysis of histones. *Nat Protoc* **2007**, *2* (6), 1445-1457.
140. Bakalarski, C. E., Haas, W., Dephoure, N. E., and Gygi, S. P., The effects of mass accuracy, data acquisition speed, and search algorithm choice on peptide identification rates in phosphoproteomics. *Anal Bioanal Chem* **2007**, *389* (5), 1409-1419.

141. Balgley, B. M., Laudeman, T., Yang, L., Song, T., and Lee, C. S., Comparative evaluation of tandem MS search algorithms using a target-decoy search strategy. *Mol Cell Proteomics* **2007**, *6* (9), 1599-1608.
142. Resing, K. A., Meyer-Arendt, K., Mendoza, A. M., Aveline-Wolf, L. D., Jonscher, K. R., Pierce, K. G., Old, W. M., Cheung, H. T., Russell, S., Wattawa, J. L., Goehle, G. R., Knight, R. D., and Ahn, N. G., Improving reproducibility and sensitivity in identifying human proteins by shotgun proteomics. *Anal Chem* **2004**, *76* (13), 3556-3568.
143. Elias, J. E., Haas, W., Faherty, B. K., and Gygi, S. P., Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. *Nat Methods* **2005**, *2* (9), 667-675.
144. Kapp, E. A., Schutz, F., Connolly, L. M., Chakel, J. A., Meza, J. E., Miller, C. A., Fenyo, D., Eng, J. K., Adkins, J. N., Omenn, G. S., and Simpson, R. J., An evaluation, comparison, and accurate benchmarking of several publicly available MS/MS search algorithms: sensitivity and specificity analysis. *Proteomics* **2005**, *5* (13), 3475-3490.
145. Price, T. S., Lucitt, M. B., Wu, W., Austin, D. J., Pizarro, A., Yocum, A. K., Blair, I. A., FitzGerald, G. A., and Grosser, T., EBP, a program for protein identification using multiple tandem mass spectrometry datasets. *Mol Cell Proteomics* **2007**, *6* (3), 527-536.
146. Alves, G., Wu, W. W., Wang, G., Shen, R. F., and Yu, Y. K., Enhancing peptide identification confidence by combining search methods. *J Proteome Res* **2008**, *7* (8), 3102-3113.
147. Jones, A. R., Siepen, J. A., Hubbard, S. J., and Paton, N. W., Improving sensitivity in proteome studies by analysis of false discovery rates for multiple search engines. *Proteomics* **2009**, *9* (5), 1220-1229.
148. Nesvizhskii, A. I., and Aebersold, R., Analysis, statistical validation and dissemination of large-scale proteomics datasets generated by tandem MS. *Drug Discovery Today* **2004**, *9* (4), 173-181.
149. Russell, S. A., Old, W., Resing, K. A., and Hunter, L., Proteomic informatics. *Int Rev Neurobiol* **2004**, *61*, 127-157.
150. Baldwin, M. A., Protein identification by mass spectrometry: issues to be considered. *Mol Cell Proteomics* **2004**, *3* (1), 1-9.
151. Xie, H., and Griffin, T. J., Trade-off between high sensitivity and increased potential for false positive peptide sequence matches using a two-dimensional linear ion trap for tandem mass spectrometry-based proteomics. *J Proteome Res* **2006**, *5* (4), 1003-1009.
152. States, D. J., Omenn, G. S., Blackwell, T. W., Fermin, D., Eng, J., Speicher, D. W., and Hanash, S. M., Challenges in deriving high-confidence protein identifications from data gathered by a HUPO plasma proteome collaborative study. *Nat Biotechnol* **2006**, *24* (3), 333-338.
153. Boguski, M. S., and McIntosh, M. W., Biomedical informatics for proteomics. *Nature* **2003**, *422* (6928), 233-237.
154. Cargile, B. J., Bundy, J. L., and Stephenson, J. L., Jr., Potential for false positive identifications from large databases through tandem mass spectrometry. *J Proteome Res* **2004**, *3* (5), 1082-1085.
155. Chen, Y., Zhang, J., Xing, G., and Zhao, Y., Mascot-Derived False Positive Peptide Identifications Revealed by Manual Analysis of Tandem Mass Spectra. *Journal of Proteome Research* **2009**, *8* (6), 3141-3147.
156. Nesvizhskii, A. I., Keller, A., Kolker, E., and Aebersold, R., A statistical model for identifying proteins by tandem mass spectrometry. *Anal Chem* **2003**, *75* (17), 4646-4658.
157. Patterson, S. D., Data analysis--the Achilles heel of proteomics. *Nat Biotechnol* **2003**, *21* (3), 221-222.
158. Nesvizhskii, A. I., A survey of computational methods and error rate estimation procedures for peptide and protein identification in shotgun proteomics. *Journal of Proteomics* **2010**, *73* (11), 2092-2123.
159. Moore, R. E., Young, M. K., and Lee, T. D., Qscore: an algorithm for evaluating SEQUEST database search results. *J Am Soc Mass Spectrom* **2002**, *13* (4), 378-386.

160. Peng, J., Elias, J. E., Thoreen, C. C., Licklider, L. J., and Gygi, S. P., Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. *J Proteome Res* **2003**, *2* (1), 43-50.
161. Elias, J. E., and Gygi, S. P., Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods* **2007**, *4* (3), 207-214.
162. Bianco, L., Mead, J. A., and Bessant, C., Comparison of Novel Decoy Database Designs for Optimizing Protein Identification Searches Using ABRF sPRG2006 Standard MS/MS Data Sets. *Journal of Proteome Research* **2009**, *8* (4), 1782-1791.
163. Wang, G., Wu, W. W., Zhang, Z., Masilamani, S., and Shen, R.-F., Decoy Methods for Assessing False Positives and False Discovery Rates in Shotgun Proteomics. *Analytical Chemistry* **2008**, *81* (1), 146-159.
164. Choi, H., and Nesvizhskii, A. I., False Discovery Rates and Related Statistical Concepts in Mass Spectrometry-Based Proteomics. *Journal of Proteome Research* **2007**, *7* (1), 47-50.
165. Fitzgibbon, M., Li, Q., and McIntosh, M., Modes of Inference for Evaluating the Confidence of Peptide Identifications. *Journal of Proteome Research* **2007**, *7* (1), 35-39.
166. Käll, L., Storey, J. D., MacCoss, M. J., and Noble, W. S., Posterior Error Probabilities and False Discovery Rates: Two Sides of the Same Coin. *Journal of Proteome Research* **2007**, *7* (1), 40-44.
167. Käll, L., Storey, J. D., MacCoss, M. J., and Noble, W. S., Assigning Significance to Peptides Identified by Tandem Mass Spectrometry Using Decoy Databases. *Journal of Proteome Research* **2007**, *7* (1), 29-34.
168. Navarro, P., and Vázquez, J. s., A Refined Method To Calculate False Discovery Rates for Peptide Identification Using Decoy Databases. *Journal of Proteome Research* **2009**, *8* (4), 1792-1796.
169. Bayes, M., and Price, M., An Essay towards Solving a Problem in the Doctrine of Chances. By the Late Rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a Letter to John Canton, A. M. F. R. S. *Philosophical Transactions* **1763**, *53*, 370-418.
170. Keller, A., Nesvizhskii, A. I., Kolker, E., and Aebersold, R., Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal Chem* **2002**, *74* (20), 5383-5392.
171. Dempster, A., Laird, N., and Rdin, D., Maximum Likelihood from Incomplete Data via the EM Algorithm. *JOURNAL OF THE ROYAL STATISTICAL SOCIETY, SERIES B* **1977**, *39* (1), 1-38.
172. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J., Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **1997**, *25* (17), 3389-3402.
173. Shevchenko, A., Sunyaev, S., Loboda, A., Bork, P., Ens, W., and Standing, K. G., Charting the proteomes of organisms with unsequenced genomes by MALDI-quadrupole time-of-flight mass spectrometry and BLAST homology searching. *Anal Chem* **2001**, *73* (9), 1917-1926.
174. Shevchenko, A., Sunyaev, S., Liska, A., and Bork, P., Nanoelectrospray tandem mass spectrometry and sequence similarity searching for identification of proteins from organisms with unknown genomes. *Methods Mol Biol* **2003**, *211*, 221-234.
175. Habermann, B., Oegema, J., Sunyaev, S., and Shevchenko, A., The power and the limitations of cross-species protein identification by mass spectrometry-driven sequence similarity searches. *Mol Cell Proteomics* **2004**, *3* (3), 238-249.
176. Martens, L., Hermjakob, H., Jones, P., Adamski, M., Taylor, C., States, D., Gevaert, K., Vandekerckhove, J., and Apweiler, R., PRIDE: the proteomics identifications database. *Proteomics* **2005**, *5* (13), 3537-3545.
177. Desiere, F., Deutsch, E. W., Nesvizhskii, A. I., Mallick, P., King, N. L., Eng, J. K., Aderem, A., Boyle, R., Brunner, E., Donohoe, S., Fausto, N., Hafen, E., Hood, L., Katze, M. G., Kennedy, K. A., Kregenow, F., Lee, H., Lin, B., Martin, D., Ranish, J. A., Rawlings, D. J., Samelson, L. E., Shiio, Y., Watts, J. D., Wollscheid, B., Wright, M. E., Yan, W., Yang, L., Yi, E. C., Zhang, H., and

- Aebersold, R., Integration with the human genome of peptide sequences obtained by high-throughput mass spectrometry. *Genome Biol* **2005**, *6* (1), R9.
178. Craig, R., Cortens, J. P., and Beavis, R. C., The use of proteotypic peptide libraries for protein identification. *Rapid Commun Mass Spectrom* **2005**, *19* (13), 1844-1850.
179. Lam, H., Deutsch, E. W., Eddes, J. S., Eng, J. K., King, N., Stein, S. E., and Aebersold, R., Development and validation of a spectral library searching method for peptide identification from MS/MS. *Proteomics* **2007**, *7* (5), 655-667.
180. Frewen, B. E., Merrihew, G. E., Wu, C. C., Noble, W. S., and MacCoss, M. J., Analysis of peptide MS/MS spectra from large-scale proteomics experiments using spectrum libraries. *Anal Chem* **2006**, *78* (16), 5678-5684.
181. Dasari, S., Chambers, M. C., Slebos, R. J., Zimmerman, L. J., Ham, A. J., and Tabb, D. L., TagRecon: high-throughput mutation identification through sequence tagging. *J Proteome Res* **2010**, *9* (4), 1716-1726.
182. Shen, Y., Tolic, N., Hixson, K. K., Purvine, S. O., Anderson, G. A., and Smith, R. D., De novo sequencing of unique sequence tags for discovery of post-translational modifications of proteins. *Anal Chem* **2008**, *80* (20), 7742-7754.
183. Shen, Y., Tolic, N., Hixson, K. K., Purvine, S. O., Pasa-Tolic, L., Qian, W. J., Adkins, J. N., Moore, R. J., and Smith, R. D., Proteome-wide identification of proteins and their modifications with decreased ambiguities and improved false discovery rates using unique sequence tags. *Anal Chem* **2008**, *80* (6), 1871-1882.
184. Li, G. Z., Vissers, J. P., Silva, J. C., Golick, D., Gorenstein, M. V., and Geromanos, S. J., Database searching and accounting of multiplexed precursor and product ion spectra from the data independent analysis of simple and complex peptide mixtures. *Proteomics* **2009**, *9* (6), 1696-1719.
185. Geromanos, S. J., Vissers, J. P., Silva, J. C., Dorschel, C. A., Li, G. Z., Gorenstein, M. V., Bateman, R. H., and Langridge, J. I., The detection, correlation, and comparison of peptide precursor and product ions from data independent LC-MS with data dependant LC-MS/MS. *Proteomics* **2009**, *9* (6), 1683-1695.
186. Zhu, W., Smith, J. W., and Huang, C. M., Mass spectrometry-based label-free quantitative proteomics. *J Biomed Biotechnol* **2010**, *2010*, 840518.
187. Issaq, H., and Veenstra, T., Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE): advances and perspectives. *Biotechniques* **2008**, *44* (5), 697-698, 700.
188. Bantscheff, M., Schirle, M., Sweetman, G., Rick, J., and Kuster, B., Quantitative mass spectrometry in proteomics: a critical review. *Analytical and Bioanalytical Chemistry* **2007**, *389* (4), 1017-1031.
189. Domon, B., and Aebersold, R., Options and considerations when selecting a quantitative proteomics strategy. *Nat Biotechnol* **2010**, *28* (7), 710-721.
190. Rabilloud, T. (2003) Electrophorèse bidimensionnelle et analyse protéomique. in *Analyse protéomique: nouvelles approches d'études des interactions protéine-protéine*,
191. Unlu, M., Morgan, M. E., and Minden, J. S., Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis* **1997**, *18* (11), 2071-2077.
192. Alban, A., David, S. O., Bjorkesten, L., Andersson, C., Sloge, E., Lewis, S., and Currie, I., A novel experimental design for comparative two-dimensional gel analysis: two-dimensional difference gel electrophoresis incorporating a pooled internal standard. *Proteomics* **2003**, *3* (1), 36-44.
193. BiogenOuest, P. P., <http://www.proteome.univ-rennes1.fr/>. Université de Rennes I.
194. Van den Bergh, G., Clerens, S., Cnops, L., Vandesande, F., and Arckens, L., Fluorescent two-dimensional difference gel electrophoresis and mass spectrometry identify age-related protein expression differences for the primary visual cortex of kitten and adult cat. *J Neurochem* **2003**, *85* (1), 193-205.

195. Marouga, R., David, S., and Hawkins, E., The development of the DIGE system: 2D fluorescence difference gel analysis technology. *Analytical and Bioanalytical Chemistry* **2005**, 382 (3), 669-678.
196. Bondarenko, P. V., Chelius, D., and Shaler, T. A., Identification and relative quantitation of protein mixtures by enzymatic digestion followed by capillary reversed-phase liquid chromatography-tandem mass spectrometry. *Anal Chem* **2002**, 74 (18), 4741-4749.
197. Chelius, D., and Bondarenko, P. V., Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry. *J Proteome Res* **2002**, 1 (4), 317-323.
198. Wang, W., Zhou, H., Lin, H., Roy, S., Shaler, T. A., Hill, L. R., Norton, S., Kumar, P., Anderle, M., and Becker, C. H., Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards. *Anal Chem* **2003**, 75 (18), 4818-4826.
199. Wiener, M. C., Sachs, J. R., Deyanova, E. G., and Yates, N. A., Differential mass spectrometry: a label-free LC-MS method for finding significant differences in complex peptide and protein mixtures. *Anal Chem* **2004**, 76 (20), 6085-6096.
200. Wang, G., Wu, W. W., Zeng, W., Chou, C. L., and Shen, R. F., Label-free protein quantification using LC-coupled ion trap or FT mass spectrometry: Reproducibility, linearity, and application with complex proteomes. *J Proteome Res* **2006**, 5 (5), 1214-1223.
201. Smith, R. D., Anderson, G. A., Lipton, M. S., Masselon, C., Pasa-Tolic, L., Udseth, H., Belov, M., Shen, Y., and Veenstra, T. D., High-performance separations and mass spectrometric methods for high-throughput proteomics using accurate mass tags. *Adv Protein Chem* **2003**, 65, 85-131.
202. Washburn, M. P., Ulaszek, R., Deciu, C., Schieltz, D. M., and Yates, J. R., 3rd, Analysis of quantitative proteomic data generated via multidimensional protein identification technology. *Anal Chem* **2002**, 74 (7), 1650-1657.
203. Liu, H., Sadygov, R. G., and Yates, J. R., 3rd, A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem* **2004**, 76 (14), 4193-4201.
204. Gilchrist, A., Au, C. E., Hiding, J., Bell, A. W., Fernandez-Rodriguez, J., Lesimple, S., Nagaya, H., Roy, L., Gosline, S. J., Hallett, M., Paiement, J., Kearney, R. E., Nilsson, T., and Bergeron, J. J., Quantitative proteomics analysis of the secretory pathway. *Cell* **2006**, 127 (6), 1265-1281.
205. Simpson, K. L., Whetton, A. D., and Dive, C., Quantitative mass spectrometry-based techniques for clinical use: biomarker identification and quantification. *J Chromatogr B Analyt Technol Biomed Life Sci* **2009**, 877 (13), 1240-1249.
206. Old, W. M., Meyer-Arendt, K., Aveline-Wolf, L., Pierce, K. G., Mendoza, A., Sevinsky, J. R., Resing, K. A., and Ahn, N. G., Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Mol Cell Proteomics* **2005**, 4 (10), 1487-1502.
207. Gao, J., Opitck, G. J., Friedrichs, M. S., Dongre, A. R., and Hefta, S. A., Changes in the protein expression of yeast as a function of carbon source. *J Proteome Res* **2003**, 2 (6), 643-649.
208. Colinge, J., Chiappe, D., Lagache, S., Moniatte, M., and Bougueleret, L., Differential proteomics via probabilistic peptide identification scores. *Anal Chem* **2005**, 77 (2), 596-606.
209. Mueller, L. N., Brusniak, M. Y., Mani, D. R., and Aebersold, R., An assessment of software solutions for the analysis of mass spectrometry based quantitative proteomics data. *J Proteome Res* **2008**, 7 (1), 51-61.
210. Rappsilber, J., Ryder, U., Lamond, A. I., and Mann, M., Large-scale proteomic analysis of the human spliceosome. *Genome Res* **2002**, 12 (8), 1231-1245.
211. Ishihama, Y., Oda, Y., Tabata, T., Sato, T., Nagasu, T., Rappsilber, J., and Mann, M., Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol Cell Proteomics* **2005**, 4 (9), 1265-1272.
212. Blondeau, F., Ritter, B., Allaire, P. D., Wasiak, S., Girard, M., Hussain, N. K., Angers, A., Legendre-Guillemain, V., Roy, L., Boismenu, D., Kearney, R. E., Bell, A. W., Bergeron, J. J., and McPherson, P. S., Tandem MS analysis of brain clathrin-coated vesicles reveals their critical involvement in synaptic vesicle recycling. *Proc Natl Acad Sci U S A* **2004**, 101 (11), 3833-3838.

213. Bergeron, J. J., and Hallett, M., Peptides you can count on. *Nat Biotechnol* **2007**, *25* (1), 61-62.
214. Bernay, B., Gaillard, M. C., Guryca, V., Emadali, A., Kuhn, L., Bertrand, A., Detraz, I., Carcenac, C., Savasta, M., Brouillet, E., Garin, J., and Elalouf, J. M., Discovering new bioactive neuropeptides in the striatum secretome using in vivo microdialysis and versatile proteomics. *Mol Cell Proteomics* **2009**, *8* (5), 946-958.
215. Neilson, K. A., Ali, N. A., Muralidharan, S., Mirzaei, M., Mariani, M., Assadourian, G., Lee, A., van Sluyter, S. C., and Haynes, P. A., Less label, more free: approaches in label-free quantitative mass spectrometry. *Proteomics* **2011**, *11* (4), 535-553.
216. Parker, S. J., Halligan, B. D., and Greene, A. S., Quantitative analysis of SILAC data sets using spectral counting. *Proteomics* **2010**, *10* (7), 1408-1415.
217. Lu, P., Vogel, C., Wang, R., Yao, X., and Marcotte, E. M., Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat Biotechnol* **2007**, *25* (1), 117-124.
218. Malmstrom, J., Beck, M., Schmidt, A., Lange, V., Deutsch, E. W., and Aebersold, R., Proteome-wide cellular protein concentrations of the human pathogen *Leptospira interrogans*. *Nature* **2009**, *460* (7256), 762-765.
219. Miguet, L., Bechade, G., Fornecker, L., Zink, E., Felden, C., Gervais, C., Herbrecht, R., Van Dorsselaer, A., Mauvieux, L., and Sanglier-Cianferani, S., Proteomic analysis of malignant B-cell derived microparticles reveals CD148 as a potentially useful antigenic biomarker for mantle cell lymphoma diagnosis. *J Proteome Res* **2009**, *8* (7), 3346-3354.
220. Park, J., Cha, D. H., Lee, S. J., Kim, Y. N., Kim, Y. H., and Kim, K. P., Discovery of the serum biomarker proteins in severe preeclampsia by proteomic analysis. *Exp Mol Med* **2011**, *43* (7), 427-435.
221. Xie, X., Feng, S., Vuong, H., Liu, Y., Goodison, S., and Lubman, D. M., A comparative phosphoproteomic analysis of a human tumor metastasis model using a label-free quantitative approach. *Electrophoresis* **2010**, *31* (11), 1842-1852.
222. Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., and Aebersold, R., Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* **1999**, *17* (10), 994-999.
223. Paša-Tolić, L., Jensen, P. K., Anderson, G. A., Lipton, M. S., Peden, K. K., Martinović, S., Tolić, N., Bruce, J. E., and Smith, R. D., High Throughput Proteome-Wide Precision Measurements of Protein Expression Using Mass Spectrometry. *Journal of the American Chemical Society* **1999**, *121* (34), 7949-7950.
224. Oda, Y., Huang, K., Cross, F. R., Cowburn, D., and Chait, B. T., Accurate quantitation of protein expression and site-specific phosphorylation. *Proc Natl Acad Sci U S A* **1999**, *96* (12), 6591-6596.
225. Kline, K. G., and Sussman, M. R., Protein Quantitation Using Isotope-Assisted Mass Spectrometry. *Annual Review of Biophysics* **2010**, *39* (1), 291-308.
226. Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., and Mann, M., Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* **2002**, *1* (5), 376-386.
227. Ong, S. E., Kratchmarova, I., and Mann, M., Properties of <sup>13</sup>C-substituted arginine in stable isotope labeling by amino acids in cell culture (SILAC). *J Proteome Res* **2003**, *2* (2), 173-181.
228. Amanchy, R., Kalume, D. E., Iwahori, A., Zhong, J., and Pandey, A., Phosphoproteome analysis of HeLa cells using stable isotope labeling with amino acids in cell culture (SILAC). *J Proteome Res* **2005**, *4* (5), 1661-1671.
229. Rikova, K., Guo, A., Zeng, Q., Possemato, A., Yu, J., Haack, H., Nardone, J., Lee, K., Reeves, C., Li, Y., Hu, Y., Tan, Z., Stokes, M., Sullivan, L., Mitchell, J., Wetzels, R., Macneill, J., Ren, J. M., Yuan, J., Bakalarski, C. E., Villen, J., Kornhauser, J. M., Smith, B., Li, D., Zhou, X., Gygi, S. P., Gu, T. L., Polakiewicz, R. D., Rush, J., and Comb, M. J., Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer. *Cell* **2007**, *131* (6), 1190-1203.

230. Olsen, J. V., Blagoev, B., Gnad, F., Macek, B., Kumar, C., Mortensen, P., and Mann, M., Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* **2006**, *127* (3), 635-648.
231. Zhang, X. X., Chan, C. S., Bao, H., Fang, Y., Foster, L. J., and Duong, F., Nanodiscs and SILAC-Based Mass Spectrometry to Identify a Membrane Protein Interactome. *Journal of Proteome Research* **2011**, *11* (2), 1454-1459.
232. Geiger, T., Wehner, A., Schaab, C., Cox, J., and Mann, M., Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins. *Molecular & Cellular Proteomics* **2012**.
233. Gouw, J. W., Tops, B. B., Mortensen, P., Heck, A. J., and Krijgsveld, J., Optimizing identification and quantitation of <sup>15</sup>N-labeled proteins in comparative proteomics. *Anal Chem* **2008**, *80* (20), 7796-7803.
234. Nelson, C. J., Huttlin, E. L., Hegeman, A. D., Harms, A. C., and Sussman, M. R., Implications of <sup>15</sup>N-metabolic labeling for automated peptide identification in *Arabidopsis thaliana*. *Proteomics* **2007**, *7* (8), 1279-1292.
235. Krijgsveld, J., Ketting, R. F., Mahmoudi, T., Johansen, J., Artal-Sanz, M., Verrijzer, C. P., Plasterk, R. H., and Heck, A. J., Metabolic labeling of *C. elegans* and *D. melanogaster* for quantitative proteomics. *Nat Biotechnol* **2003**, *21* (8), 927-931.
236. Wu, C. C., MacCoss, M. J., Howell, K. E., Matthews, D. E., and Yates, J. R., 3rd, Metabolic labeling of mammalian organisms with stable isotopes for quantitative proteomic analysis. *Anal Chem* **2004**, *76* (17), 4951-4959.
237. Zhang, Y., Webhofer, C., Reckow, S., Filiou, M. D., Maccarrone, G., and Turck, C. W., A MS data search method for improved <sup>15</sup>N-labeled protein identification. *Proteomics* **2009**, *9* (17), 4265-4270.
238. Gouw, J. W., Krijgsveld, J., and Heck, A. J., Quantitative proteomics by metabolic labeling of model organisms. *Mol Cell Proteomics* **2010**, *9* (1), 11-24.
239. Blagoev, B., Ong, S. E., Kratchmarova, I., and Mann, M., Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics. *Nat Biotechnol* **2004**, *22* (9), 1139-1145.
240. Park, K. S., Mohapatra, D. P., Misonou, H., and Trimmer, J. S., Graded regulation of the Kv2.1 potassium channel by variable phosphorylation. *Science* **2006**, *313* (5789), 976-979.
241. Blagoev, B., Kratchmarova, I., Ong, S. E., Nielsen, M., Foster, L. J., and Mann, M., A proteomics strategy to elucidate functional protein-protein interactions applied to EGF signaling. *Nat Biotechnol* **2003**, *21* (3), 315-318.
242. Ong, S. E., and Mann, M., A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). *Nat Protoc* **2006**, *1* (6), 2650-2660.
243. Van Hoof, D., Pinkse, M. W., Oostwaard, D. W., Mummery, C. L., Heck, A. J., and Krijgsveld, J., An experimental correction for arginine-to-proline conversion artifacts in SILAC-based quantitative proteomics. *Nat Methods* **2007**, *4* (9), 677-678.
244. Reynolds, K. J., Yao, X., and Fenselau, C., Proteolytic <sup>18</sup>O labeling for comparative proteomics: evaluation of endoprotease Glu-C as the catalytic agent. *J Proteome Res* **2002**, *1* (1), 27-33.
245. Rao, K. C., Carruth, R. T., and Miyagi, M., Proteolytic <sup>18</sup>O labeling by peptidyl-Lys metalloendopeptidase for comparative proteomics. *J Proteome Res* **2005**, *4* (2), 507-514.
246. Schnolzer, M., Jedrzejewski, P., and Lehmann, W. D., Protease-catalyzed incorporation of <sup>18</sup>O into peptide fragments and its application for protein sequencing by electrospray and matrix-assisted laser desorption/ionization mass spectrometry. *Electrophoresis* **1996**, *17* (5), 945-953.
247. Johnson, K. L., and Muddiman, D. C., A method for calculating <sup>16</sup>O/<sup>18</sup>O peptide ion ratios for the relative quantification of proteomes. *J Am Soc Mass Spectrom* **2004**, *15* (4), 437-445.

248. Ramos-Fernandez, A., Lopez-Ferrer, D., and Vazquez, J., Improved method for differential expression proteomics using trypsin-catalyzed 18O labeling with a correction for labeling efficiency. *Mol Cell Proteomics* **2007**, *6* (7), 1274-1286.
249. Ye, X., Luke, B., Andresson, T., and Blonder, J., 18O Stable Isotope Labeling in MS-based Proteomics. *Briefings in Functional Genomics & Proteomics* **2009**, *8* (2), 136-144.
250. Smolka, M. B., Zhou, H., Purkayastha, S., and Aebersold, R., Optimization of the isotope-coded affinity tag-labeling procedure for quantitative proteome analysis. *Anal Biochem* **2001**, *297* (1), 25-31.
251. Oda, Y., Owa, T., Sato, T., Boucher, B., Daniels, S., Yamanaka, H., Shinohara, Y., Yokoi, A., Kuromitsu, J., and Nagasu, T., Quantitative chemical proteomics for identifying candidate drug targets. *Anal Chem* **2003**, *75* (9), 2159-2165.
252. Hansen, K. C., Schmitt-Ulms, G., Chalkley, R. J., Hirsch, J., Baldwin, M. A., and Burlingame, A. L., Mass spectrometric analysis of protein mixtures at low levels using cleavable 13C-isotope-coded affinity tag and multidimensional chromatography. *Mol Cell Proteomics* **2003**, *2* (5), 299-314.
253. Yi, E. C., Li, X. J., Cooke, K., Lee, H., Raught, B., Page, A., Aneliunas, V., Hieter, P., Goodlett, D. R., and Aebersold, R., Increased quantitative proteome coverage with (13)C/(12)C-based, acid-cleavable isotope-coded affinity tag reagent and modified data acquisition scheme. *Proteomics* **2005**, *5* (2), 380-387.
254. Yi, E. C., Li, X.-j., Cooke, K., Lee, H., Raught, B., Page, A., Aneliunas, V., Hieter, P., Goodlett, D. R., and Aebersold, R., Increased quantitative proteome coverage with 13C/12C-based, acid-cleavable isotope-coded affinity tag reagent and modified data acquisition scheme. *Proteomics* **2005**, *5* (2), 380-387.
255. Schmidt, A., Kellermann, J., and Lottspeich, F., A novel strategy for quantitative proteomics using isotope-coded protein labels. *Proteomics* **2005**, *5* (1), 4-15.
256. Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A., and Pappin, D. J., Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics* **2004**, *3* (12), 1154-1169.
257. Choe, L., D'Ascenzo, M., Relkin, N. R., Pappin, D., Ross, P., Williamson, B., Guertin, S., Pribil, P., and Lee, K. H., 8-plex quantitation of changes in cerebrospinal fluid protein expression in subjects undergoing intravenous immunoglobulin treatment for Alzheimer's disease. *Proteomics* **2007**, *7* (20), 3651-3660.
258. Thompson, A., Schafer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., Neumann, T., Johnstone, R., Mohammed, A. K., and Hamon, C., Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem* **2003**, *75* (8), 1895-1904.
259. Desiderio, D. M., and Kai, M., Preparation of stable isotope-incorporated peptide internal standards for field desorption mass spectrometry quantification of peptides in biologic tissue. *Biomed Mass Spectrom* **1983**, *10* (8), 471-479.
260. Gerber, S. A., Rush, J., Stemman, O., Kirschner, M. W., and Gygi, S. P., Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc Natl Acad Sci U S A* **2003**, *100* (12), 6940-6945.
261. Pan, S., Zhang, H., Rush, J., Eng, J., Zhang, N., Patterson, D., Comb, M. J., and Aebersold, R., High throughput proteome screening for biomarker detection. *Mol Cell Proteomics* **2005**, *4* (2), 182-190.
262. Pratt, J. M., Simpson, D. M., Doherty, M. K., Rivers, J., Gaskell, S. J., and Beynon, R. J., Multiplexed absolute quantification for proteomics using concatenated signature peptides encoded by QconCAT genes. *Nat Protoc* **2006**, *1* (2), 1029-1043.
263. Rivers, J., Simpson, D. M., Robertson, D. H., Gaskell, S. J., and Beynon, R. J., Absolute multiplexed quantitative analysis of protein expression during muscle development using QconCAT. *Mol Cell Proteomics* **2007**, *6* (8), 1416-1427.

264. Brun, V., Dupuis, A., Adrait, A., Marcellin, M., Thomas, D., Court, M., Vandenesch, F., and Garin, J., Isotope-labeled protein standards: toward absolute quantitative proteomics. *Mol Cell Proteomics* **2007**, *6* (12), 2139-2149.
265. Gallien, S., Duriez, E., and Domon, B., Selected reaction monitoring applied to proteomics. *J Mass Spectrom* **2011**, *46* (3), 298-312.
266. Gstaiger, M., and Aebersold, R., Applying mass spectrometry-based proteomics to genetics, genomics and network biology. *Nat Rev Genet* **2009**, *10* (9), 617-627.
267. Deutsch, E. W., Lam, H., and Aebersold, R., PeptideAtlas: a resource for target selection for emerging targeted proteomics workflows. *EMBO Rep* **2008**, *9* (5), 429-434.
268. Jones, P., Cote, R. G., Cho, S. Y., Klie, S., Martens, L., Quinn, A. F., Thorneycroft, D., and Hermjakob, H., PRIDE: new developments and new datasets. *Nucleic Acids Res* **2008**, *36* (Database issue), D878-883.
269. Fusaro, V. A., Mani, D. R., Mesirov, J. P., and Carr, S. A., Prediction of high-responding peptides for targeted protein assays by mass spectrometry. *Nat Biotechnol* **2009**, *27* (2), 190-198.
270. Mallick, P., Schirle, M., Chen, S. S., Flory, M. R., Lee, H., Martin, D., Ranish, J., Raught, B., Schmitt, R., Werner, T., Kuster, B., and Aebersold, R., Computational prediction of proteotypic peptides for quantitative proteomics. *Nat Biotechnol* **2007**, *25* (1), 125-131.
271. Sanders, W. S., Bridges, S. M., McCarthy, F. M., Nanduri, B., and Burgess, S. C., Prediction of peptides observable by mass spectrometry applied at the experimental set level. *BMC Bioinformatics* **2007**, *8 Suppl 7*, S23.
272. Webb-Robertson, B. J., Cannon, W. R., Oehmen, C. S., Shah, A. R., Gurumoorthi, V., Lipton, M. S., and Waters, K. M., A support vector machine model for the prediction of proteotypic peptides for accurate mass and time proteomics. *Bioinformatics* **2008**, *24* (13), 1503-1509.
273. Picotti, P., Lam, H., Campbell, D., Deutsch, E. W., Mirzaei, H., Ranish, J., Domon, B., and Aebersold, R., A database of mass spectrometric assays for the yeast proteome. *Nat Methods* **2008**, *5* (11), 913-914.
274. Kiyonami, R., Schoen, A., Prakash, A., Peterman, S., Zabrouskov, V., Picotti, P., Aebersold, R., Huhmer, A., and Domon, B., Increased selectivity, analytical precision, and throughput in targeted proteomics. *Mol Cell Proteomics* **2011**, *10* (2), M110 002931.
275. van den Berg, B. H., and Tholey, A., Mass spectrometry-based proteomics strategies for protease cleavage site identification. *Proteomics* **2012**, *12* (4-5), 516-529.
276. Barrett, A. J., Bioinformatics of proteases in the MEROPS database. *Curr Opin Drug Discov Devel* **2004**, *7* (3), 334-341.
277. Quesada, V., Ordonez, G. R., Sanchez, L. M., Puente, X. S., and Lopez-Otin, C., The Degradome database: mammalian proteases and diseases of proteolysis. *Nucleic Acids Res* **2009**, *37* (Database issue), D239-243.
278. Rawlings, N. D., Barrett, A. J., and Bateman, A., MEROPS: the peptidase database. *Nucleic Acids Res* **2010**, *38* (Database issue), D227-233.
279. Frohlich, E., Proteases in cutaneous malignant melanoma: relevance as biomarker and therapeutic target. *Cell Mol Life Sci* **2010**, *67* (23), 3947-3960.
280. Kessenbrock, K., Plaks, V., and Werb, Z., Matrix metalloproteinases: regulators of the tumor microenvironment. *Cell* **2010**, *141* (1), 52-67.
281. Nisman, B., Biran, H., Heching, N., Barak, V., Ramu, N., Nemirovsky, I., and Peretz, T., Prognostic role of serum cytokeratin 19 fragments in advanced non-small-cell lung cancer: association of marker changes after two chemotherapy cycles with different measures of clinical response and survival. *Br J Cancer* **2008**, *98* (1), 77-79.
282. Lopez-Otin, C., and Hunter, T., The regulatory crosstalk between kinases and proteases in cancer. *Nat Rev Cancer* **2010**, *10* (4), 278-292.
283. Chalaris, A., Adam, N., Sina, C., Rosenstiel, P., Lehmann-Koch, J., Schirmacher, P., Hartmann, D., Cichy, J., Gavrilova, O., Schreiber, S., Jostock, T., Matthews, V., Hasler, R., Becker, C., Neurath, M. F., Reiss, K., Saftig, P., Scheller, J., and Rose-John, S., Critical role of the

- disintegrin metalloprotease ADAM17 for intestinal inflammation and regeneration in mice. *J Exp Med* **2010**, *207* (8), 1617-1624.
- 284.** Xia, W., Amyloid metabolism and secretases in Alzheimer's disease. *Curr Neurol Neurosci Rep* **2001**, *1* (5), 422-427.
- 285.** Grammas, P., Neurovascular dysfunction, inflammation and endothelial activation: implications for the pathogenesis of Alzheimer's disease. *J Neuroinflammation* **2011**, *8*, 26.
- 286.** Bird, P. I., Trapani, J. A., and Villadangos, J. A., Endolysosomal proteases and their inhibitors in immunity. *Nat Rev Immunol* **2009**, *9* (12), 871-882.
- 287.** Spielmann, T., and Gilberger, T. W., Protein export in malaria parasites: do multiple export motifs add up to multiple export pathways? *Trends Parasitol* **2010**, *26* (1), 6-10.
- 288.** Osawa, R., Williams, K. L., and Singh, N., The inflammasome regulatory pathway and infections: role in pathophysiology and clinical implications. *J Infect* **2011**, *62* (2), 119-129.
- 289.** Doucet, A., Butler, G. S., Rodriguez, D., Prudova, A., and Overall, C. M., Metadegradomics: toward in vivo quantitative degradomics of proteolytic post-translational modifications of the cancer proteome. *Mol Cell Proteomics* **2008**, *7* (10), 1925-1951.
- 290.** Lopez-Otin, C., and Overall, C. M., Protease degradomics: a new challenge for proteomics. *Nat Rev Mol Cell Biol* **2002**, *3* (7), 509-519.
- 291.** Timmer, J. C., Enoksson, M., Wildfang, E., Zhu, W., Igarashi, Y., Denault, J. B., Ma, Y., Dummitt, B., Chang, Y. H., Mast, A. E., Eroshkin, A., Smith, J. W., Tao, W. A., and Salvesen, G. S., Profiling constitutive proteolytic events in vivo. *Biochem J* **2007**, *407* (1), 41-48.
- 292.** Xu, G., Shin, S. B., and Jaffrey, S. R., Global profiling of protease cleavage sites by chemoselective labeling of protein N-termini. *Proc Natl Acad Sci U S A* **2009**, *106* (46), 19310-19315.
- 293.** Schilling, O., and Overall, C. M., Proteome-derived, database-searchable peptide libraries for identifying protease cleavage sites. *Nat Biotechnol* **2008**, *26* (6), 685-694.
- 294.** Paes Leme, A. F., Escalante, T., Pereira, J. G., Oliveira, A. K., Sanchez, E. F., Gutierrez, J. M., Serrano, S. M., and Fox, J. W., High resolution analysis of snake venom metalloproteinase (SVMP) peptide bond cleavage specificity using proteome based peptide libraries and mass spectrometry. *J Proteomics* **2011**, *74* (4), 401-410.
- 295.** Doucet, A., and Overall, C. M., Broad coverage identification of multiple proteolytic cleavage site sequences in complex high molecular weight proteins using quantitative proteomics as a complement to edman sequencing. *Mol Cell Proteomics* **2011**, *10* (5), M110 003533.
- 296.** Gevaert, K., Goethals, M., Martens, L., Van Damme, J., Staes, A., Thomas, G. R., and Vandekerckhove, J., Exploring proteomes and analyzing protein processing by mass spectrometric identification of sorted N-terminal peptides. *Nat Biotechnol* **2003**, *21* (5), 566-569.
- 297.** Van Damme, P., Martens, L., Van Damme, J., Hugelier, K., Staes, A., Vandekerckhove, J., and Gevaert, K., Caspase-specific and nonspecific in vivo protein processing during Fas-induced apoptosis. *Nat Methods* **2005**, *2* (10), 771-777.
- 298.** Staes, A., Van Damme, P., Helsens, K., Demol, H., Vandekerckhove, J., and Gevaert, K., Improved recovery of proteome-informative, protein N-terminal peptides by combined fractional diagonal chromatography (COFRADIC). *Proteomics* **2008**, *8* (7), 1362-1370.
- 299.** Arnesen, T., Van Damme, P., Polevoda, B., Helsens, K., Evjenth, R., Colaert, N., Varhaug, J. E., Vandekerckhove, J., Lillehaug, J. R., Sherman, F., and Gevaert, K., Proteomics analyses reveal the evolutionary conservation and divergence of N-terminal acetyltransferases from yeast and humans. *Proc Natl Acad Sci U S A* **2009**, *106* (20), 8157-8162.
- 300.** Impens, F., Van Damme, P., Demol, H., Van Damme, J., Vandekerckhove, J., and Gevaert, K., Mechanistic insight into taxol-induced cell death. *Oncogene* **2008**, *27* (33), 4580-4591.
- 301.** Lamkanfi, M., Kanneganti, T. D., Van Damme, P., Vanden Berghe, T., Vanoverberghe, I., Vandekerckhove, J., Vandenabeele, P., Gevaert, K., and Nunez, G., Targeted peptidecentric proteomics reveals caspase-7 as a substrate of the caspase-1 inflammasomes. *Mol Cell Proteomics* **2008**, *7* (12), 2350-2363.

302. Van Damme, P., Maurer-Stroh, S., Plasman, K., Van Durme, J., Colaert, N., Timmerman, E., De Bock, P. J., Goethals, M., Rousseau, F., Schymkowitz, J., Vandekerckhove, J., and Gevaert, K., Analysis of protein processing by N-terminal proteomics reveals novel species-specific substrate determinants of granzyme B orthologs. *Mol Cell Proteomics* **2009**, *8* (2), 258-272.
303. Van Damme, P., Staes, A., Bronsoms, S., Helsens, K., Colaert, N., Timmerman, E., Aviles, F. X., Vandekerckhove, J., and Gevaert, K., Complementary positional proteomics for screening substrates of endo- and exoproteases. *Nat Methods* **2010**, *7* (7), 512-515.
304. McDonald, L., Robertson, D. H., Hurst, J. L., and Beynon, R. J., Positional proteomics: selective recovery and analysis of N-terminal proteolytic peptides. *Nat Methods* **2005**, *2* (12), 955-957.
305. McDonald, L., and Beynon, R. J., Positional proteomics: preparation of amino-terminal peptides as a strategy for proteome simplification and characterization. *Nat Protoc* **2006**, *1* (4), 1790-1798.
306. Enoksson, M., Li, J., Ivancic, M. M., Timmer, J. C., Wildfang, E., Eroshkin, A., Salvesen, G. S., and Tao, W. A., Identification of proteolytic cleavage sites by quantitative proteomics. *J Proteome Res* **2007**, *6* (7), 2850-2858.
307. Doucet, A., and Overall, C. M., Protease proteomics: revealing protease in vivo functions using systems biology approaches. *Mol Aspects Med* **2008**, *29* (5), 339-358.
308. Prudova, A., auf dem Keller, U., Butler, G. S., and Overall, C. M., Multiplex N-terminome Analysis of MMP-2 and MMP-9 Substrate Degradomes by iTRAQ-TAILS Quantitative Proteomics. *Molecular & Cellular Proteomics* **2010**, *9* (5), 894-911.
309. Impens, F., Colaert, N., Helsens, K., Plasman, K., Van Damme, P., Vandekerckhove, J., and Gevaert, K., MS-driven protease substrate degradomics. *Proteomics* **2010**, *10* (6), 1284-1296.
310. Veenstra, T. D., Conrads, T. P., and Issaq, H. J., What to do with "one-hit wonders"? *Electrophoresis* **2004**, *25* (9), 1278-1279.
311. Olsen, J. V., Ong, S. E., and Mann, M., Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol Cell Proteomics* **2004**, *3* (6), 608-614.
312. Taylor, C. F., Binz, P. A., Aebersold, R., Affolter, M., Barkovich, R., Deutsch, E. W., Horn, D. M., Huhmer, A., Kussmann, M., Lilley, K., Macht, M., Mann, M., Muller, D., Neubert, T. A., Nickson, J., Patterson, S. D., Raso, R., Resing, K., Seymour, S. L., Tsugita, A., Xenarios, I., Zeng, R., and Julian, R. K., Jr., Guidelines for reporting the use of mass spectrometry in proteomics. *Nat Biotechnol* **2008**, *26* (8), 860-861.



## Partie II. Développements méthodologiques

---

### **Chapitre 1**

Mise au point de la dérivation N-terminale des protéines  
dans le gel : La stratégie « Ingel N-TOP »

### **Chapitre 2**

Mise au point de la stratégie qN-TOP : Dérivation avec du  
TMPP <sup>13</sup>C pour la quantification du protéome N-terminal

### **Chapitre 3**

Apport du marquage avec un nouveau TMPP plus  
hydrophobe pour enrichir d'avantage en peptide N-terminaux.



## Introduction

Dans le cadre du développement d'une méthode d'aide à l'annotation des génomes, et plus particulièrement des codons d'initiation, une stratégie protéogénomique basée sur la dérivation des amines N-terminales des protéines a été développée au laboratoire. Le réactif utilisé pour cette dérivation est le bromure de N-(Succinimidylloxycarbonylméthyl)tris(2,4,6-triméthoxyphényl) phosphine (TMPP-Ac-OSu). Cette méthode consiste à dériver spécifiquement les amines N-terminales des protéines et puis d'identifier les séquences des peptides N-terminaux en effectuant des recherches directement dans les banques génomiques. Ceci a permis d'annoter et de corriger les erreurs de prédiction des codons d'initiation de plus de 300 gènes de *Mycobacterium smegmatis* [1]. Dans ce travail de thèse, nous nous sommes particulièrement intéressés au clivage protéolytique. Nous avons donc élargie le champs d'action de cette technologie de marquage en l'appliquant à l'étude de divers processus protéolytique tel que le clivage des peptides de transit des protéines mitochondriales, le protéome N-terminal du parasite de la malaria et les processus de dégradation des protéines lors du jeûne prolongé chez le rat. Dans cette partie je décrirai trois améliorations apportées à cette technologie de marquage ; mais avant ceci je décrirai ici brièvement l'historique du réactif et de la stratégie N-TOP (N-Terminal Oriented Proteomics).

### 1. Le TMPP-Ac-OSu et son utilisation en analyse de protéines par spectrométrie de masse

L'utilisation des dérivés de phosphine pour l'analyse des peptides par spectrométrie de masse fut introduite par l'équipe de Watson en 1991. Ils ont décrit une amélioration de la détection et de la fragmentation de peptides dérivés à l'éthyl-triphenylphosphine en FAB-CAD-MS/MS [2]. En 1995, Liao et Allison publie une lettre dans Journal of Mass Spectrometry décrivant une amélioration de la détection en MALDI-MS des peptides dérivés par des réactifs portant des charges permanentes dont les phenylphosphines [3]. L'équipe de Watson a ensuite introduit des dérivés du TMPP en 1997 pour la dérivation des amines N-terminales des peptides et leur analyse en FAB et MALDI-MS [4], puis le TMPP-Ac-OSu en 1999 pour la fragmentation améliorée des peptides en MALDI-PSD (Post Source Decay)-MS [5]. La même année Adamczyk décrit la fragmentation de peptides dérivés au TMPP-Ac-OSu

en ESI-trappe d'ions [6]. Une autre application pour la caractérisation MALDI-MS des sites de O-glycosylation fut introduite par l'équipe de Lemoine en 2004 [7].

On constate ainsi, que si les conditions de réaction de réactif sur des peptides modèles sont déjà décrites dans la littérature, elles ne sont pas adaptées aux extraits protéiques complexes en présence de détergents. De plus le contrôle de la spécificité du marquage est nécessaire : par exemple si le marquage spécifique des amines N-terminales des protéines est souhaité, des réactions non spécifiques sur les amines des chaînes latérales des lysines et les groupements hydroxyle des tyrosines peuvent avoir lieu. En 2009, notre laboratoire publie la première étude à l'échelle protéomique utilisant le TMPP-Ac-OSu et ceci est dans un cadre protéogénomique pour l'annotation du génome de *M. smegmatis* et la correction des prédictions des codons d'initiation [1]. En 2010, l'équipe d'Armengaud du CEA réalise la même étude sur la bactérie extrêmophile *D. deserti* et met en évidence des codons d'initiation non canoniques [8]. La stratégie de marquage des peptides N-terminaux par le TMPP-Ac-OSu est appelée N-TOP (N-Terminal Oriented Proteomics).

## 2. Les avantages de la dérivation N-terminale au TMPP pour l'identification des peptides N-terminaux

La stratégie N-TOP est une stratégie originale d'enrichissement et de marquage des amines N-terminales des protéines. Elle consiste à modifier sélectivement l'acide aminé N-terminal des protéines de manière à générer, après digestion enzymatique, des peptides N-terminaux marqués tout en laissant intacts les peptides internes. Le réactif TMPP-Ac-OSu utilisé pour la dérivation et le marquage de ces amines permet :

- Une amélioration de l'efficacité d'ionisation des peptides N-terminaux en mode électrospray grâce à la charge permanente apportée par la phosphine ;
- Une fragmentation particulière des peptides marqués en MS/MS ;
- Une augmentation de leur hydrophobie et par suite de leur temps de rétention en RP-HPLC ce qui les sépare des peptides internes et favorise leur échantillonnage par le spectromètre de masse pour la MS/MS en mode DDA (Data Dependent Acquisition).

Lorsque cette réaction est effectuée d'une manière sélective sur les amines N-terminales (Figure 1) des protéines à pH contrôlé, les peptides N-terminaux ainsi dérivés profitent des améliorations citées ci-dessus. Ceci permet donc d'identifier et de caractériser sans ambiguïté les extrémités N-terminales des protéines.

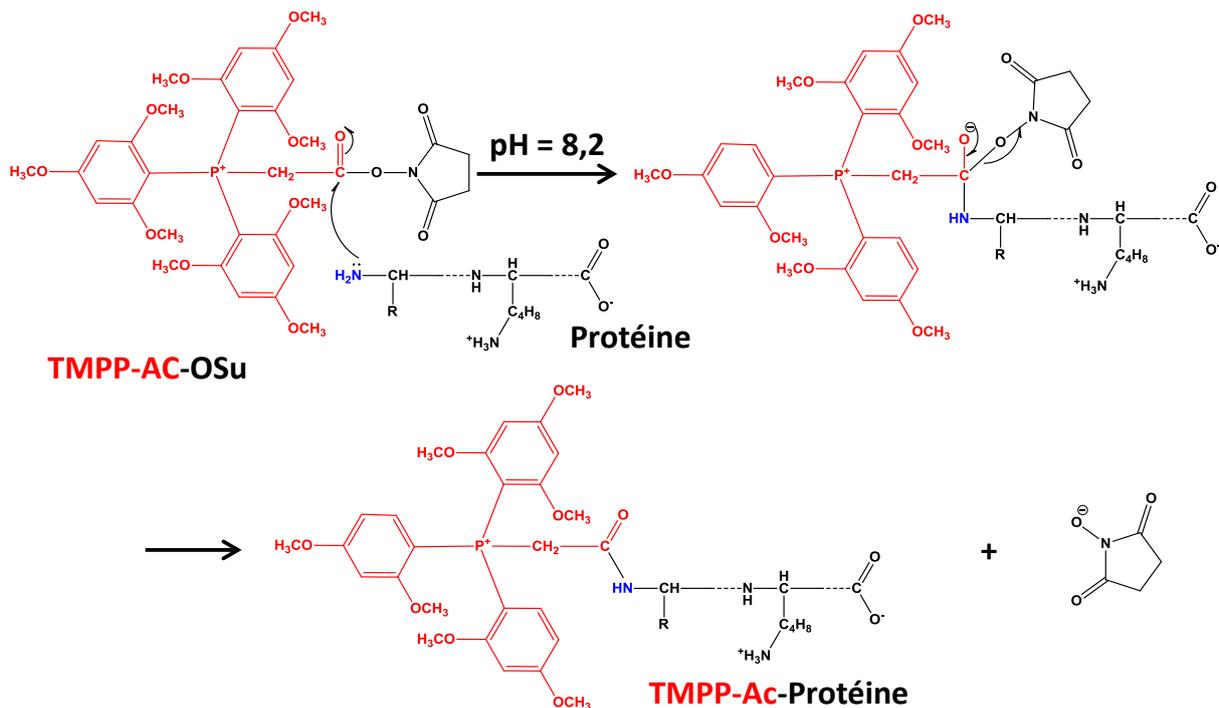


Figure 1 : Réaction de dérivation des amines N-terminales des protéines par le TMPP-Ac-OSu. Le contrôle du pH à 8,2 permet le marquage sélectif et spécifique des amines N-terminales en laissant celles des chaînes latérales des lysines intactes.

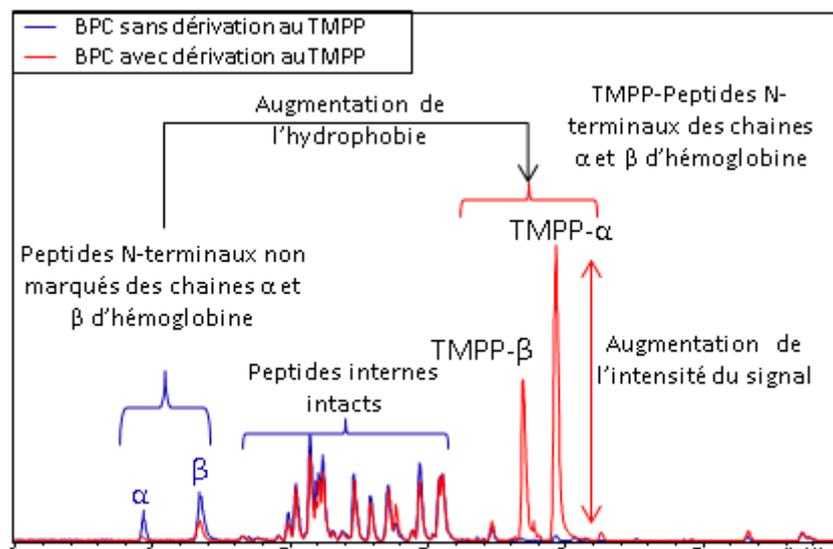


Figure 2 : Comportement des peptides marqués au TMPP par rapport aux peptides non marqués en LC-MS.

Cette réaction de marquage des protéines dans un mélange complexe nécessite l'utilisation d'un large excès de réactif et un contrôle minutieux du pH. Une étape de gel d'électrophorèse 1D est nécessaire après la réaction de marquage pour l'élimination de l'excès de réactif et de ces produits de dégradation. Cette étape de gel permet aussi de fractionner l'échantillon en séparant les protéines selon leurs poids moléculaires. La figure 2 illustre les avantages du marquage N-terminal des protéines au TMPP-Ac-OSu. On remarque

l'augmentation considérable de l'intensité des peptides marqués par rapport à leurs homologues non marqués. L'élution tardive de ces peptides permet de les séparer de l'ensemble des peptides internes. On note aussi que la détection des peptides internes n'est pas altérée puisque ceux-ci sont conservés et continueront à être sélectionnés, fragmentés et identifiés par MS/MS.

Afin de profiter des avantages de cette stratégie, nous avons étendu son champ d'application à l'étude des processus protéolytiques. En effet, contrairement aux approches présentées dans le chapitre 4 de la partie I, cette stratégie permet de marquer spécifiquement les extrémités N-terminales des protéines et d'améliorer leur détection tout en conservant les peptides internes. De plus, elle est simple à mettre en place et n'ajoute qu'une étape de marquage (1 heure) à un protocole d'analyse protéomique classique. Nous avons apporté trois nouveaux développements à cette stratégie pour l'adapter à différentes problématiques en biologie. Ces développements sont décrits dans les trois chapitres suivants.

## Chapitre 1 Mise au point de la dérivation N-terminale des protéines dans le gel : La stratégie « Ingel N-TOP »

Dans de nombreux cas, la caractérisation de l'extrémité N-terminale d'une ou de quelques protéines particulières permet de répondre à certaines questions en biologie. Il s'agit souvent de protéines isolées dans une bande ou un spot de gel, de protéines purifiées par immunoprécipitation ou de protéines recombinantes. Dans certains cas la détermination de la séquence N-terminale et du site de clivage d'une protéine permet d'élucider une voie de signalisation induite par un évènement, ou le mécanisme de translocation de cette protéine (voire Partie III Chapitre 2). Dans d'autres, il s'agit de déterminer la bonne maturation d'une protéine recombinante à visé thérapeutique telle que les anticorps monoclonaux. Or dans un grand nombre de cas, ces protéines se trouvent purifiées dans des tampons qui ne sont pas compatibles avec la réaction de marquage au TMPP. Les anticorps monoclonaux sont par exemple souvent formulés dans un tampon histidine et à pH acide. Dans d'autre, il s'agit d'une protéine faiblement abondante et qui se trouve dans un mélange de protéines contaminantes et dans un tampon non compatible avec la réaction de marquage suite à une immunoprécipitation peu spécifiques. Ceci est le cas de la protéine STEVOR du parasite de la malaria *Plasmodium falciparum*. Cette protéine était détectable au Western Blot mais n'était pas visible avec une coloration au bleu de Coomassie. Pour répondre à ces problématiques, j'ai adapté la stratégie N-TOP aux protéines séparées sur un gel d'électrophorèse SDS PAGE. La réaction de marquage au TMPP-AC-OSu est effectuée d'une manière à cibler la bande de gel 1D contenant la protéine d'intérêt.

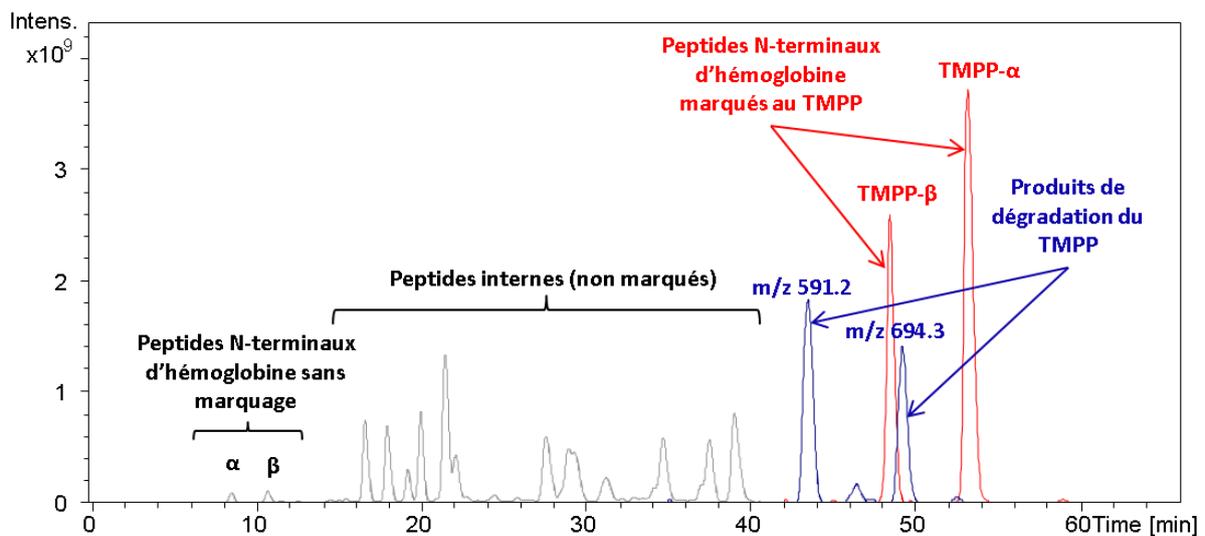
### 1. Développement de la stratégie Ingel N-TOP

Dans le but de développer cette stratégie, nous avons effectué nos optimisations sur des protéines modèles, les chaînes alpha et beta de l'hémoglobine. Les protéines sont déposées et migrées sur un gel d'électrophorèse SDS PAGE et en suite, les bandes sont découpées, décolorées et déshydratées à l'aide d'un système automatisé. Plusieurs tampons pour la réaction furent testés, le bicarbonate d'ammonium, le tris-HCl, le tris-HCl avec des agents chaotropiques telle que l'urée, toujours à pH strictement contrôlé à 8,2. Les meilleurs résultats furent obtenus avec le Tris-HCl additionné d'urée. Nous avons aussi remarqué un

meilleur rendement de la réaction en ajoutant tout d'abord le TMPP dans de l'eau/acétonitrile 80/20 directement sur le gel déshydraté suivi de l'ajout du tampon à pH = 8,2. Ceci permet probablement une pénétration plus efficace du réactif de marquage dans le gel.

La quantité de TMPP-Ac-OSu à ajouter a été optimisée à 400 équivalents, ceci est dans le cas où l'on connaît la quantité de protéine dans la bande de gel. Dans le cas contraire, plusieurs essais nous ont conduit à considérer que lorsqu'on observe une bande, même légère, au bleu de Coomassie, 100 µg de TMPP est une quantité raisonnable, bien que ce soit bien au-delà des 400 équivalents requis, vu que l'étape ultérieure d'élimination de l'excès de réactif fût bien optimisée à posteriori.

En ce qui concerne le maillage du gel, différents pourcentages en acrylamide allant de 4 jusqu'à 16 % furent essayés sans qu'une différence significative soit observée. Il semble donc que le réactif arrive à pénétrer efficacement dans le gel indépendamment du pourcentage d'acrylamide. Le temps de réaction optimal observé est d'une heure pour le marquage. Une étape d'incubation d'une demi heure avec de l'hydroxylamine fût ajoutée pour quencher la réaction et éliminer les éventuels marquages non spécifiques sur les groupements hydroxyles des tyrosines.



**Figure 1 : Chromatogramme LC-MS d'un digest d'hémoglobine  $\alpha$  et  $\beta$  montrant les produits de dégradation du TMPP (en bleu) qui élue dans la zone d'élution des peptides N-terminaux marqués au TMPP (en rouge).**

Le problème qui s'est posé pour ce protocole était l'élimination des produits de dégradation du réactif de marquage. En effet, le TMPP-Ac-OSu n'est pas stable dans l'eau et dans le temps, il s'hydrolyse facilement pour donner essentiellement une forme acide (TMPP-Ac-OH) à 591,2 m/z. Nous avons observé, aussi une espèce à 694,3 m/z qui correspondrait à la masse de la réaction du TMPP-Ac-OSu avec le Tris à savoir (TMPP-Ac-NH-C(CH<sub>2</sub>OH)<sub>3</sub>). Ces espèces parasites sont éluées dans la même zone d'élution que les peptides N-terminaux marqués au TMPP lors des analyses en LC-MS et LC-MS/MS (Figure 1).

Finalement la solution était de multiplier, jusqu'à quatre fois, les cycles de lavage et de déshydratation des gels en les mettant dans des puits propres à la fin de chaque cycle. Cette étape a été automatisée grâce au robot Massprep (Waters, Manchester). Chaque cycle comprend quatre lavages avec de l'hydrogencarbonate d'ammonium 25 mM et de l'acétonitrile et une étape de déshydratation des morceaux de gel à l'aide de l'acétonitrile.

La figure 2 montre les étapes du protocole de la stratégie Ingel N-TOP comparées avec celles de la stratégie de dérivation en solution N-TOP.

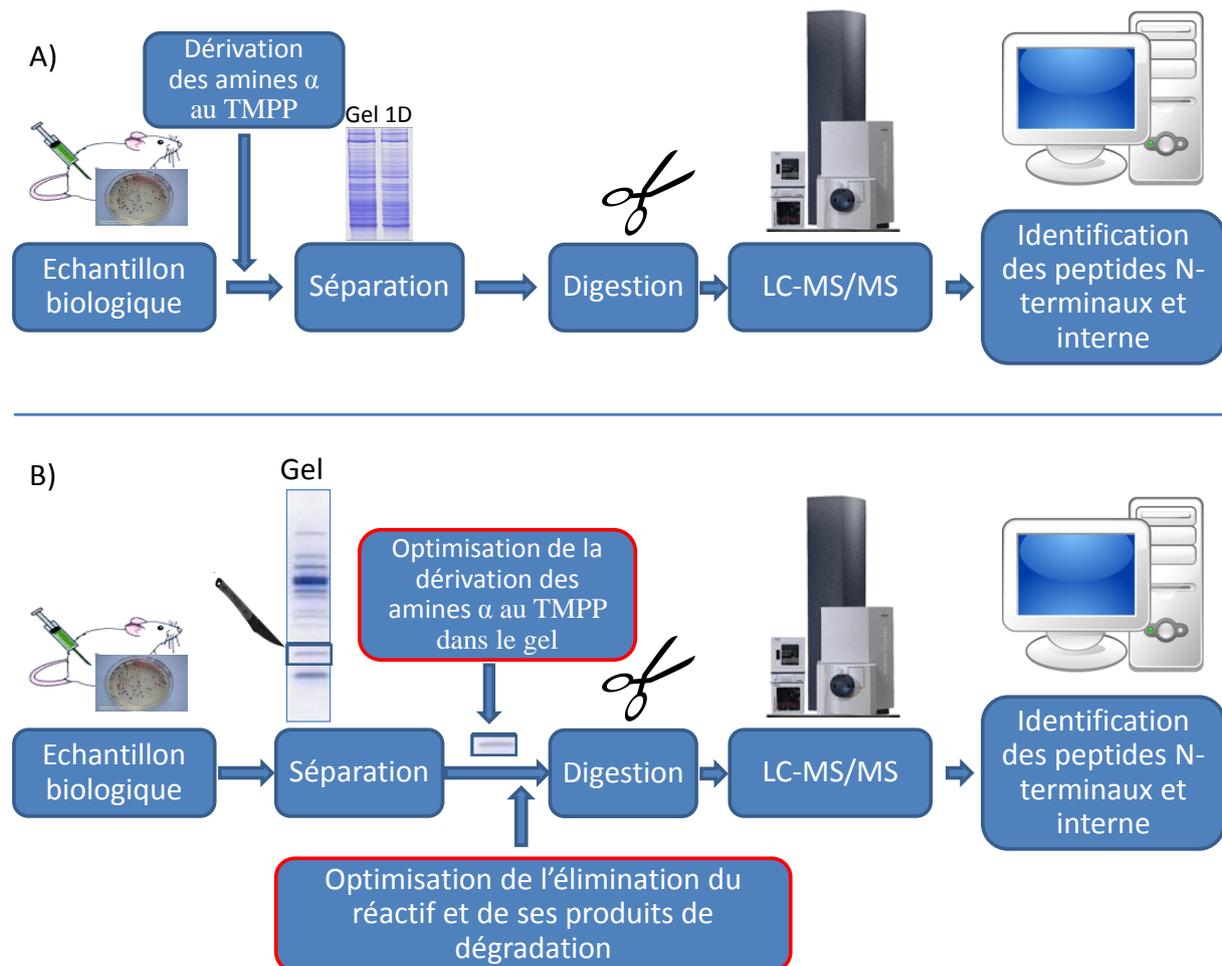


Figure 2 : A) Protocole expérimental de la stratégie N-TOP de dérivation en solution des amines  $\alpha$  des protéines et B) protocole expérimental de la stratégie Ingel N-TOP de dérivation ciblée dans le gel avec les principales étapes optimisées.

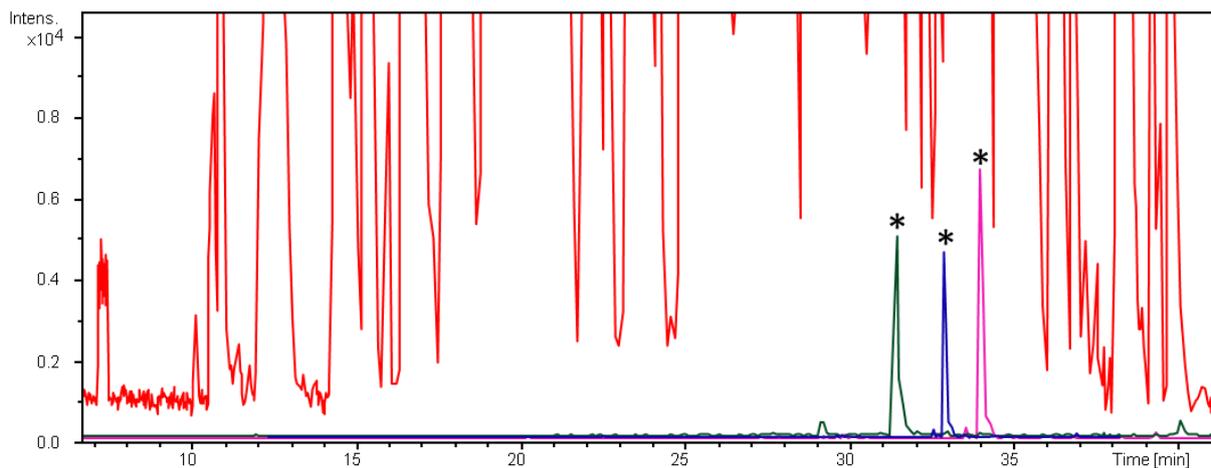
## 2. Rendement de la réaction

Les résultats obtenus sur les protéines modèles montrent que la réaction de dérivation se fait avec succès. Le rendement de la réaction est comparable à celui du protocole de dérivation en solution. En effet, après dérivation dans le gel et digestion trypsique puis analyse en LC-MS, l'extraction des courants d'ions des deux peptides N-terminaux des chaînes alpha et beta de l'hémoglobine, marqués et non marqués au TMPP,

permet d'observer que la réaction est presque totale pour la chaîne alpha et que le pic du peptide N-terminal non marqué de la chaîne beta est de faible intensité. En plus, l'élimination de l'excès de réactif s'est fait avec succès, les pics correspondants aux produits d'hydrolyse du TMPP-Ac-OSu sont de très faible intensité.

### 3. La régiosélectivité et spécificité du marquage au TMPP dans le gel

La régiosélectivité de la dérivation semble être toujours respectée. La dérivation aspécifique sur les chaînes latérales des lysines est mineure et du même ordre observé dans le protocole de dérivation en solution. Ceci est bien évidemment grâce au contrôle du pH à 8,2 durant la réaction de dérivation qui confère une régiosélectivité de 95 % sur les amines N-terminales libres dont le pKa approximatif est de 7,8 vis-à-vis des amines libres des lysines dont le pKa est autour de 11. Comme le montre la figure 3, il y a effectivement quelques peptides qui voient leur lysine modifiée par le TMPP. Cependant, le taux de modification des lysines reste faible, l'intensité des pics correspondants à des peptides internes ayant des lysines modifiées représente environ 3 % de celle des peptides N-terminaux. Ces réactions parasites restent donc mineures.



**Figure 3 :** Zoom sur une superposition du chromatogramme BPC d'une analyse LC-MS/MS d'un digest d'hémoglobine dérivée au TMPP dans le gel et de l'extrait des courants d'ions correspondant à des peptides internes d'hémoglobine ayant subi une réaction parasite sur la chaîne latérale de leur lysine.

Toutefois, il faudra quand même tenir compte de ces réactions parasites lors de l'interprétation des résultats et vérifier, lorsqu'il y a une possible ambiguïté dans la séquence du peptide (des lysines proches du N-terminus du peptide), s'il y a assez de fragment N-terminaux sur le spectre MS/MS pour identifier avec confiance un marquage spécifique et donc un peptide N-terminal ou bien une réaction parasite. Ceci est facilité par la fragmentation particulière des peptides marqués au TMPP en mode CID. En effet, la fragmentation CID de ces peptides génère essentiellement des fragments dont la charge est portée par le côté N-terminal (fragments de type a et b). Nous avons vu dans le chapitre 1 de la partie bibliographique (paragraphe 1.4.2.1) que la fragmentation CID des peptides tryptiques génère principalement des fragments b et y expliqués par le modèle du proton

mobile. Or sur un peptide dérivé au TMPP, les deux charges sont séquestrées, d'une part la charge est permanente sur le groupement phosphine et d'autre part, la deuxième charge apportée par le proton est séquestrée sur la chaîne latérale de la lysine ou de l'arginine. La mobilisation et la délocalisation de ce proton sur le squelette peptidique permet au peptide de fragmenter et de générer des ions b et y avec des ions b majoritaires à cause de la charge retenue sur le phosphonium N-terminal. Cependant, cette mobilisation nécessite une énergie conséquente (voir Figure 14, chapitre 1 de la partie Bibliographie), et par suite le mécanisme de fragmentation distant de la charge devient plus compétitif en terme d'énergie requise, ce qui semble expliquer cette fragmentation particulière et la présence majoritaire de fragments a et b. La figure 4 illustre cette différence dans le type de fragments observés pour un même peptide sans et avec marquage au TMPP.

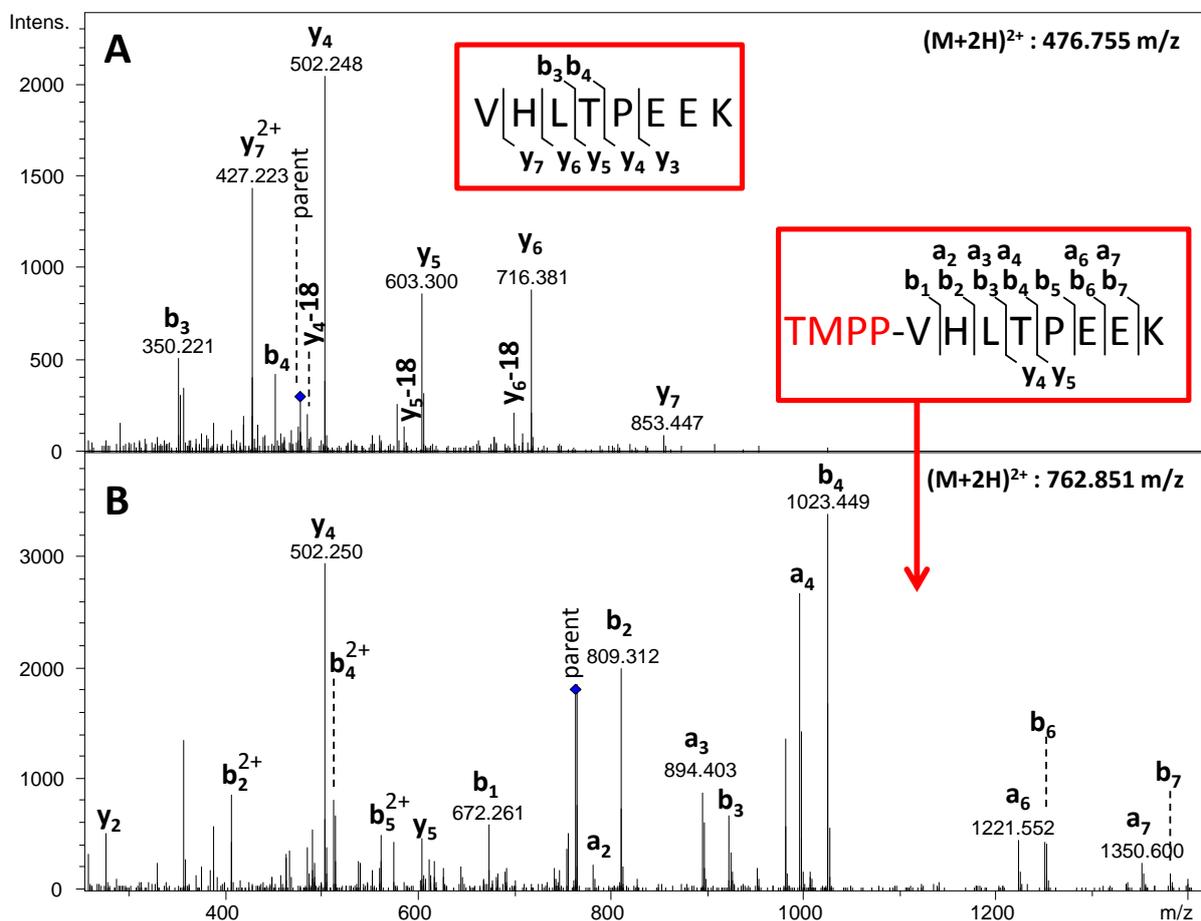


Figure 4 : Fragmentation CID du peptide N-terminal VHLTPEEK sans dérivation (A) et avec dérivation au TMPP (B). On observe sur le spectre A majoritairement des fragments y. Sur le spectre B on observe majoritairement des fragments a et b et très peu de fragments y, mis à part le fragment y<sub>4</sub> qui présente un site de fragmentation favorable dû au résidu proline.

## 4. Conclusion

Cette stratégie Ingel N-TOP permet donc de caractériser sans ambiguïté les extrémités N-terminales des protéines. C'est une approche ciblée de la stratégie N-TOP ouvrant la voie pour la caractérisation de protéines purifiées en très petites quantités. Nos résultats montrent que le rendement et la régiosélectivité de la réaction sont comparables à ceux de la réaction de dérivation en solution. Cette stratégie est donc particulièrement intéressante lorsqu'il s'agit d'étudier un substrat de protéase en particulier ou la maturation d'une protéine particulière. Une application dans le cadre de l'étude du mécanisme de maturation et de translocation de la protéine STEVOR de *Plasmodium falciparum* est décrite dans le chapitre 2 de la Partie III.

## Chapitre 2 Mise au point de la stratégie qN-TOP : Dérivation avec du TMPP $^{13}\text{C}$ pour la quantification du protéome N-terminal

Fort de notre expérience dans la dérivation et la caractérisation des extrémités N-terminales des protéines ; nous avons décidé d'ajouter une dimension quantitative à la stratégie N-TOP à l'aide d'un réactif TMPP marqué aux isotopes stables. Cette stratégie appelée qN-TOP pour « quantitative N-terminal Oriented Proteomics » permettrait d'effectuer des études protéomiques différentielles des processus protéolytiques. Elle est donc, à la différence des stratégies de quantification par marquage chimique isotopique classique telles que ICPL et ICAT, orientée vers la quantification des peptides N-terminaux et ne servira pas pour la quantification des niveaux d'expression des protéines. Néanmoins, elle présente tous les avantages de la stratégie N-TOP classique, notamment la conservation des peptides internes.

### 1. Réactifs et marquage différentiel qN-TOP

Comme toute technique de marquage isotopique différentielle, cette stratégie s'appuie sur l'utilisation de deux formes d'un même réactif, une forme dite lourde contenant des isotopes stables et l'autre dite légère et contenant les isotopes naturels (voir partie bibliographique, chapitre 3, paragraphe 3.3.1). Pour le TMPP-Ac-OSu, nous avons fait synthétiser une forme lourde dans laquelle nous avons substitué neuf atomes de carbone  $^{12}\text{C}$  par neuf atomes de carbone  $^{13}\text{C}$ . En effet, contrairement au marquage isotopique au deutérium, le marquage au  $^{13}\text{C}$  est connu pour ne pas induire des shifts dans les temps d'élution et de permettre une coélution parfaite avec son homologue au  $^{12}\text{C}$  [9]. Le choix de substituer 9 atomes de carbone, tient compte des contraintes de la synthèse d'une part, et permet d'autre part d'obtenir une différence de masse suffisante pour différencier, par spectrométrie de masse, les enveloppes isotopiques de deux peptides marqués par l'une et l'autre des formes du TMPP. Cette synthèse a été réalisée par la société « Alsachim » à Strasbourg. La figure 1 illustre les deux réactifs, léger et lourd, du TMPP-AC-OSu utilisées dans cette stratégie de quantification des protéomes N-terminaux.

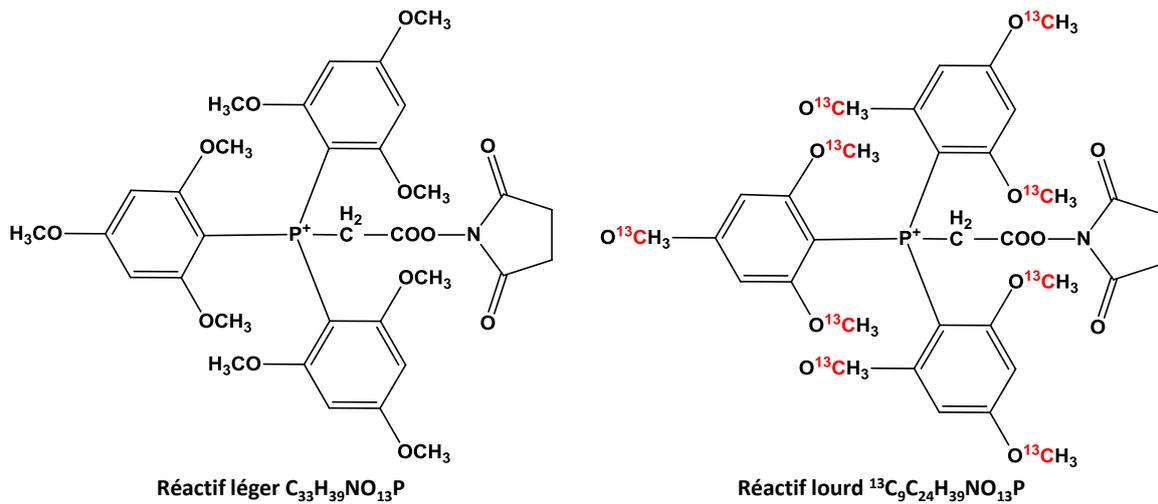


Figure 1 : Structure chimique de la forme légère (à gauche) du TMPP-Ac-OSu et de la forme lourde (à droite) avec 9 atomes de carbone 13, cette forme sera désignée par  $[^{13}C_9]$ TMPP-Ac-OSu.

## 2. Propriétés du réactif $[^{13}C_9]$ TMPP-Ac-OSu

Afin de valider ce nouveau réactif  $[^{13}C_9]$  TMPP-Ac-OSu, nous avons tout d'abord vérifié sa pureté par divers méthodes analytiques (LC-MS, ESI-MS, RMN). Nous avons ensuite vérifié son comportement dans une analyse N-TOP. Pour cela, nous avons réalisé des expériences tests en marquant des protéines modèles avec ce réactif. Nos résultats montrent que le TMPP lourd possède un comportement similaire au TMPP classique au niveau réactivité, rendement et réponse en MS. Ceci est bien évidemment attendu vu que les propriétés physico-chimiques des deux réactifs, mis à part leur masse moléculaire, sont strictement identiques. Les peptides marqués au TMPP lourd et léger coéluent parfaitement en chromatographie liquide comme le montre la figure 2. Cette coélution a été vérifiée dans des matrices simples (digest de protéines modèles) et des matrices complexes (digest de lysat d'un tissu).

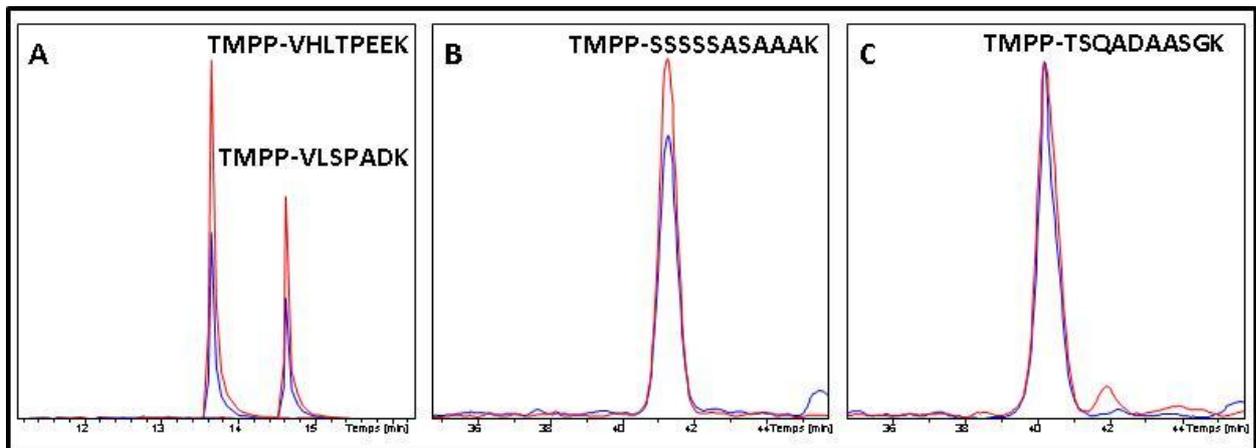


Figure 2 : Superposition d'extraits de courant d'ions de peptides homologues marqués d'une manière différentielle au TMPP lourd (bleu) et léger (rouge). Le chromatogramme A représente les peptides N-terminaux de protéines modèles (chaines  $\alpha$  et  $\beta$  de l'hémoglobine). Les chromatogrammes B et C proviennent de l'analyse d'un lysat entier de tissus de foie de rats.

### 3. Déroulement de la stratégie qN-TOP

Le déroulement d'une expérience qN-TOP est similaire à celui d'une expérience de marquage chimique différentielle classique avec cependant quelques particularités. Les protéines sont extraites des deux échantillons à comparer puis dosées. Un échantillon est marqué avec du  $[^{13}\text{C}_9]$ -TMPP et l'autre est marqué au TMPP léger. Les deux extraits protéiques sont ensuite mélangés en quantités égales, puis les protéines sont séparées sur un gel SDS-PAGE monodimensionnel. Cette étape de gel, comme dans la stratégie N-TOP classique reste indispensable pour l'élimination des produits de dégradation des réactifs de marquage et pour décomplexifier l'échantillon. Le gel est ensuite découpé d'une façon systématique et puis une digestion enzymatique des protéines est réalisée dans le gel. Les peptides ainsi générés sont extraits des bandes de gel et analysés en LC-MS/MS. Les protéines sont ensuite identifiées grâce aux peptides internes. Leurs extrémités N-terminales sont caractérisées grâce au marquage N-terminal au TMPP. Les peptides N-terminaux ainsi identifiés, sont ensuite quantifiés de façon relative par comparaison des intensités des signaux correspondant aux peptides homologues marqués par la forme lourde et légère du TMPP (Figure 3). Cette quantification relative des peptides N-terminaux permet en effet de quantifier les processus protéolytiques que subit le protéome d'un échantillon par rapport à un autre dans divers conditions.

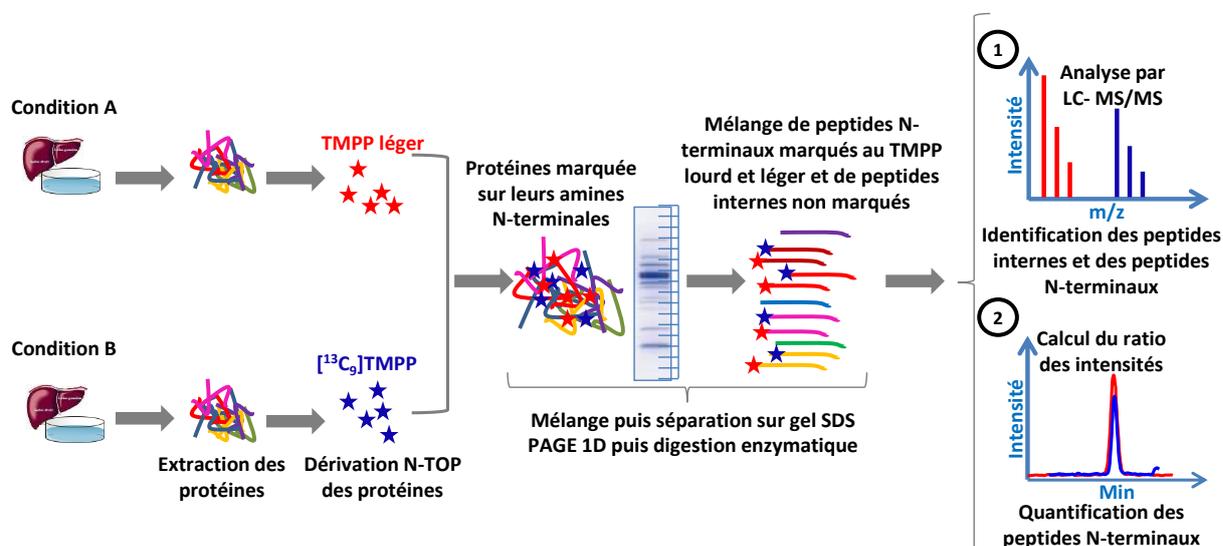


Figure 3 : Illustration de la stratégie qN-TOP pour la quantification relative des processus protéolytiques.

Le choix de l'instrument pour l'analyse LC-MS/MS s'est avéré très important pour la stratégie qN-TOP. Ceci sera détaillé dans le paragraphe suivant.

### 4. Choix et paramétrages instrumentaux pour la stratégie qN-TOP

#### 4.1. La bonne définition du pic chromatographique

Pour effectuer une quantification précise, il faut disposer d'un nombre suffisant de points pour décrire le pic chromatographique qui va servir pour la quantification. On estime

qu'il faut au moins 9 points pour décrire correctement un pic de forme gaussienne. Ces points correspondent aux spectres MS enregistrés. Pour ceci, une augmentation de la vitesse de balayage d'un spectromètre de masse permettra d'augmenter le nombre de spectres MS et donc le nombre de points pour décrire les pics et ainsi d'améliorer la quantification. Le réglage de la vitesse de balayage du spectromètre de masse en fonction des performances chromatographiques est donc critique. Cependant, avant de quantifier les peptides, il faut les identifier. Or dans une analyse LC-MS/MS le spectromètre de masse alterne les balayages MS et MS/MS. La qualité des spectres MS/MS dépend directement de la durée des balayages MS/MS. Il est donc nécessaire de trouver un compromis pour la durée des cycles MS et MS/MS permettant d'obtenir des spectres MS/MS informatifs et suffisamment de spectres MS pour décrire le pic chromatographique.

### 4.2. La linéarité de la réponse en analyse de masse

Au cours de notre développement de la stratégie N-TOP, nous disposions au laboratoire d'instruments de type trappe d'ions (Bruker) et d'un instrument de type Q-TOF (Synapt HDMS, Waters). Il était donc évident pour nous de choisir d'effectuer nos analyses sur l'instrument de type Q-TOF pour bénéficier de la résolution apportée par cet instrument (20000). Or nous nous sommes rendu compte en travaillant avec des protéines modèles que la gamme dynamique de cet instrument est très restreinte. En effet, cet instrument est équipé d'un détecteur MCP avec un digitaliseur de type TDC (« Time to Digital Converter »). Le digitaliseur a pour rôle de convertir les signaux électriques enregistrés par le détecteur en valeurs digitales, ici en l'occurrence, convertir les courants mesurés en spectres de masse. Un digitaliseur TDC est un module qui fournit une représentation digital des événements (pulsations, courants) enregistrés en fonction du temps. La saturation d'un digitaliseur TDC par un signal intense se traduit par des erreurs de mesure de masse, de résolution et d'intensité. Pour cet instrument, cette saturation est observée à partir d'une dynamique de détection de l'ordre de 10 au dessus du seuil de détection d'un composé. Ceci se traduit, en plus de l'erreur de mesure de la masse (déplacement du pic vers la gauche) par de fortes erreurs sur la mesure des ratios des intensités de deux peptides homologues (qui eux sont détectés simultanément). Il s'est donc avéré que ce spectromètre de masse n'est pas le mieux adapté pour une quantification avec une stratégie de dilution isotopique telle que qN-TOP.

Les instruments de type trappe d'ions présentent une gamme dynamique plus étendue que le Synapt HDMS. En effet, la quantité d'ions accumulés dans la trappe peut être contrôlée. Dès que cette quantité définie est atteinte, les ions accumulés sont éjectés et la trappe recommence un cycle de mesure. Si cette quantité n'est pas atteinte dans un temps prédéfini par l'utilisateur (de l'ordre de 100 ms), la trappe éjecte quand même les ions et recommence un cycle. Ceci fait que la trappe n'est jamais saturée et aucune saturation spectrale n'est atteinte. Cependant, un des inconvénients majeurs de ce type d'instrument est sa faible résolution. Ceci va entraîner des problèmes d'interférence très importants qui

vont pénaliser la justesse de la quantification. Ceci sera discuté plus en détail dans le chapitre 4 de la partie III.

A la fin de l'année 2010, le laboratoire a fait l'acquisition d'un instrument de type Q-TOF de dernière génération (MaXis, Bruker). Cet instrument affiche une résolution de 60000 et une gamme dynamique largement plus étendue que les instruments de génération antérieure. Cet instrument est équipé d'un détecteur de type ADC (Analog to Digital Converter) ou Convertisseur Analogique-Numérique. Ce détecteur travaille à une fréquence de digitalisation de 4 GHz. Un ADC, comme son nom l'indique, a pour rôle de convertir un signal analogique (dans notre cas l'intensité issue du détecteur) en une valeur numérique. L'association de cette valeur numérique au signal analogique est réalisée par intégration du signal sur une période d'échantillonnage qui correspond à l'inverse de la fréquence de digitalisation. Chaque période correspond à un temps de vol qui permet de calculer le  $m/z$ . A chaque fenêtre de temps de vol est ainsi associée une valeur d'intensité. Dans un balayage donné, l'ensemble des intensités intégrées associées à chaque fenêtre de  $m/z$  est sommé pour constituer une intensité totale en fonction du  $m/z$  associé au temps de vol et former ainsi un spectre de masse. La précision de ce processus est limitée par l'erreur de quantification [10]. En effet, lorsque l'intensité issue du détecteur est supérieure au dernier niveau d'intensité encodé, l'ADC est à saturation et associera sa valeur d'intensité limite maximale. Le signal aura un sommet aplati. Ceci se traduit donc par une erreur sur l'intensité du signal. Cependant, la gamme dynamique de détection du MaXis de Bruker est plus étendue que celle du Synapt HDMS de Waters, elle est évaluée à 3 logs par le constructeur.

Nous avons d'abord évalué au laboratoire la gamme dynamique linéaire du MaXis en injectant des quantités croissantes d'un mélange de peptides tryptiques et en suivant l'intensité de deux ions différents. Cette expérience a montré que la réponse est linéaire sur une gamme étendue sur deux décades. Nous avons ensuite évalué la linéarité avec des peptides marqués au TMPP lourd et léger. Pour ce faire, nous avons effectué des mélanges d'une protéine modèle (chaîne  $\alpha$  de l'hémoglobine) marquée avec les deux formes de TMPP (léger et lourd) avec des ratios léger/lourd allant de 0,1 jusqu'à 100. Chaque solution correspondant à un ratio léger/lourd a été injectée trois fois afin d'évaluer la reproductibilité. Les résultats sont visibles sur le graphique de la figure 4. On observe que les ratios mesurés pour chaque mélange sont reproductibles avec des écarts types inférieurs à 3 %. D'autre part, on observe une réponse linéaire jusqu'au ratio 20. Au-delà de ce ratio, une déviation significative des ratios expérimentaux est observée par rapport aux ratios théoriques. La gamme linéaire est donc limitée à un ratio de 20 contrairement à ce qui a été obtenu dans la première expérience où nous avons eu une réponse linéaire sur une gamme de 2 logs. Ceci peut être expliqué d'une part par le fait que dans le cas d'une détection simultanée (peptides marqués au TMPP lourd et léger) la gamme linéaire est plus restreinte que dans le cas de l'injection de quantité croissante d'un peptide donné. D'autre part, la préparation des mélanges de peptides marqués léger et lourd nécessite des dilutions successives pour les ratios élevés ce qui peut introduire des erreurs sur ces ratios.

Ces résultats montrent que le MaXis présente en plus de sa résolution qui va de paire avec la précision de mesure de masse, une gamme dynamique linéaire compatible avec l'application de la stratégie qN-TOP. La gamme dynamique de linéarité de cette stratégie qN-TOP est comparable avec celles reportées pour les autres stratégies de quantification (1 à 2 logs) avec marquage chimique isotopique (ICPL, ICAT, ...) dans la littérature [11]. Il est cependant à noter que dans une expérience de quantification de modifications post-traductionnelles tel que le taux d'un clivage protéolytique, le calcul du ratio d'abondance est basé sur une seule observation (1 seul peptide) ce qui peut induire des erreurs de quantifications contrairement aux expériences de quantification de niveau d'expression des protéines où le plusieurs peptides contribuent au calcul du ratio d'abondance de leurs protéines correspondantes.

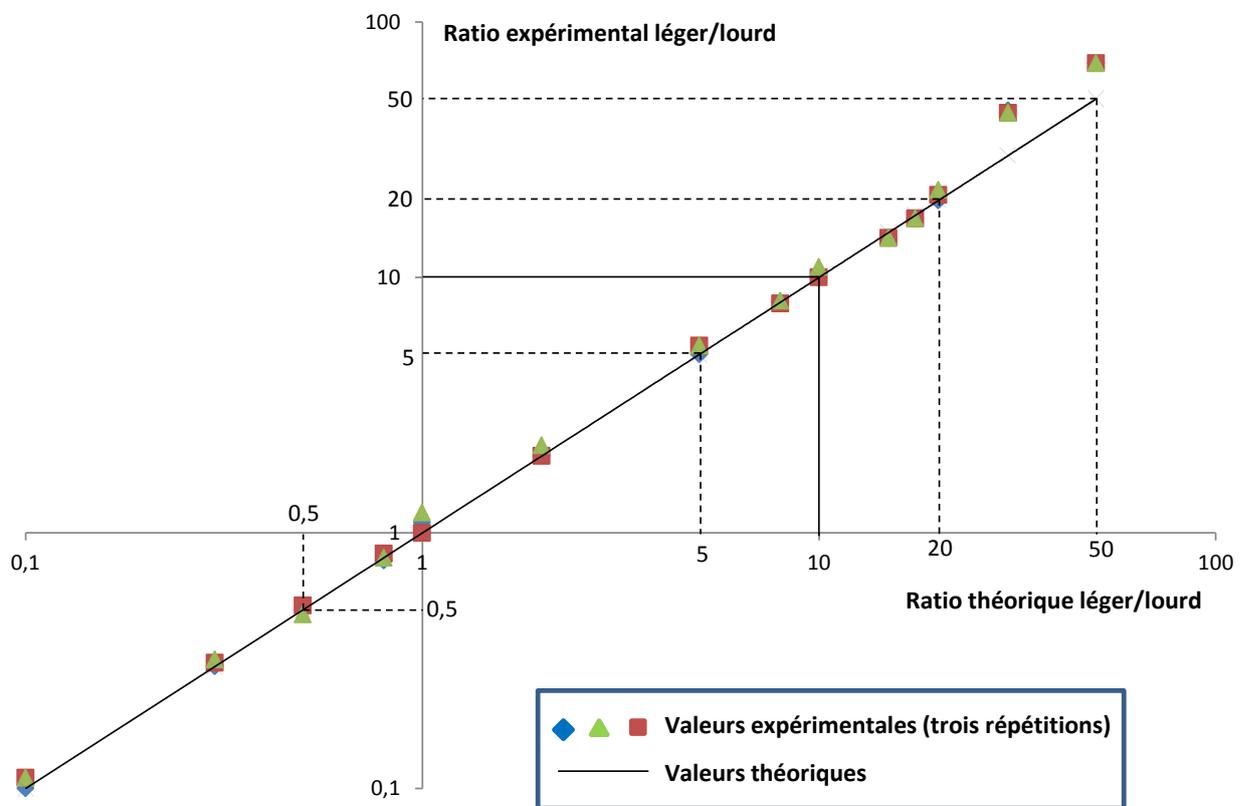


Figure 4 : Evaluation de la reproductibilité, justesse et linéarité de mesure des ratios d'abondance de deux peptides modèles homologues marqués par le TMPP léger et le  $[^{13}\text{C}_9]$ -TMPP.

## 5. Conclusion

Cette dimension quantitative ajoutée à la stratégie N-TOP permet une quantification relative des peptides N-terminaux et donc de comparer les taux de protéolyse entre deux échantillons biologiques. Ceci permettra d'élucider un nombre de mécanismes cellulaires, d'identifier des substrats de protéase et d'effectuer des études de dégradomique afin d'apporter des réponses à certaines questions biologiques. Cette méthode a l'avantage de conserver les peptides internes par rapport aux autres stratégies d'analyse des clivages

protéolytiques décrites dans le chapitre 4 de la partie I. De plus, elle est plus simple à appliquer puisqu'une seule étape de dérivation est ajoutée au protocole d'analyse protéomique. Il est cependant important d'accorder une attention particulière au choix du spectromètre de masse et au paramétrage instrumental pour les analyses LC-MS/MS. Une première application est décrite dans le chapitre 3 de la partie III dans le cadre d'une étude du jeûne prolongé chez le rat.



## Chapitre 3 Apport du marquage avec un nouveau TMPP plus hydrophobe pour enrichir d'avantage en peptides N-terminaux

La stratégie N-TOP permet d'enrichir en peptides N-terminaux en déplaçant l'élution des ces peptides vers la fin du gradient. Or d'après S. Gallien et coll., il reste quand même une partie de peptides internes plus hydrophobes qui coélue avec les TMPP-Ac-Peptides les plus hydrophiles (figure 1).

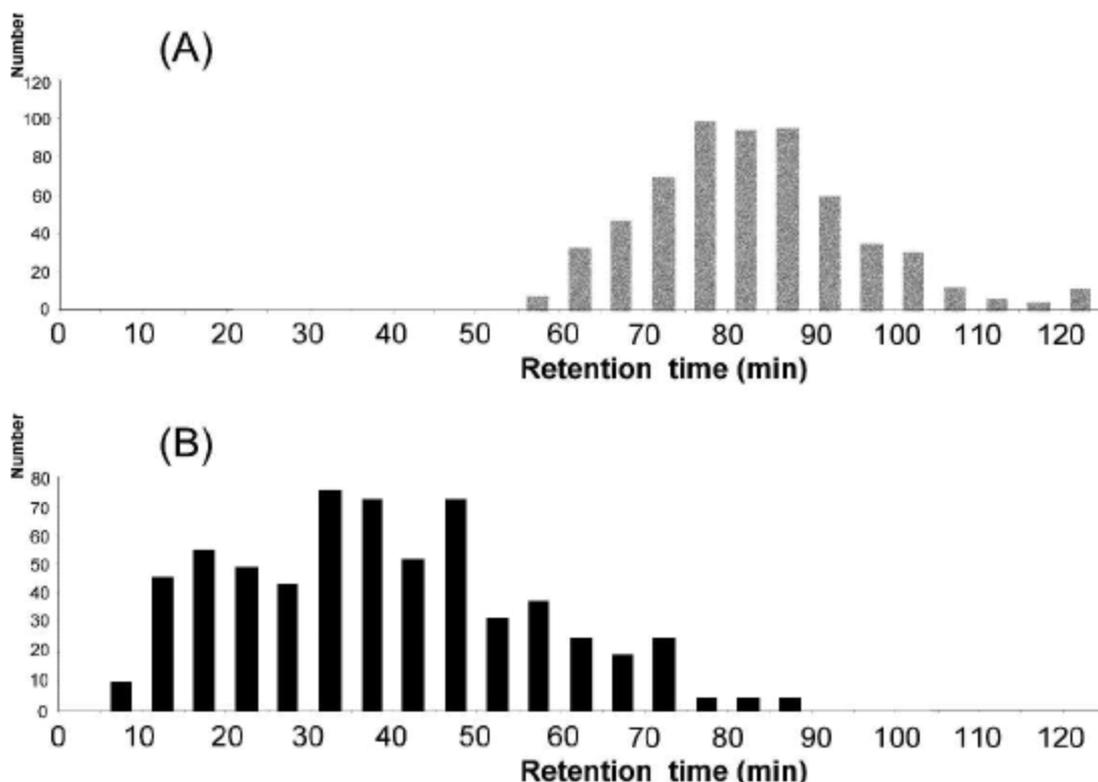


Figure 1 : Figure issue de [1] montrant une zone de coélution entre les peptides internes hydrophobes et le TMPP-Ac-Peptides N-terminaux. (A) temps de rétention des 591 TMPP-Ac-Peptide identifiés et (B) temps de rétention de 600 peptides internes choisis de manière aléatoire.

Dans le but d'augmenter davantage le rendement d'identification de peptides N-terminaux par la stratégie N-TOP, nous avons exploré la possibilité de modifier le réactif de marquage (le TMPP-Ac-OSu) en modifiant son hydrophobie. Ceci permettra d'augmenter

l'enrichissement chromatographique en peptide N-terminaux marqués et par suite de maximiser leur échantillonnage en mode d'acquisition donnée dépendant. La synthèse de ces nouveaux réactifs a été effectuée en collaboration avec le Dr. Z. Asfari (IPHC-DSA, UMR 7178, CNRS-UdS, Strasbourg).

## 1. Addition d'un groupement méthylène : le TMPP-Propanoyl-OSu

Nous avons tout d'abord commencé par l'ajout d'un groupement méthylène au TMPP-Ac-OSu de manière à ne pas modifier la réactivité du réactif. Le produit ainsi synthétisé est le bromure de (N-Succinimidylloxycarbonyl**ethyl**)tris(2,4,6-triméthoxyphényl)phosphine (Figure 2).

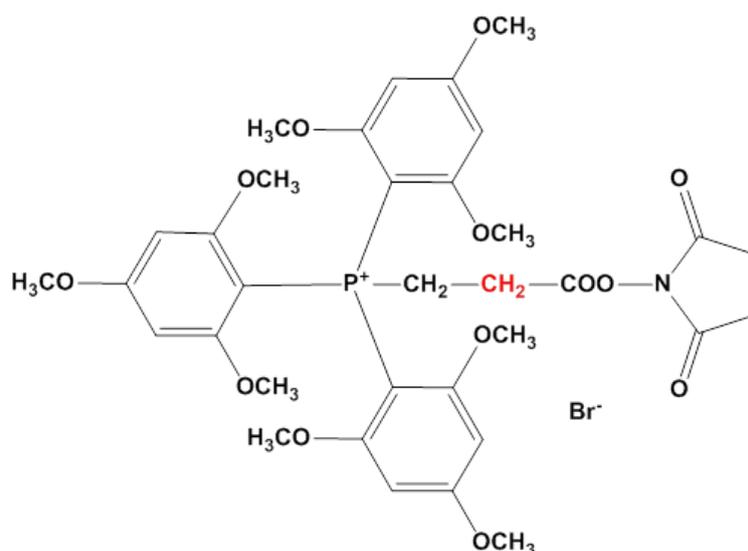


Figure 2 : Structure chimique du bromure de TMPP-Propanoyl-OSu

L'utilisation de ce nouveau réactif a montré qu'il a un comportement comparable à celui du TMPP-Ac-OSu au niveau rendement, sélectivité et fragmentation. Le temps d'élution des peptides marqués avec ce réactif n'ont pas augmenté d'une façon permettant de tirer un bénéfice de ce nouveau marquage. En effet, les peptides marqués avec le TMPP-Propanoyl-OSu éluent seulement une minute plus tard que les peptides marqués au TMPP-Ac-OSu. Cependant, ce résultat nous a été encourageant, puisqu'il montre que l'ajout d'un groupement méthylène dans cette position a un effet, même s'il n'est pas suffisant, sur le temps d'élution des peptides ainsi marqués. Ceci nous a donc permis de considérer d'ajouter au même endroit un groupement encore plus hydrophobe, un propylène.

## 2. Ajout d'un groupement propyl : le TMPP-Pentanoyl-OSu

Après optimisation de la synthèse du réactif TMPP-Pentanoyl-OSu ou Bromure de (N-Succinimidylloxycarbonyl**butyl**)tris(2,4,6-triméthoxyphényl)phosphine (Figure 3) ; le Dr. Asfari a réussi à le produire et à nous le fournir avec la pureté requise pour notre réaction. A la différence du TMPP-Ac-OSu et du TMPP-Propanoyl-OSu qui se présentaient sous forme de

poudre, ce nouveau réactif avait un aspect pâteux. Ceci rendait la pesée et la dissolution de ce produit moins facile, mais une solution a été trouvée en préparant des aliquots prêts à l'emploi.

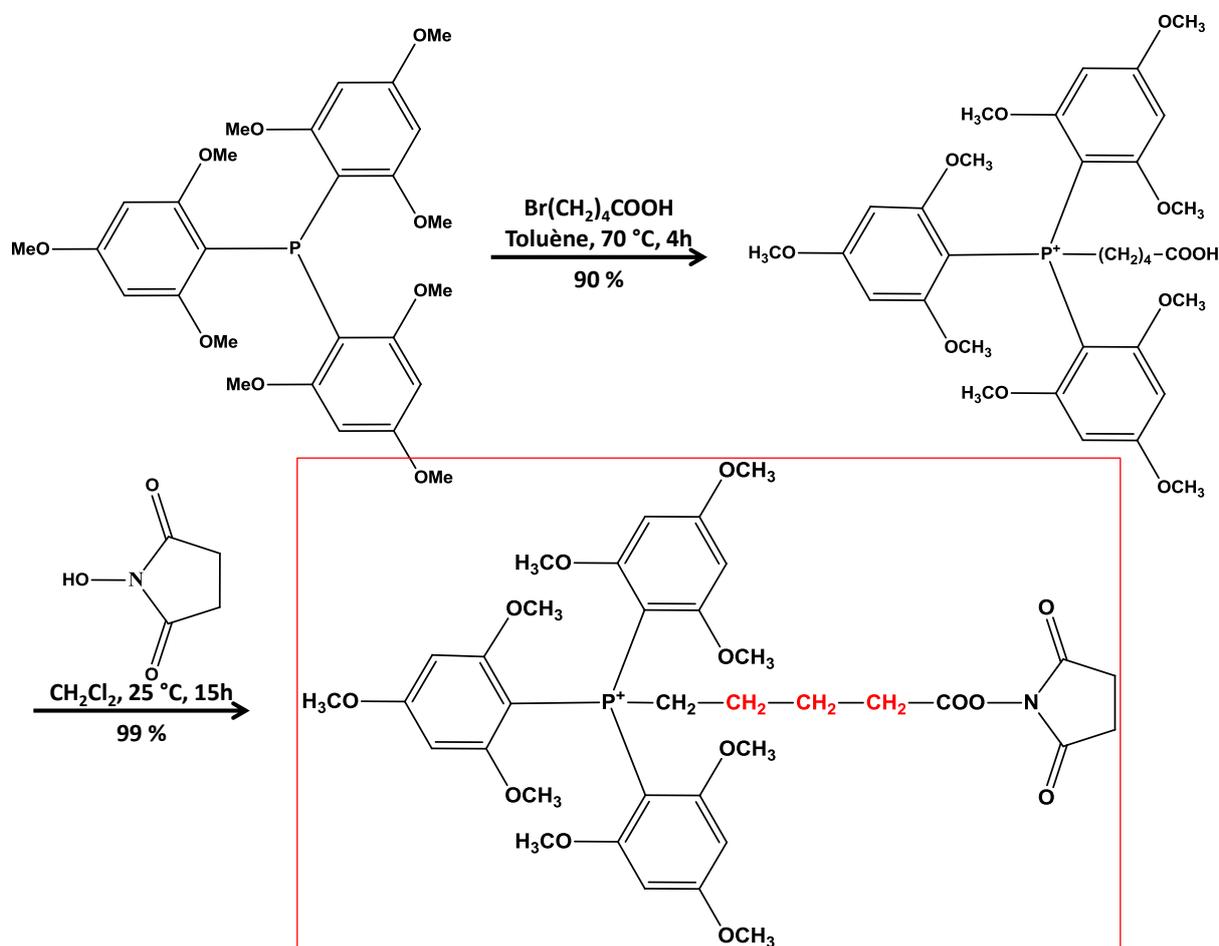


Figure 2 : Schéma de synthèse du TMPP-Pentanoyl-OSu

## 2.1. Application à des protéines modèles

Nous avons donc testé ce nouveau réactif plus hydrophobe sur nos protéines modèles habituelles, les chaînes  $\alpha$  et  $\beta$  de l'hémoglobine humaine. Le protocole de marquage utilisé pour le TMPP-Ac-OSu a été conservé. La réaction est donc effectuée dans un tampon contenant 6 M urée et 2 M thiourée et 1 % SDS à pH strictement contrôlé à 8,2 et avec 200 excès de TMPP-Pentanoyl-OSu pendant 1 heure à température ambiante. Afin d'éliminer les produits d'hydrolyse de ce réactif, l'étape de gel 1D est conservée. Comme dans le cas du TMPP-Ac-OSu, ces produits d'hydrolyse sont de natures comparables au réactif et éluent dans la même zone chromatographique que les peptides marqués et par suite peuvent gêner leur ionisation et leur sélection pour la MS/MS.

Ayant ce groupement propylène avec un effet inductif donneur inséré entre la phosphine (groupement attracteur d'électron) et l'ester, nous craignons que la fonction ester sera moins activée et par suite moins réactive à l'attaque nucléophile de l'amine N-

terminale des peptides. Cependant, nos essais ne montrent pas une réactivité plus faible de ce réactif par rapport au TMPP-Ac-OSu. Ceci est probablement dû à la forte réactivité intrinsèque de la fonction ester-succinimide (le succinimide étant un très bon nucléophile) et aux conditions de la réaction avec un excès considérable de ce réactif et un pH favorable à cette attaque. De plus la sélectivité du produit est conservée grâce au contrôle du pH à 8,2.

### 3. Augmentation de l'hydrophobie des peptides marqués avec le TMPP-Pentanoyl-OSu

Nos résultats montrent que les peptides marqués avec notre nouveau réactif sont effectivement plus hydrophobes et leur temps d'éluion est significativement augmenté. Comme le montrent la figure 3, les TMPP-Pentanoyl-peptides éluent 7 à 8 minutes plus tard que les TMPP-Ac-peptides si l'on utilise un gradient d'environ 0,5 % par minute d'acétonitrile dans l'eau. Ce gradient relativement lent est le gradient de choix dans l'analyse d'échantillon complexe avec la stratégie N-TOP [1,12].

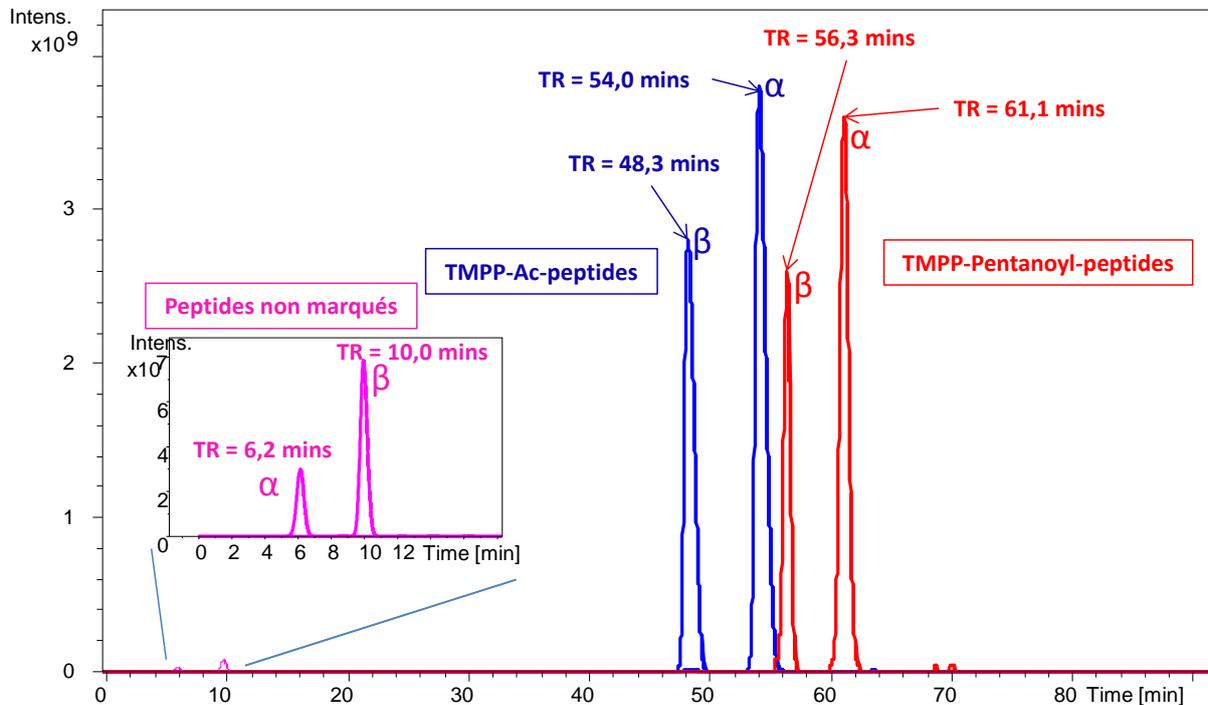


Figure 3 : Extraction des courants d'ions des peptides N-terminaux d'un mélange de deux digests d'hémoglobine marqués au TMPP-Ac-OSu et TMPP-Pentanoyl-OSu et analysés par LC-MS. Les peptides non marqués des chaînes  $\alpha$  et  $\beta$  sont en rose, ceux marqués avec le TMPP-Ac-OSu sont en bleu et ceux marqués avec le TMPP-Pentanoyl-OSu sont en rouge. Les temps de rétention (TR) de chacune des espèces sont indiqués sur la figure.

Cette augmentation du temps d'éluion permettra d'éloigner davantage les peptides N-terminaux marqués de la zone d'éluion des peptides marqués, même les plus hydrophobes. Ceci permettra de diminuer la compétition entre ces deux groupes de peptides et favorisera de part et d'autre leur détection. Or si l'on considère qu'un spectromètre de masse récent effectue en moyenne un demi-cycle MS et 6 MS/MS par

seconde, ceci permettra en théorie d'effectuer 210 cycles supplémentaires dans cette zone de 7 minutes permettant d'augmenter le rendement d'identification.

#### **4. Conclusion**

L'utilisation de notre nouveau réactif plus hydrophobe devra permettre d'après nos résultats d'augmenter, par rapport au réactif classique, le rendement d'identification en peptide N-terminaux grâce à leur séparation plus efficace des peptides internes. La stratégie N-TOP en soit n'est pas modifiée et le remplacement du réactif classique par le nouveau réactif se fait sans aucune autre modification dans les étapes de la stratégie. Une étude sur un mélange de protéine plus complexe (lysate cellulaire totale) est en cours. Elle permettra d'évaluer l'apport de ce réactif par rapport au TMPP-Ac-OSu dans une étude d'un protéome N-terminal.



## Références Bibliographiques

1. Gallien, S., Perrodou, E., Carapito, C., Deshayes, C., Reyrat, J. M., Van Dorsselaer, A., Poch, O., Schaeffer, C., and Lecompte, O., Ortho-proteogenomics: multiple proteomes investigation through orthology and a new MS-based protocol. *Genome Res* **2009**, *19* (1), 128-135.
2. Throck Watson, J., Wagner, D. S., Chang, Y.-S., Strahler, J. R., Hanash, S. M., and Gage, D. A., Characterization of the ethyl-triphenylphosphonium derivative of model peptides by fast atom bombardment collisionally-activated dissociation tandem mass spectrometry using B/E linked scans. *International Journal of Mass Spectrometry and Ion Processes* **1991**, *111* (0), 191-209.
3. Liao, P.-C., and Allison, J., Enhanced detection of peptides in matrix-assisted laser desorption/ionization mass spectrometry through the use of charge-localized derivatives. *Journal of Mass Spectrometry* **1995**, *30* (3), 511-512.
4. Huang, Z. H., Wu, J., Roth, K. D., Yang, Y., Gage, D. A., and Watson, J. T., A picomole-scale method for charge derivatization of peptides for sequence analysis by mass spectrometry. *Anal Chem* **1997**, *69* (2), 137-144.
5. Huang, Z. H., Shen, T., Wu, J., Gage, D. A., and Watson, J. T., Protein sequencing by matrix-assisted laser desorption ionization-postsource decay-mass spectrometry analysis of the N-Tris(2,4,6-trimethoxyphenyl)phosphine-acetylated tryptic digests. *Anal Biochem* **1999**, *268* (2), 305-317.
6. Adamczyk, M., Gebler, J. C., and Wu, J., Charge derivatization of peptides to simplify their sequencing with an ion trap mass spectrometer. *Rapid Commun Mass Spectrom* **1999**, *13* (14), 1413-1422.
7. Czeszak, X., Morelle, W., Ricart, G., Tetaert, D., and Lemoine, J., Localization of the O-glycosylated sites in peptides by fixed-charge derivatization with a phosphonium group. *Anal Chem* **2004**, *76* (15), 4320-4324.
8. Baudet, M., Ortet, P., Gaillard, J.-C., Fernandez, B., Guérin, P., Enjalbal, C., Subra, G., de Groot, A., Barakat, M., Dedieu, A., and Armengaud, J., Proteomics-based Refinement of *Deinococcus deserti* Genome Annotation Reveals an Unwonted Use of Non-canonical Translation Initiation Codons. *Molecular & Cellular Proteomics* **2010**, *9* (2), 415-426.
9. Chelius, D., and Bondarenko, P. V., Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry. *J Proteome Res* **2002**, *1* (4), 317-323.
10. Allen, P. E., and Holberg, D. R. (eds). (2002) *CMOS Analog Circuit Design*, Oxford University Press
11. Bantscheff, M., Schirle, M., Sweetman, G., Rick, J., and Kuster, B., Quantitative mass spectrometry in proteomics: a critical review. *Analytical and Bioanalytical Chemistry* **2007**, *389* (4), 1017-1031.
12. Gallien, S. (2009) Nouvelle méthodologies protéomiques d'aide à l'annotation des génomes et à la validation des séquences protéiques. in *Thèse de l'Institut Pluridisciplinaire Hubert Curien - Département des Sciences Analytiques*, Université de Strasbourg, Strasbourg



# Partie III. Applications de l'Analyse Protéomique en Biologie

---

## **Chapitre 1**

Etude du protéome mitochondrial humain : Application de la stratégie N-TOP pour la caractérisation des peptides de transit

## **Chapitre 2**

Etude du protéome du parasite *Plasmodium falciparum* et de l'export des protéines parasitaires à l'érythrocyte infecté. Application de la stratégie N-TOP et Ingel N-TOP

## **Chapitre 3**

Analyse protéomique du tissu adipeux du rat en réponse à une privation de nourriture plus ou moins prolongée

## **Chapitre 4**

Application de la stratégie qN-TOP à l'étude du protéome du foie du rat soumis au jeûne prolongé

## **Chapitre 5**

Etude fonctionnelle de ReP1-NCXSQ, une protéine qui régule l'échangeur  $\text{Na}^+/\text{Ca}^{2+}$  du nerf du calamar



# Chapitre 1 Etude du protéome mitochondrial humain : application de la stratégie N-TOP pour la caractérisation des peptides de transit

*Collaboration avec le CEA de Grenoble (Dr. Thierry Rabilloud) et le groupe CALIPHO de l'Institut Suisse de Bioinformatique (BIS) de Genève (Prof. Amos Bairoch)*

## 1. Résumé

Les mitochondries sont des organites cellulaires essentiels, impliquées, non seulement dans la production d'énergie, mais aussi dans plusieurs pathologies et voies métaboliques. La majorité des protéines mitochondriales sont produites à l'extérieur des mitochondries et importées dans ces dernières après plusieurs étapes de maturation, souvent par clivage protéolytique. Cette maturation peut altérer les caractéristiques fonctionnelles des protéines. Or les séquences des protéines présentes dans les banques protéiques sont issues de la traduction bioinformatiques des génomes et ne sont donc pas exhaustives, ni exemptes d'erreurs de séquence notamment au niveau des prédictions des séquences N-terminales. Il est donc important de valider expérimentalement la prédiction des séquences protéiques à partir du génome pour générer des banques protéiques « propres » à haute valeur ajoutée et les mettre au service de la communauté scientifique. Cette étude fait partie de l'effort collectif entrepris par le groupe CALIPHO (SIB, Genève) pour mettre en place une nouvelle génération de base de données protéiques de haute qualité centrée sur le protéome humain (neXtProt) et ceci dans le cadre du Projet Protéome Humain HPP (« Human Proteome Project ») tel qu'il est défini par HUPO (« HUman Proteome Organization »).

Dans ce contexte, nous avons appliqué la stratégie N-TOP de marquage sélectif des amines N-terminales des protéines par le TMPP-Ac-OSu développée au laboratoire (voir Partie I). Afin d'augmenter le nombre de protéines identifiées, nous avons réalisé des expériences complémentaires à chaque étape du protocole d'analyse que ce soit au niveau de la digestion, de l'analyse LC-MS/MS et du traitement des données.

Deux digestions enzymatiques différentes ont été effectuées pour augmenter le recouvrement des séquences protéiques et avoir accès à un maximum de peptides N-terminaux. Les analyses de nanoLC-MS/MS sont effectuées sur les instruments les plus performants au laboratoire en termes d'identification de peptides/proteines (Q-TOF et trappe ionique). De plus, les analyses sur la trappe ionique ont été réalisées avec deux modes de fragmentation différents : dissociation induite par collision (CID) et dissociation par transfert d'électrons (ETD). Deux moteurs de recherche ont été utilisés : Mascot qui est un logiciel propriétaire et OMSSA un logiciel libre (« open source »). Cette stratégie a permis l'identification d'environ 1600 protéines et de quelques centaines de peptides N-terminaux. Ces résultats ont été comparés avec les listes des protéines mitochondriales connues et leurs peptides de transit prédits (peptides clivés lors de la maturation) pour valider et corriger ces prédictions.

## **2. Cadre de l'étude dans le Projet Protéome Humain (HPP, Human Proteome organization (HuPO))**

Le but du Projet Protéome Humain (HPP) est de caractériser l'ensemble des 21000 gènes du génome humain connus afin de mettre en place une carte de l'architecture moléculaire de l'organisme humain basée sur les protéines. Cette carte sera une source d'informations précieuse qui permettra d'élucider la fonction moléculaire, le diagnostic préalable des maladies et le développement de traitements [1]. Ce projet a connu beaucoup de difficultés au départ, surtout après l'échec du Projet Plasma Humain [2]. Cependant, les progrès techniques et expérimentaux de la dernière décennie vont permettre une caractérisation poussée du protéome humain [3]. Le succès du Projet Génome Humain [4,5] a fourni une cartographie des protéines codées dans le génome et qui sont actives dans les centaines de types de cellules qui constituent le corps humain. Cette cartographie recense environ 20000 gènes codant pour des protéines [6,7]. Néanmoins, parmi ces 20000 gènes, seulement 12000 ont été observés expérimentalement au niveau protéique et environ 8000 (38%) n'ont jamais été observés et n'ont été mis en évidence qu'au niveau ARN messager. De plus, peu d'informations sont disponibles sur l'abondance, la distribution, la localisation cellulaire et la fonction pour un grand nombre des protéines validées.

Par opposition au génome, la dynamique du protéome humain est indéfinie due à son énorme complexité [8]. Le produit de l'expression de ces 20000 gènes atteint environ un million de protéines différentes dû à l'épissage alternatif et les différents types de modifications post-traductionnelles qui varient dans le temps et l'espace en fonction des états physiologiques et pathologiques. Pour cette cartographie du protéome humain permettant de recueillir des informations sur les protéines exprimées à partir de chaque locus du génome, HPP propose une approche centrée sur les gènes. Cette cartographie et toutes les informations collectées seront mises à la disposition de toute la communauté scientifique et sans restrictions, ce qui facilitera les études approfondies pour comprendre la biologie et les pathologies humaines [9].

Dans une telle approche centrée sur les gènes, les protéines sont définies en tant que produits des gènes de façon à classer tous les variants de protéines issus d'un gène donné en tant qu'une seule entité. Cela permettra à HPP de bénéficier pleinement des progrès du Projet Génome Humain [10] et de fournir les preuves si ces 8000 gènes restants codent pour des protéines *bona fide*. Pour cela, la mise en place d'une base de données regroupant les connaissances sur les protéines humaines constitue le socle du projet [11]. En même temps, il est de plus en plus évident qu'une grande partie des données de protéomique disponible n'est pas validée voire erronée, et ne peut donc pas être utilisée d'une façon fiable par la communauté des biologistes [2]. La nouvelle base de données neXtProt a donc comme objectif de remédier à ce problème, elle intégrera l'ensemble des informations de haute qualité disponible sur les protéines humaines, et sera construite sur la base d'une échelle de qualité des données. Dans ce contexte, et en collaboration avec le groupe CALIPHO qui a entrepris la mise en place de neXtProt et l'équipe du Dr. Thierry Rabilloud du CEA de Grenoble spécialiste des mitochondries, nous avons monté un projet dont le but est d'améliorer nos connaissances du protéome mitochondrial humain.

Comme il a été récemment montré [12,13], des analyses permettant l'identification des protéines par quelques peptides protéotypiques ou quelques épitopes sont incapables de fournir certaines informations biologiquement importantes. Ces informations incluent la durée de vie des protéines, les clivages protéolytiques qu'elles subissent, leurs localisations et leurs dynamiques ainsi que les différentes modifications post-traductionnelles qui modulent leurs fonctions, leurs activités et leurs interactions avec d'autres molécules biologiques. De plus, les projets visant à déterminer les niveaux d'expression de tous les produits des gènes dans différents tissus et cellules humaines, comme cela a été réalisé pour la levure [14], ainsi que les projets visant à fournir des informations importantes et de haute qualité sur les sous protéomes humains, occupent une place importante dans le cadre du Projet Protéome Humain. Notre étude s'inscrit dans cette catégorie de projets en visant à améliorer les connaissances sur les protéines des mitochondries humaines.

Cependant, ce projet ne doit pas être restreint à un simple recueil des protéines mitochondriales provenant d'un seul type de cellules humaines. Nous pensons qu'un projet d'une importante utilité doit être mis en place pour effectuer des analyses de protéomique comparative sur différents types cellulaires. Ces différents types de cellules peuvent être liés à des recherches cliniques vue que les mitochondries sont impliquées dans des maladies génétiques telle que MELAS, MERRF et le syndrome de Leber [15] et qu'elles sont suspectées d'être impliquées dans différentes pathologies multifactorielles comme certaines maladies neurodégénératives [16] et certains cancers [17].

### 3. Etat des connaissances au début du projet

Plusieurs études protéomiques des organites cellulaires en général et des mitochondries en particulier ont été effectuées. La première publication concernant le

protéome mitochondrial remonte à l'année 1998 [18]. Depuis, il y a eu plusieurs études dédiées à l'analyse protéomique des mitochondries visant à augmenter le recouvrement du protéome mitochondrial. Cependant, la plupart des données récentes de « haute qualité » ont été obtenues sur des tissus de rongeurs [19-21] ou sur la levure [22]. L'étude la plus récente effectuée sur des mitochondries humaines a permis d'identifier 150 protéines environ [23]. Une des difficultés majeures des études du protéome mitochondriale réside dans les quantités très faibles de tissus ou de cellules humaines disponibles pour préparer des mitochondries de pureté acceptable pour les analyses protéomiques. La pureté des préparations de mitochondries diminue drastiquement avec la diminution de la quantité de l'échantillon de départ. Une étude qui date de 2003 sur les mitochondries de cœur humain effectué par Taylor et coll. restera une exception et a permis d'identifier 615 protéines. Les auteurs indiquent qu'ils disposaient de trois cœurs humains et donc de suffisamment de matériel pour permettre une purification de mitochondries avec un rendement important [24]. Un examen poussé des données montrent une importante contamination par des protéines cytoplasmiques. Ceci constitue un obstacle pour l'obtention de données protéomiques sur les autres tissus et pour l'étude des variations intercellulaires des protéomes mitochondriaux.

Dans une étude intéressante [21], Pagliarini et coll. ont réalisé un large recueil de protéines mitochondrial provenant de quatorze tissus différents de souris. Pour différencier les protéines mitochondriales des contaminants, ils ont appliqué une approche dite de protéomique soustractive. Cette approche consiste à comparer un échantillon de mitochondries grossièrement purifié avec un échantillon de mitochondries issu d'une purification poussée. Les protéines mitochondriales devraient donc être enrichies dans le deuxième échantillon par rapport au premier. Enfin, dans le but de confirmer la nature mitochondriale de leurs protéines, ils ont effectué des expériences de marquage à la GFP et de microscopie de fluorescence. Dans leur approche quantitative, ils ont utilisé les pourcentages de recouvrement de séquence des protéines et la somme des aires des pics chromatographiques de toutes les séquences de peptides assignées aux protéines identifiées. Le pourcentage de recouvrement étant utile pour estimer les abondances relatives des différentes protéines entre elles, on observe néanmoins des erreurs significatives lorsqu'il s'agit de protéines abondantes. La somme des aires des pics est en corrélation avec les abondances relatives d'un échantillon à l'autre pour une protéine donnée mais difficile à utiliser pour comparer les abondances de différentes protéines entre elles vu que le nombre de peptides varie considérablement entre deux protéines différentes. Ils ont donc utilisé le pourcentage de recouvrement pour la comparaison des protéines entre elles et les aires des pics pour la comparaison des variations d'abondance d'un tissu à l'autre. Une des possibilités pour réaliser une cartographie des protéines mitochondriales humaines est d'utiliser ces données des mitochondries murines et procéder par orthologie avec celles des humains. Cependant les auteurs soulignent eux même clairement que leur étude souffre de certains biais importants :

- Premièrement, leur étude s'appuie fortement sur des comparaisons quantitatives pour étudier les variations du protéome mitochondrial entre les différents tissus et la vraisemblance de la nature mitochondriale des protéines identifiées (en comparant les extraits bruts avec les extraits purifiés). Or leurs approches quantitatives qui reposent sur l'analyse spectrale sont probablement peu précises et pas suffisamment reproductibles conduisant à un taux de 50 % de résultats non concluants.
- Deuxièmement, la confirmation de la localisation mitochondriale des protéines est effectuée par le marquage à la GFP et les auteurs ont eux même indiqué que cette technique est peu sensible et peut induire des artefacts dans la translocation des protéines.

Alors qu'il est connu que des différences dans les protéomes mitochondriaux issus de différents tissus existent, des travaux récents montrent néanmoins que des sous-populations de mitochondries avec des protéomes différents peuvent exister au sein d'un même tissu [25]. Très peu d'informations existent sur les protéomes de ces sous-populations de mitochondries.

Chez les mammifères, toutes les protéines mitochondriales à l'exception de treize (codées par le génome mitochondrial) sont codées par le génome nucléaire. Elles sont traduites dans le cytosol et importées dans la mitochondrie. La plupart de ces protéines portent un peptide signal à leurs extrémités N-terminales appelé peptide de transit qui assure sa translocation correcte dans l'organite. Une minorité de protéines portent une séquence signal interne [26] et implique donc une série de translocases [27,28]. Deux voies d'importation principales existent (Figure 1) [29]. Dans la première voie, le peptide signal dirige la protéine à travers le complexe TOM (Translocase of the Outer Membrane) qui se situe dans la membrane externe de la mitochondrie, puis à travers le complexe TIM23 (Translocase of the Inner Membrane) et le moteur PAM (Presequence translocase-Associated Motor) qui se situent dans la membrane interne jusqu'à la matrice où la MPP (Mitochondrial Processing Peptidase) clive le peptide de transit. Pour la deuxième voie, les protéines avec une séquence signal interne sont reconnues par le récepteur Tom70 et transportées par le complexe TOM et par les chaperones Tim9 et Tim10 (elles se situent dans l'espace intermembranaire) vers le complexe TIM22. Le complexe TIM22 favorise l'insertion des protéines transporteurs dans la membrane interne.

Il a été récemment montré chez la levure, que la voie d'importation impliquant un clivage protéolytique et libérant une nouvelle extrémité N-terminale peut modifier la stabilité de la protéine mitochondriale maturée [22]. Par ailleurs, cette étude montre que des données expérimentales solides sont nécessaires pour renforcer nos connaissances sur ce procédé de clivage N-terminal. Il existe effectivement des algorithmes de prédiction pour prédire le site de clivage du peptide de transit [30,31], mais ces algorithmes s'appuient sur un ensemble relativement petit de protéines dont le site de clivage a été déterminé expérimentalement. Il est donc nécessaire de valider expérimentalement ces prédictions.



Premièrement, plusieurs lignées cellulaires humaines sont disponibles à différents stades de différenciations. Ceci permettra d'effectuer des études comparatives plus facilement et d'évaluer les variations des protéomes d'une lignée à une autre.

Deuxièmement, on remarque que les cellules sanguines sont sous représentées dans la littérature en ce qui concerne la protéomique mitochondriale des mammifères, ceci est probablement dû au fait que la plupart de ces études sont réalisées sur la souris ce qui limite la quantité de sang disponible.

Troisièmement, le sang est un biofluide non invasif.

Enfin, et pour les mêmes raisons citées ci-dessus, les cellules du sang sont des cellules de choix pour l'étude de maladies liées aux mitochondries de façon permettant de transférer la technique vers la recherche clinique au cas où des candidats biomarqueurs émergent de cette étude.

La caractérisation des extrémités N-terminales des protéines est le seul moyen d'avoir accès aux séquences exactes des peptides N-terminaux des protéines mitochondriales humaines. Même s'il est connu que l'abondance des protéines mitochondriales change d'un type de cellule à l'autre, aucune information n'existe sur les variations dans la maturation et les clivages protéolytiques que subissent ces protéines dans ces types de cellules. D'un autre côté, peu de données expérimentales existent sur les N-terminaux exactes dans la littérature. En effet, comme ça a été indiqué ci-dessus, les extrémités N-terminales exactes ont été déterminées expérimentalement seulement pour une partie mineure des protéines mitochondriales. Par conséquent, les algorithmes de prédiction utilisés [34] ne peuvent que s'appuyer sur un petit nombre de protéines pour éduquer les logiciels et par suite la prédiction du site de clivage exacte est souvent erronée [22]. L'élargissement du nombre de données expérimentales permettra donc d'améliorer ces algorithmes et leur précision.

## 5. Stratégie d'analyse

### 5.1. Préparation de l'échantillon

Des cellules hématopoïétiques humaines de la lignée U266B1 sont collectées par centrifugation et suspendues dans un tampon de lyse hypotonique (triéthanolamine-HCl 10 mM). Elles sont ensuite cassées à l'aide d'un homogénéiseur de type Dounce. Une première centrifugation à basse vitesse (1200 xg) est effectuée pour éliminer les cellules non cassées, les noyaux et les débris. Les mitochondries brutes sont ensuite collectées par centrifugation à 10000 xg et resuspendues dans une solution de lavage. Plusieurs cycles sont effectués de cette façon. Le culot final est suspendu dans un tampon HEPES (HEPES NaOH 10 mM, Saccharose 0,25 M et EDTA 1 mM). Ceci constitue notre préparation de mitochondries brutes. La concentration en protéine est estimée par la méthode de Bradford.

Pour la purification des mitochondries, 7 traitements ont été testés :

- 1) Digestion à la protéinase K en conditions hypotoniques ;
- 2) Digestion à la protéinase K en conditions isotoniques ;
- 3) Digestion à la protéinase K en condition isotonique suivie d'un traitement à la digitonine qui permet d'éliminer la membrane externe des mitochondries abimée par cette digestion ;
- 4) Digestion à la protéinase K en condition isotonique puis traitement au Triton X-165 pour éliminer la membrane externe des mitochondries ;
- 5) Digestion à la protéinase K en condition isotonique puis traitement au Triton X-305 pour éliminer la membrane externe des mitochondries ;
- 6) Digestion à la protéinase K puis centrifugation avec gradient de densité ;
- 7) Centrifugation avec un gradient de densité suivie de digestion à la protéinase K.

La protéinase K est une exoprotéase qui permet de dégrader les protéines contaminantes. Après chacun de ces traitements décrits ci-dessus, les préparations sont lavées avant d'être resuspendues dans le tampon HEPES.

Nous avons ensuite évalué la pureté de ces préparations en calculant les pourcentages des protéines mitochondriales identifiées dans chacune par une analyse « shotgun » LC-MS/MS. Pour ce faire, un aliquot de chaque préparation a été repris dans un tampon urée, puis, après digestion liquide à l'aide de la trypsine a été analysé par LC-MS/MS. Nous avons obtenu les pourcentages de protéines mitochondriales suivant et qui traduisent la pureté des préparations : 33 % pour la préparation brute puis respectivement 53 %, 58 %, 34 %, 61 %, 59 %, 73 % et 74%. Il s'agit ici d'une estimation de pureté minimum puisque nous nous sommes basés sur les protéines qui sont déjà annotées comme étant mitochondriales dans les banques de données SwissProt pour calculer ces pourcentages.

La préparation 7 s'est donc avérée la plus enrichie en mitochondrie mais elle présente un rendement très faible puisque 95 % des mitochondries sont perdues après ce traitement. Nous avons alors décidé d'analyser la préparation brute et la préparation 7. D'une part, parce que la préparation brute contient une quantité plus importante de protéines et par suite permettra d'identifier un nombre plus important d'extrémités N-terminales et d'autre part parce que la préparation 7 est la plus riche en protéines mitochondriales.

## 5.2. Dérivation des extrémités N-terminales et analyses LC-MS/MS

Pour identifier les protéines mitochondriales et caractériser les sites de clivage des peptides de transit, la stratégie N-TOP (N-terminal Oriented Proteomics) fut utilisée (voire Partie II). Après réduction à la tributylphosphine et alkylation des cystéines, les amines N-terminales des protéines sont dérivées avec le réactif TMPP-Ac-OSu. Une étape de gel SDS PAGE est mise en place pour éliminer l'excès de réactif et décomplexifier l'échantillon suivie d'une digestion enzymatique. L'extrait de peptides ainsi obtenu est ensuite analysé par LC-MS/MS (Figure 2).

La stratégie est conçue pour obtenir le maximum de protéines et de peptides N-terminaux identifiés. Ainsi deux digestions enzymatiques indépendantes sont utilisées en parallèle pour augmenter le nombre d'identifications et surtout le nombre de peptides N-terminaux. Les deux



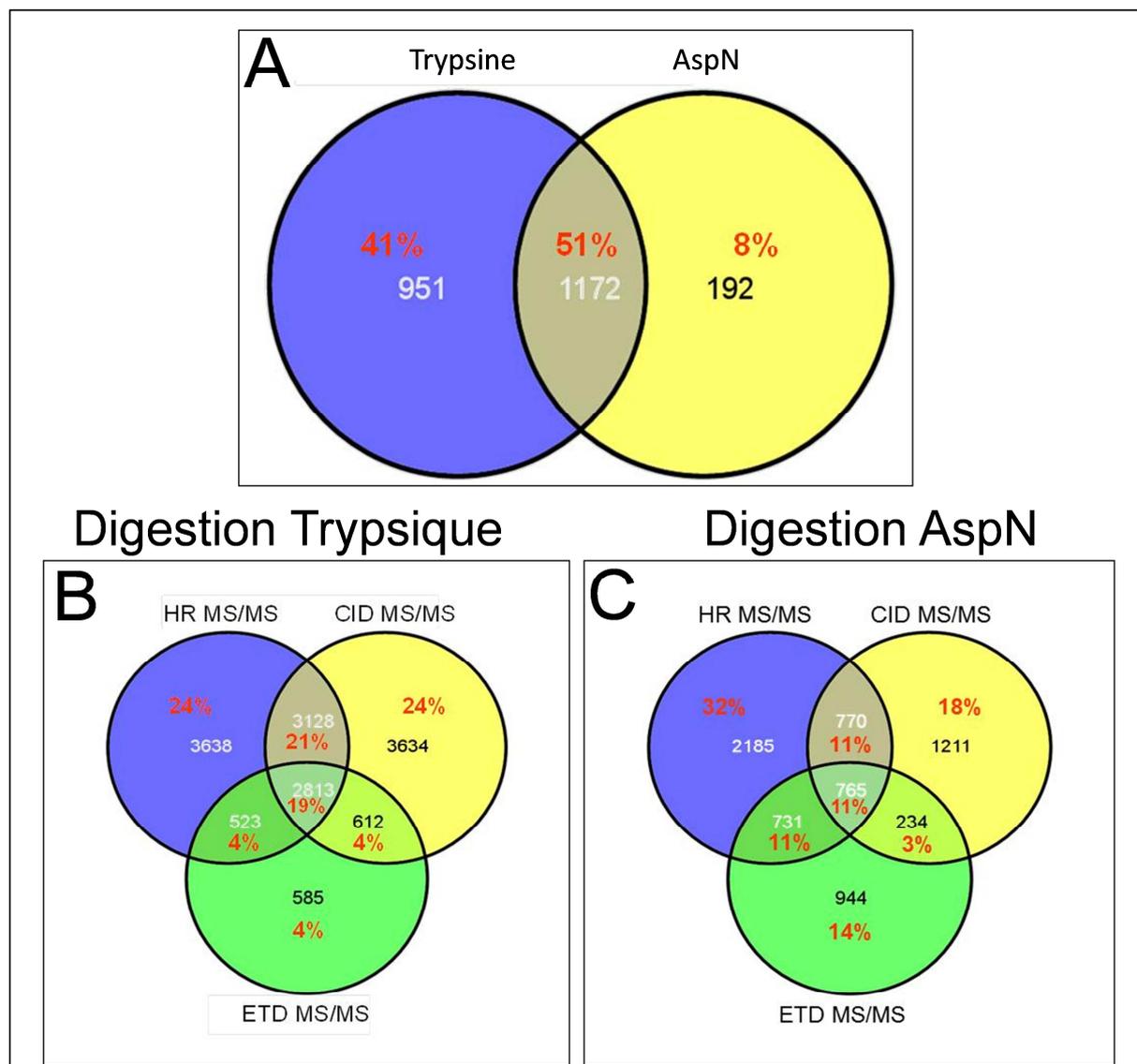
un deuxième temps, l'interrogation de la banque est relancée avec un paramétrage tenant compte de la modification des extrémités N-terminales des peptides et recherchant ceux dont seulement l'extrémité C-terminale correspond à un clivage spécifiques de l'enzyme utilisée (Enzyme semi-spécifique). Cette interrogation en deux temps permet de restreindre l'espace de recherche pour l'identification des protéines. En effet, la digestion *in silico* avec une enzyme semi-spécifique de toutes les protéines de la banque de données augmente considérablement le nombre de peptides théoriques issus de cette digestion. Cette augmentation, de plusieurs ordres de grandeur, augmente ainsi l'espace de recherche des données théoriques à comparer avec les données expérimentales. Ceci, comme expliqué dans le chapitre 2 de la partie bibliographique, augmente le risque de fausses découvertes. C'est pour ces raisons que nous avons décidé de procéder en deux temps, identifiant ainsi les protéines à l'aide d'une digestion *in silico* spécifique dans un premier temps. Puis, pour les peptides N-terminaux marqués au TMPP qui eux, n'ont qu'une seule extrémité spécifique d'un clivage enzymatique, nous avons paramétré les moteurs de recherche pour une recherche dans un deuxième temps avec clivage enzymatique semi-spécifique. Pour cette deuxième recherche, les moteurs de recherches ont aussi été paramétrés pour intégrer les fragments « a » dans la liste des fragments théoriques issus de la dissociation des peptides afin de tenir compte de la fragmentation particulière des peptides dérivés au TMPP. Ces fragments « a » constituent une indication précieuse lors de la validation manuelle de des identifications des peptides N-terminaux. Ils permettent de lever les ambiguïtés sur la véracité et la spécificité du marquage N-terminale des peptides contenant des lysines ou des tyrosines proche de leur extrémité N-terminales.

L'interprétation des résultats d'identification a été effectuée à l'aide du logiciel Scaffold (Proteome Software) qui permet de visualiser, trier, filtrer les résultats et appliquer la stratégie Target-Decoy et les approches empiriques de Bayes.

## 6. Résultats et discussion

Grâce à notre stratégie d'analyse, nous avons pu identifier plus de 2300 protéines et plusieurs centaines de sites de clivage protéolytique (325 sites uniques). Ces résultats sont illustrés dans la figure 2. Le taux de faux positifs (FDR) évalué pour ces identifications à l'aide de la stratégie « Target-Decoy » est inférieur à 1 %. On remarque sur cette figure que la digestion à la trypsine a permis à elle seule l'identification de 92 % des protéines laissant seulement 8 % de nouvelles identifications de protéines apportées par la digestion à l'AspN. 51 % des protéines sont identifiées communément avec les deux enzymes (Figure 3-A). Cependant l'utilisation de cette enzyme nous a permis d'augmenter le pourcentage de recouvrement de séquence des protéines identifiées d'environ 20 % en moyenne, ce qui, d'un côté renforce les identifications, et de l'autre côté permet de caractériser davantage les séquences des protéines. L'utilisation de deux instruments différents, le Q-TOF et la Trappe Ionique avec ses deux modes de fragmentation CID et ETD ont quant à eux permis une considérable augmentation du nombre d'identifications avec 24 % d'identifications apportées exclusivement par chacun des deux instruments (Q-TOF et trappe d'ions en mode CID) pour la trypsine et 32 % et 18 % respectivement après la digestion AspN. La fragmentation ETD n'a apporté que peu de nouvelles identifications de peptides tryptiques (4 %) et 14 % de peptides AspN (Figure 3-B et 3-C). Cependant, en comparant les résultats obtenus sur la trappe d'ions, 36 % des peptides sont identifiés exclusivement par fragmentation ETD, 43 % exclusivement avec le mode CID

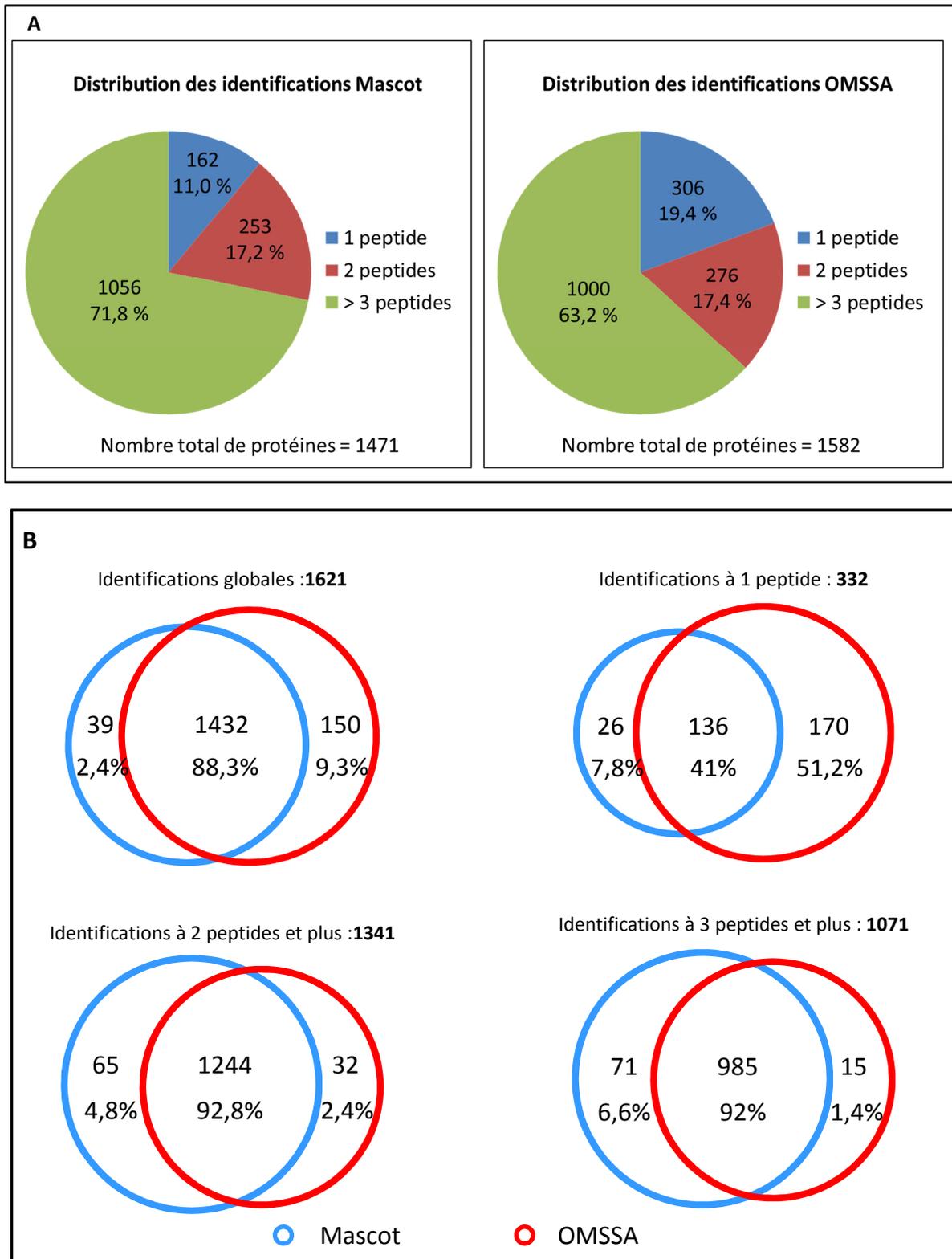
et 21 % sont des peptides identifiés en commun. Ces données illustrent la complémentarité des deux modes de fragmentation.



**Figure 3 :** Nombre total des protéines identifiées par des peptides tryptiques en comparaison avec des peptides AspN (A). Nombre total de peptides tryptiques (B) et AspN (C) identifiés par le système nanoLC-Q-TOF-MS/MS (haute résolution) en comparaison avec le système à trappe d'ions (basse résolution) en mode CID et en mode ETD.

L'utilisation des deux moteurs de recherche a été motivée par les nouvelles guidelines des journaux de protéomique pour les larges ensembles de données. Ces journaux commencent à exiger que les algorithmes d'analyses des données protéomiques soient disponibles au public, non seulement pour que tout le monde puisse reproduire les analyses, mais aussi pour que l'algorithme puisse être indépendamment vérifié [35]. L'utilisation de moteurs de recherche multiples est aussi encouragée. OMSSA est un logiciel libre dont l'algorithme est publié [36] ; il s'inscrit donc dans cette catégorie de moteurs de recherche. De plus, l'utilisation de deux moteurs de recherche différents qui utilisent des règles de « scoring » différentes permet une meilleure confiance dans les résultats, surtout pour les identifications à un seul peptide. En effet, le risque qu'une identification d'une protéine soit un faux positif augmente considérablement lorsque celle-ci n'est identifiée que par un

seul peptide [37,38]. La figure 4 montre une comparaison des résultats obtenus par les deux moteurs de recherche pour les analyses effectuées sur le Q-TOF après digestion trypsique.



**Figure 3 : Comparaisons des résultats d'identification par les deux moteurs de recherche Mascot et OMSSA. A) Distribution du nombre de peptides par protéine identifiée pour chacun des moteurs de recherche. B) Diagrammes de Venn montrant le recouvrement dans les résultats d'identification entre les deux moteurs de recherche pour la totalité des protéines, les protéines identifiée à 1, 2 ou 3 peptides et plus.**

Ces résultats (Figure 3) montrent tout d'abord que la majorité des protéines sont identifiées par les deux moteurs de recherche (88,3 %). On remarque aussi que OMSSA identifie légèrement plus de protéines (~ 7 %) mais la majorité de ces nouvelles identifications relève de protéines identifiées avec un seul peptide. D'un autre côté, Mascot identifie un peu plus de protéines avec 3 peptides et plus. On note que pour les identifications à 1 seul peptide, seulement 41 % des protéines identifiées l'ont été par les deux moteurs. Ceci permet donc de renforcer ces identifications par rapport à celles obtenues avec un seul moteur de recherche. Ces résultats montrent par ailleurs la complémentarité des deux moteurs de recherche permettant de présenter des résultats validés par deux algorithmes d'interrogation et de « scoring » différents. La liste des protéines identifiées par cette stratégie au cours de ce travail est en cours d'intégration dans la banque de données neXtProt.

Notre stratégie de dérivation N-TOP nous a permis de caractériser sans ambiguïté 325 peptides de transit uniques de protéines mitochondriales. Parmi ces peptides N-terminaux dérivés au TMPP, nous avons obtenu les quatre cas de figure illustrés par les exemples suivants :

- Le peptide identifié a permis la validation d'un site de clivage déjà observé et annoté dans la littérature (exemple de la protéine « Cytochrome c oxidase subunit 7C » dont le peptide transit a été observé par Van Kuilenburg et coll. [39] et un autre exemple dans la figure 6) ;
- Le peptide identifié a permis la validation expérimentale d'un site de clivage prédit in silico (exemple dans Figure 4) ;
- Le peptide identifié a permis de corriger une erreur dans le site de clivage prédit (exemple dans Figure 5) ;
- Le peptide identifié a permis de détecter un nouveau site de clivage de peptide transit non prédit (comme par exemple la protéine Sulfide:quinone oxidoreductase (Q9Y6N5) qui n'avait pas de peptide signal prédit).

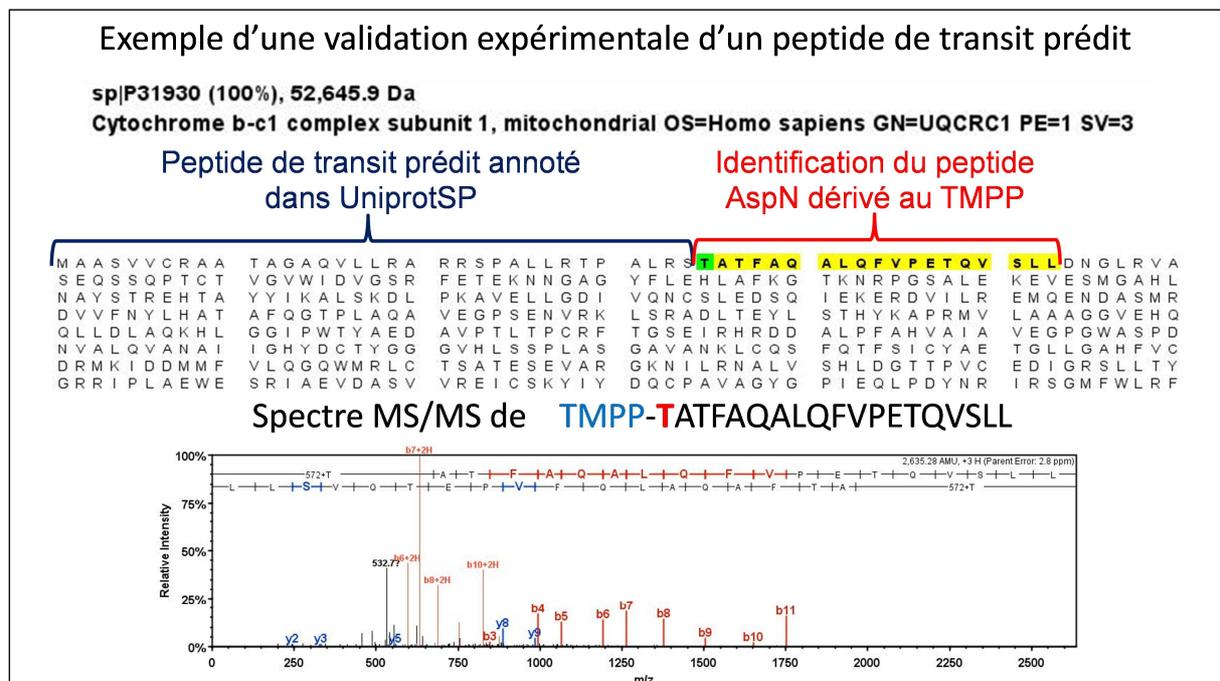
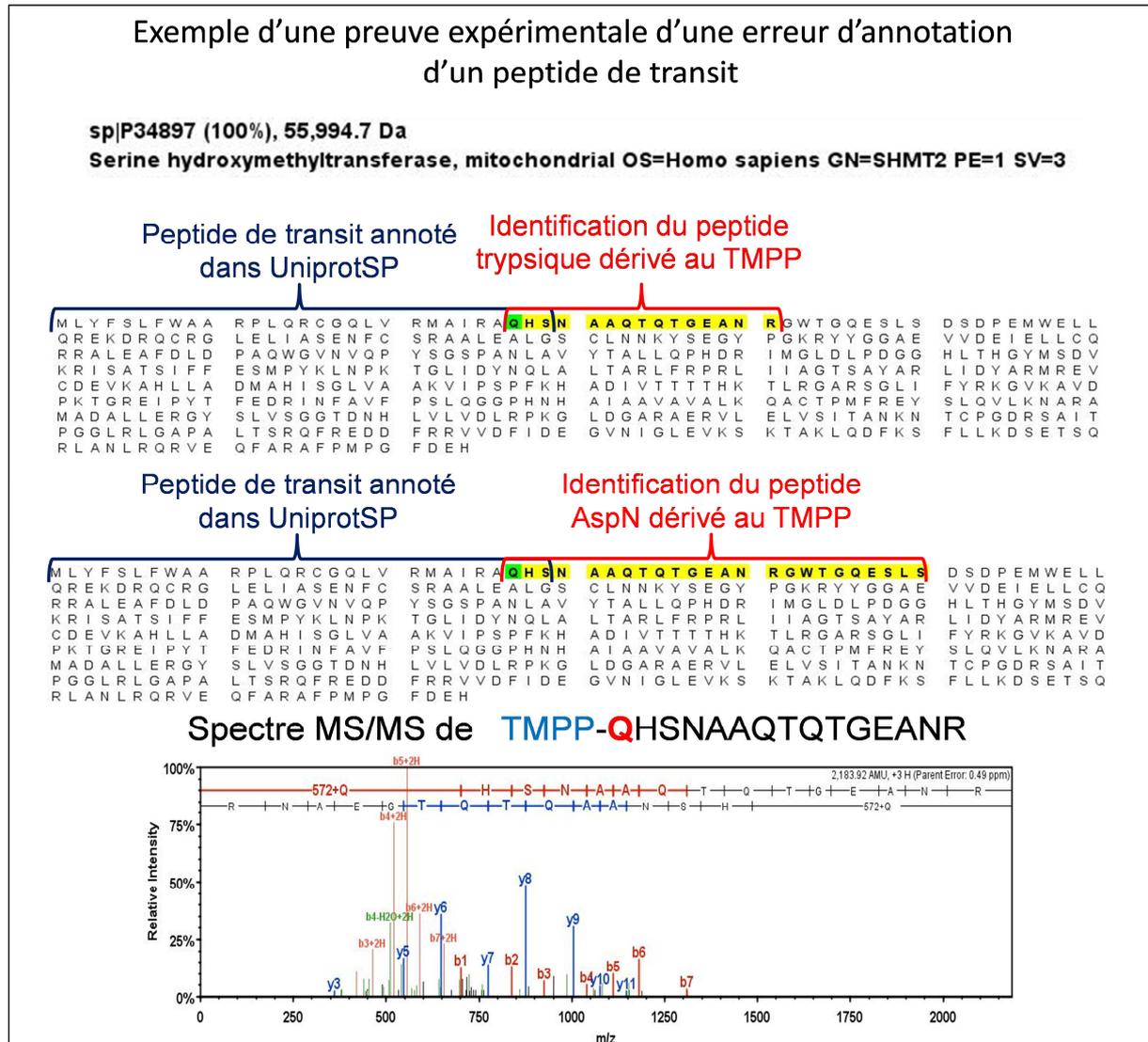


Figure 4 : Exemple de validations expérimentales d'un peptide de transit prédit.



**Figure 5 : Exemple de correction d'une erreur de prédiction d'un peptide de transit.**

Ces données expérimentales présentent un complément d'informations précieux pour l'annotation des protéines mitochondriales dans les banques de données du protéome humain et les séquences des peptides de transit sont en cours d'implémentation dans neXtProt (Figure 6). Nous avons aussi soumis nos données expérimentales sous forme de fichiers Scaffold (Proteome software) à TRANCHE (<https://proteomecommons.org/tranche/>). TRANCHE est un répertoire pour le dépôt de données protéomiques brutes. Les fichiers Scaffold contiennent notamment les spectres et les scores Mascot et OMSSA associés à chaque identification.



**CLPP » Putative ATP-dependent Clp protease proteolytic subunit, mitochondrial**

Displayed isoform: Iso 1

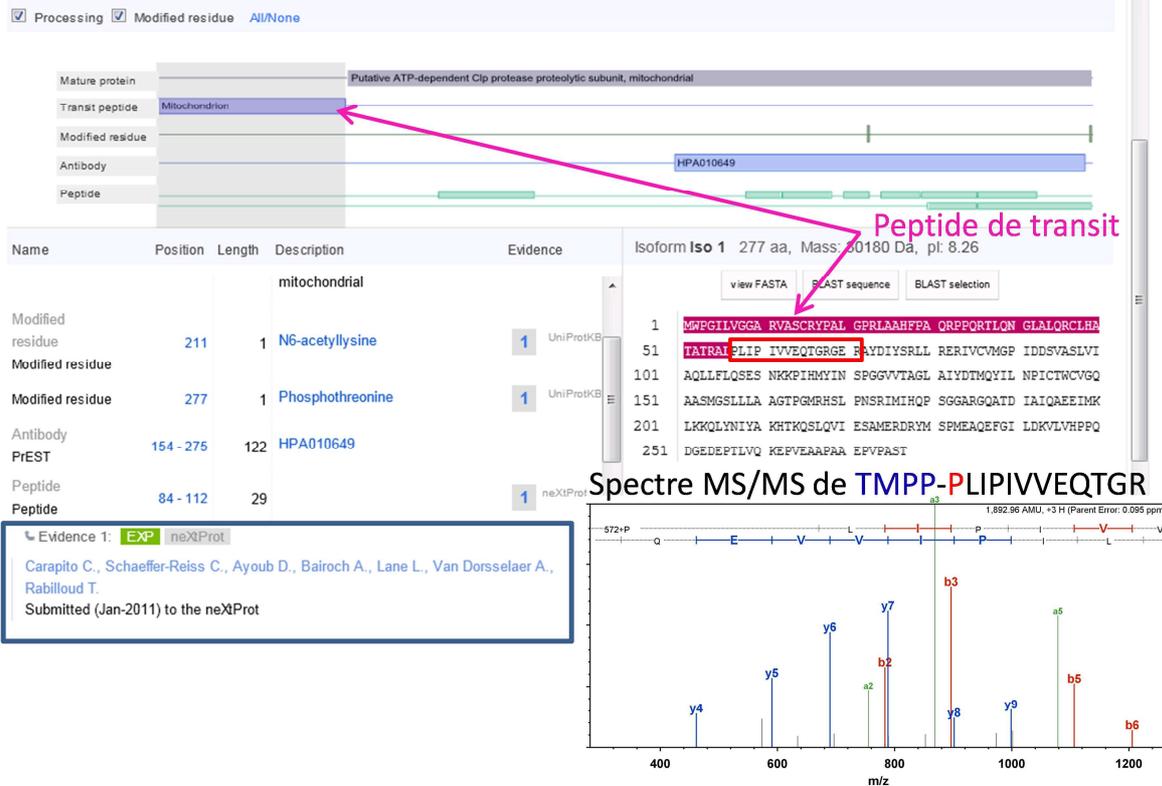


Figure 6 : Exemple de l'implémentation des données expérimentales de notre étude dans la banque de données neXtProt ; capture d'écran des informations concernant la protéine mitochondriale « CLPP » dans neXtProt. Confirmation du peptide de transit à l'aide du spectre MS/MS.

## 7. Conclusion et perspectives

Notre étude a permis jusqu'à présent l'identification de plusieurs centaines de protéines mitochondriales. Elle a aussi permis la correction et la validation expérimentales de plusieurs centaines de sites de clivage de peptides de transit. Ces résultats sont en cours d'implémentation dans la banque de données du protéome humain neXtProt. Notre stratégie de dérivation au TMPP a non seulement permis d'identifier avec confiance les peptides N-terminaux, mais a également permis l'identification de plus de 16000 peptides internes permettant l'identification de plus de 2300 protéines. Ceci est dû à l'avantage que présente cette technique de dérivation au TMPP qui induit un déplacement des temps de rétention des peptides N-terminaux vers la région hydrophobe du gradient sans perdre les peptides internes qui sont indispensables pour l'identification des protéines et l'augmentation des taux de recouvrement des séquences.

D'autre part, ces résultats nous ouvrent des perspectives pour compléter cette étude toujours dans la quête de données de haute qualité dans le cadre du Projet du Protéome Humain HPP et ceci suivant deux axes. Le premier s'articule autour de la réalisation d'analyses de protéomique comparative du « N-terminome » mitochondrial et ceci à l'aide de notre stratégie qN-TOP (Résultats, Partie I) pour étudier les variations dans le cas de pathologies liées aux

mitochondries. Le deuxième axe s'articule autour d'une approche ciblée de quantification, la SRM, pour effectuer des quantifications absolues des protéines mitochondriales d'intérêt. Ceci est possible grâce à des peptides de synthèse marqués aux isotopes stables et permettra de cibler les peptides N-terminaux ou les peptides protéotypiques des protéines d'intérêt. Cette approche pourra ouvrir la voie à la réalisation d'études quantitatives à haut débit de plusieurs centaines de protéines et permettra de les suivre de manière différentielle à travers différents types de cellules. Ces résultats ont été présentés dans une communication orale au congrès HuPO 2011 à Genève.

## Chapitre 2 Etude du protéome du parasite *Plasmodium falciparum* et de l'export des protéines parasitaires à l'érythrocyte infecté. Application de la stratégie N-TOP et Ingel N-TOP

*Collaboration avec le Département de Parasitologie de l'Université de Heidelberg (Prof. Dr. M. Lanzer et coll.)*

### 1. Résumé

*Plasmodium falciparum* est le parasite responsable de la forme la plus sévère du paludisme. Ce parasite possède un cycle de vie assez complexe. Durant son stade érythrocytaire asexué, il vit à l'intérieur de l'érythrocyte, qui est une cellule anucléée dépourvue d'un appareil sécrétoire. Pour assurer sa survie dans cet environnement hostile, il installe des fonctions vitales en exportant des protéines à travers sa membrane cytoplasmique et la vacuole parasitophore jusqu'à la cellule hôte [40,41]. Ces protéines exportées possèdent un motif dont la séquence RxLxE/Q/D est conservée dans plus de 300 protéines du parasite [42]. Ce motif est appelé « Plasmodium Protein EXport ELEMENT » (PEXEL) [43]. Il a été récemment montré qu'il contient un site de clivage protéolytique et d'acétylation N-terminale [44,45].

Pour apporter une meilleure connaissance des mécanismes intervenant dans l'export des protéines, ainsi que les mécanismes intervenants dans la survie de *P. falciparum*, nous avons entrepris trois études protéomiques de ce parasite.

Nous avons étudié un modèle représentatif de ces protéines à motif PEXEL de la famille STEVOR, un antigène de surface qui est exporté à l'érythrocyte infecté [46]. Des mutations ménagées de chaque acide aminé du motif PEXEL ont été réalisées afin de déterminer quels sont les acides aminés essentiels à la maturation et à l'export de cette protéine. Notre stratégie de marquage N-terminal réalisée dans le gel a permis de caractériser les peptides N-terminaux issus de ces expériences de mutation et ainsi de mettre en évidence non seulement le site de clivage mais également la présence d'une N-acétylation du nouveau N-terminal dans le cas de la protéine sauvage et de deux mutants dont l'un est exporté correctement. Ces expériences montrent que le clivage de la protéine a lieu sur l'alanine 51 et est suivi d'une acétylation de cette alanine. Ces deux

modifications post-traductionnelles sont donc nécessaires pour son export. Elle a aussi permis de caractériser la séquence des autres mutants et de mettre en évidence une erreur dans la séquence de la protéine.

La deuxième étude a consisté en une analyse approfondie du protéome du parasite dans son stade érythrocytaire en ciblant les séquences N-terminales. Notre stratégie de marquage par le TMPP permet d'« enrichir » en peptides N-terminaux tout en conservant les peptides internes et les peptides N-acétylés. Dans cette étude nous avons utilisé 3 enzymes de digestion différentes pour augmenter le nombre d'identifications. Nos résultats ont conduit à l'identification d'environ 1000 protéines et 550 peptides N-terminaux uniques issus ou non d'une maturation par clivage protéolytique. L'ensemble de ces résultats nous a permis d'établir une cartographie du protéome parasitaire et des clivages protéolytiques associés.

La troisième étude porte sur une protéine membranaire de la vacuole digestive du parasite : La Chloroquine Résistance Transporter (PfCRT). Cette protéine est impliquée dans la résistance du parasite à un antibiotique, La chloroquine. La phosphorylation est une modification post-traductionnelle qui régule la translocation des protéines entre divers compartiments cellulaires, les interactions protéine/protéine et l'activation ou l'inactivation de certaines fonctions cellulaires. Elle est souvent une clé régulatrice cruciale pour la signalisation cellulaire [47]. Nous avons effectué une étude de caractérisation des éventuels sites de phosphorylation de cette protéine qui seraient responsables de sa translocation à la membrane de la vacuole digestive. En utilisant plusieurs enzymes de digestion nous avons pu mettre en évidence trois sites de phosphorylation sur S<sub>33</sub>, S<sub>411</sub> et T<sub>416</sub>, et que lorsque T<sub>416</sub> est phosphorylée, S<sub>411</sub> ne l'est pas et vice versa. Or les résultats biochimiques prouvent que T<sub>416</sub> est un acide aminé nécessaire à cette translocation. Il est donc possible que cette protéine passe par une étape de phosphorylation transitoire lors de sa translocation.

## 2. Le paludisme en bref

Le paludisme est une parasitose infectieuse à laquelle, selon les estimations de l'Organisation Mondiale de la Santé (OMS), 1,2 milliards de personnes sont exposés avec un risque élevé d'infection (2,1 milliards encourtent un risque faible). Les estimations font état de 216 millions épisodes de paludisme en 2010 [48]. En 2010, le paludisme était responsable de la mort d'environ 1,2 millions de personnes [49]. La majorité des cas touchent les enfants de moins de 5 ans [50]. Cette maladie constitue un énorme fardeau pour les pays endémiques tant sur le plan économique que social.

Les parasites responsables du paludisme sont du genre *Plasmodium* dont le vecteur est le moustique femelle du genre *Anophèle*. Vingt-quatre espèces de *Plasmodium* sont répertoriées mais seulement cinq peuvent causer la maladie chez l'Homme : *Plasmodium falciparum*, *Plasmodium vivax*, *Plasmodium malariae*, *Plasmodium ovale* et enfin *Plasmodium knowlesi*. Le paludisme est endémique dans plus de 90 pays situés en majorité dans une bande large de part et d'autre de l'équateur [51,52]. La maladie sévissait en Europe depuis l'antiquité [53] et n'a été éradiquée qu'au milieu du siècle dernier grâce au pesticide DDT (DichloroDiphénylTricholoéthane) et aux conditions sanitaires. Les cas de paludisme constatés en Europe sont exclusivement des cas importés. Cependant, le vecteur, la femelle du moustique anophèle, est présent dans presque toutes les régions de l'Europe ; on parle d'« anophélisme non palustre » [54,55]. Avec le réchauffement climatique et le mouvement des populations, certains scientifiques expriment une crainte quant au

retour du paludisme en Europe [56,57], mais ceci est sujet à une grande discordance dans la communauté scientifique [58-60].

Le Paludisme affecte l'espèce humaine depuis plus de 50 000 années. Le protozoaire *Plasmodium* a été pathogène pour l'Homme tout au long de l'histoire de l'espèce [61]. Le cycle du parasite contient une étape dans l'organisme de l'anophèle qui est son hôte primaire et une étape humaine (dans le cas du paludisme humain) qui comporte deux phases : une phase hépatique et une phase érythrocytaire. La figure 1 illustre le cycle de vie de *Plasmodium*.

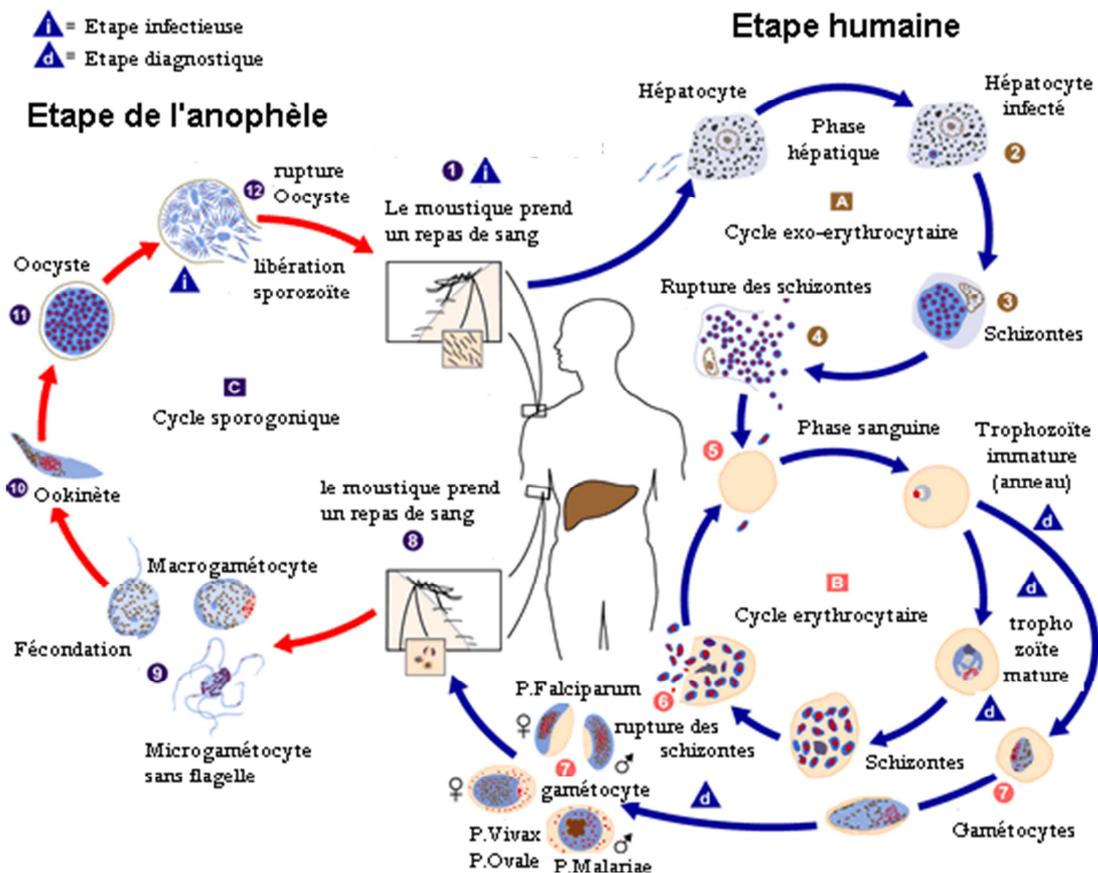


Figure 1 : Cycle de vie du parasite *Plasmodium* (d'après le Centre de Contrôle et de Prévention des Maladies, États-Unis d'Amérique [62]).

Avant l'infection, le parasite est au stade sporozoïte et réside dans la salive du moustique femelle infectée. Il est transmis à l'Homme lorsque celle-ci prend un repas de sang. Les sporozoïtes infectent vite les hépatocytes et après une brève période se différencient et se multiplient pour former des milliers de mérozoïtes. Les hépatocytes sont ensuite lysés libérant les mérozoïtes dans le sang qui vont infecter les érythrocytes. A l'intérieur des érythrocytes, les mérozoïtes se différencient en trophozoïtes et recommencent des cycles de multiplications asexuées formant un schizont contenant plusieurs mérozoïtes érythrocytaires. Ces schizonts finissent par éclater, lysant l'érythrocyte et libérant les mérozoïtes qui iront infecter d'autres érythrocytes. Certains trophozoïtes n'entament pas un cycle schizogonique mais un cycle gamogonique se différenciant ainsi en gamétocytes mâles ou femelles. Ces gamétocytes restent à l'intérieur de l'érythrocyte jusqu'au moment où un moustique femelle prend un repas de sang. Ainsi ingérés par l'insecte, ils entament

un cycle de multiplication sexuée dans son estomac pour former des sporozoïtes qui migreront jusqu'aux glandes salivaires en attendant d'infecter à nouveau un être humain.

L'éclatement des érythrocytes à la fin d'un cycle asexué est synchrone (toutes les 48h) causant des poussées de fièvre périodiques tout au long de la progression de la pathologie. Les symptômes vont de maux de tête à l'hypoglycémie, l'anémie, l'insuffisance rénale, le paludisme cérébral et la mort. La forte morbidité et mortalité du paludisme sont associées avec le stade intra-érythrocytaire [63]. Durant son développement asexué à l'intérieur de l'érythrocyte humain, *P. falciparum* modifie drastiquement sa cellule hôte [64,65].

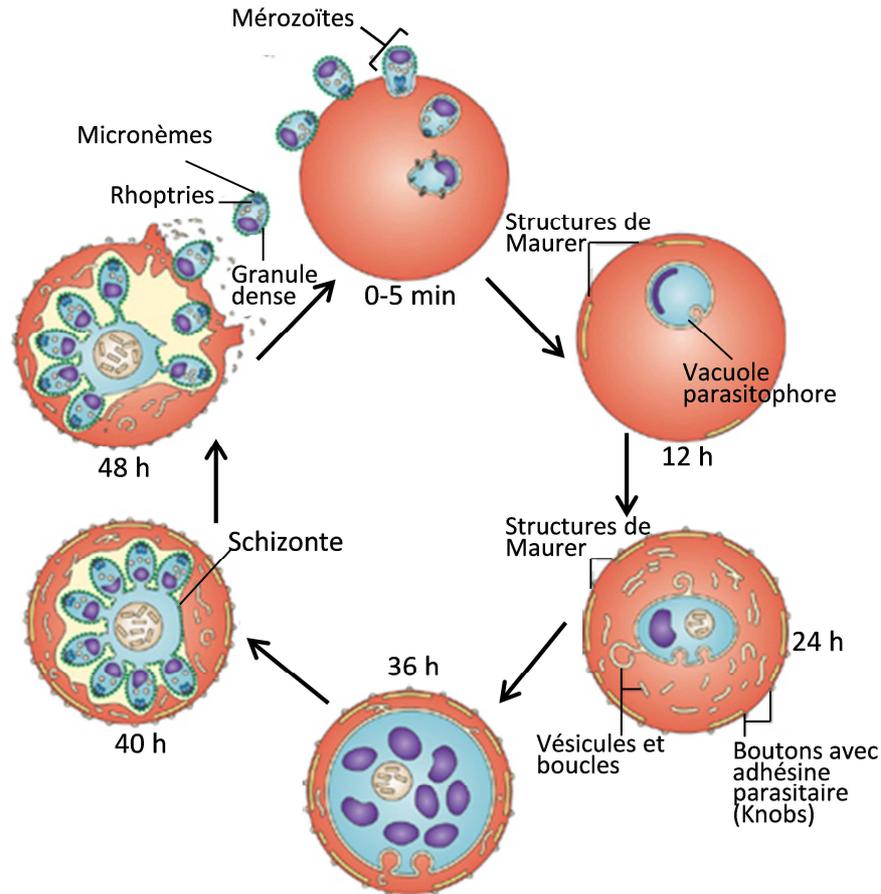
Au cours de cette thèse, nous avons entrepris trois études protéomiques concernant ce parasite durant son stade érythrocytaire. La première consiste à étudier le mécanisme d'export des protéines parasitaires à l'érythrocyte en se servant de protéines modèles et de notre stratégie « Ingel N-TOP ». La deuxième étude porte sur une caractérisation approfondie du protéome de *P. falciparum* durant son stade érythrocytaire en ciblant particulièrement les séquences N-terminales des protéines à l'aide de la stratégie N-TOP. Et enfin, la troisième étude s'intéresse au mécanisme de translocation d'une protéine de la vacuole digestive du parasite responsable de la résistance au principal antibiotique antiplaudique « la chloroquine ». Dans ce chapitre, nous présenterons les principaux résultats obtenus au cours de ces trois études.

### **3. Etude du mécanisme d'export des protéines à motif PEXEL du parasite *Plasmodium falciparum* à l'érythrocyte infecté. Application de la stratégie Ingel N-TOP**

#### **3.1. Contexte de l'étude**

Le parasite protozoaire *Plasmodium falciparum* a accompli au cours de son évolution une tâche d'adaptation remarquable en réussissant à vivre et à se développer au sein de l'érythrocyte humain. En faisant ce choix évolutif, le parasite s'est rendu vulnérable à l'élimination par la rate [66]. En effet, les érythrocytes ont été décrits comme étant des sacs d'hémoglobine adaptés par l'évolution au transport du dioxygène et du dioxyde de carbone [64]. De ce fait, ce sont des cellules très spécialisées, anucléées, dépourvues d'appareil sécrétoire et ne possédant qu'un répertoire limité de transporteurs d'ions et de métabolites. Le parasite se prive ainsi d'une source de nutriments extérieurs nécessaire à son développement et à sa reproduction [67]. Pour survivre dans cet environnement hostile, le parasite y installe des fonctions et des appareils vitaux en exportant plus de 300 de ses protéines dans la cellule hôte [42]. Certaines de ces protéines exportées créent de nouvelles voies et canaux de perméation pour les métabolites ; d'autres affectent les propriétés biomécaniques des cellules hôtes, tandis qu'une troisième classe de ces protéines exportées est constituée de variants antigéniques qui modulent la cytoadhésivité des érythrocytes infectés dans le réseau vasculaire, ou contribuent par d'autres moyens à l'évasion du système immunitaire et à la chronicité de l'infection.

La figure 2 résume le cycle érythrocytaire du moment de l'invasion à l'hémolyse et la rupture des schizontes.



**Figure 2 : Développement de *P. falciparum* dans un érythrocyte humain. Les mérozoïtes s'attachent à la membrane de l'érythrocyte avant de l'envahir (0-5min) pour se développer dans la vacuole parasitophore à travers le stade de l'anneau (24h). Le parasite passe ensuite au stade trophozoïte (24-36h) puis schizonte (40-48h). Le schizonte et l'érythrocyte sont ensuite éclatés pour libérer 16 à 32 mérozoïtes fils qui iront infecter de nouveaux érythrocytes. Au cours des stades matures du parasite (24 heures après l'invasion), des structures attachées à la membrane apparaissent dans le cytoplasme de l'érythrocyte et des déformations en forme de boutons se forment sur la membrane de ce dernier. Il s'agit respectivement des structures (aussi appelées vésicules) de Maurer (Maurer's Cleft en Anglais) et des « knobs ». Figure adaptée de [64] et de [68].**

Lors de l'invasion, une vacuole parasitophore est formée par invagination d'une partie de la membrane de l'érythrocyte dépourvue de protéines membranaires [69,70] qui sont préalablement dégradées par des enzymes parasitaires [71,72] sécrétées par des organites spécifiques des apicomplexa appelées rhoptries [73], micronèmes [74] et mononèmes [75]. Après formation de cette vacuole qui constitue le siège de développement du parasite, on voit apparaître des structures en forme de boutons sur la membrane érythrocytaire. Ces structures appelés « Knobs » agissent comme une plateforme pour la présentation à la surface de la membrane d'une adhésine parasitaire, la PfEMP1 (« *Plasmodium falciparum* Erythrocyte Membrane Protein 1 »), qui est responsable de la séquestration de l'érythrocyte par son adhésion aux cellules endothéliales des vaisseaux sanguins, échappant ainsi à l'élimination par la rate [64]. Cette cytoadhérence est aussi responsable des complications fatales tel que le paludisme cérébral (ou neuropaludisme) et le paludisme placentaire [76,77]. D'autres nouveaux organites cellulaires commencent aussi à apparaître. Parmi ceux-ci, des extensions et des spires émanent de la membrane de la vacuole parasitophore pour former un réseau tubulovasculaire [76,78]. D'autres structures, avec une composition protéique particulière, sont connues sous le nom de vésicules ou structures de Maurer (Maurer's Cleft). Ces vésicules de

Maurer ont pour rôle de concentrer les protéines virulentes pour les livrer à la membrane de l'érythrocyte hôte [40,67,79,80].

L'export des protéines vers l'érythrocyte est un mécanisme sophistiqué et assez spécifique de cette espèce. Il présente des aspects propres à ce parasite qu'on ne trouve dans aucun des organismes modèles utilisés dans l'étude de l'export et de la translocation des protéines. En effet, chez les autres cellules eucaryotes, les protéines destinées à la sécrétion sont généralement dirigées au réticulum endoplasmique (RE) rugueux à partir duquel elles sont transportées dans des vésicules, souvent via l'appareil de Golgi, à la surface de la cellule puis libérer [81,82]. Les protéines de *P. falciparum* destinées à l'export à la cellule hôte passe aussi par cette voie sécrétoire classique, un processus régi par l'intermédiaire de séquences signal (N-terminales ou internes) [40,83,84]. Cependant, cette voie permet au parasite de livrer ses protéines au travers sa membrane cytoplasmique et à la lumière de la vacuole parasitophore mais pas au-delà [40,41,85]. L'export et la translocation outre cette vacuole dans laquelle vit le parasite nécessitent des moyens supplémentaires tel que des séquences signal et des outils capables de reconnaître ces séquences signal que le parasite doit installer *de novo* dans la cellule hôte et au-delà de ces propres frontières que constitue la vacuole parasitophore.

Les vésicules de Maurer citées ci-dessus, sont des structures essentielles dans cet appareil sécrétoire extracellulaire que le parasite installe dans sa cellule hôte. On pense que ces vésicules bourgeonnent puis se détachent de la membrane de la vacuole parasitophore et s'étendent jusqu'à la membrane cytoplasmique de l'érythrocyte. Elles semblent être ancrées dans le cytosquelette de l'hôte. Des travaux suggèrent que les protéines transmembranaires et probablement des protéines solubles destinées à l'export seraient associées aux vésicules de Maurer d'une manière transitoire avant d'être envoyées à leur destination finale. Des expériences réalisées avec des parasites mutants déficitaires en protéines constituant des vésicules de Maurer montrent des altérations dans l'export et la translocation des protéines à la membrane cytoplasmique de l'érythrocyte [67,80].

La majorité des protéines exportées, contiennent une séquence signal conservée HCT (Host Cell Targeting signal) qui est impliquée dans la translocation à travers la membrane de la vacuole parasitophore. Marti et coll. ont suggéré que HCT est un motif pentamérique avec RXLXE/Q/D comme séquence consensus [43]. Ce motif est appelé motif PEXEL/HT (Plasmodium Protein EXport ELement/Host Targetting) ou tout simplement motif PEXEL. Hiller et coll. ont quant à eux proposé une version plus étendue de 11 acides aminés avec RXRSRILAEXXX comme séquence consensus appelée VTS (Vacuolar Transport Signal) [86]. Si la mutagenèse dirigée a confirmé l'importance des résidus R, L et E/D/Q au sein du motif [43,86,87], la taille minimale de la séquence signal HCT est encore discutée. De plus, des études protéomiques ont identifié plusieurs protéines exportées avec des séquences HCT putatives qui divergent de la séquence consensus [88]. Ceci conduit à se poser la question sur l'étendue de la conservation de ce motif PEXEL. C'est dans ce contexte que nous avons, en collaboration avec l'équipe du Prof. Michael Lanzer de l'université de Heidelberg, effectué une étude dont l'objectif est de mieux définir le signal HCT et son rôle dans l'export des protéines.

### 3.2. Etat des connaissances concernant le motif PEXEL

Pour modifier sa cellule hôte, le parasite *P. falciparum* exporte plus de 5 % de son protéome vers cette cellule [89]. Or ce parasite se développe et vit dans une vacuole parasitophore à l'intérieur même de l'érythrocyte se séparant ainsi du cytosol de sa cellule hôte par deux barrières

membranaires, sa propre membrane cytoplasmique et la membrane de la vacuole parasitophore. La membrane de la vacuole parasitophore (MVP) constitue une barrière qui protège le parasite des protéines de son hôte. Les interactions entre le parasite et son hôte passent obligatoirement à travers cette membrane [90] et les protéines exportées par le parasite doivent donc traverser cette membrane. Le motif d'export conservé PEXEL [43] a permis de prédire plus de 300 protéines comme appartenant à l'exportome [86] ou le sécrétome [91] du parasite. Une étude récente a montré, par des expériences de « knockout » des gènes codant pour des protéines à motif PEXEL, qu'environ 25 % des produits de ces gènes sont des protéines essentielles pour le développement *in vitro* et une partie d'entre elles joue un rôle probable dans l'absorption des nutriments et l'échange de solutés [65]. Cependant, le mécanisme par lequel ces protéines sont exportées reste peu connu et suscite beaucoup d'interrogations et d'intérêt puisqu'il est totalement propre au genre *Plasmodium* et il peut révéler des cibles thérapeutiques contre le paludisme [92].

Les études de ce mécanisme d'export ont montré que le motif PEXEL est clivé par une protéase non caractérisée et est probablement acétylé à posteriori [45]. Ce clivage se produit vraisemblablement dans le réticulum endoplasmique avant l'export [44,45]. Boddey et coll. ont effectué une étude de modèle de protéines contenant un motif PEXEL en mutant les trois acides aminés conservés de ce motif [45]. Les résultats montrent que les résidus R et L du motif sont indispensables pour l'export des protéines tandis que lorsque le résidu Q/E/D est muté, la protéine est partiellement exportée. De plus, ils ont suggéré que l'acétylation des protéines clivées est probablement un processus commun et qu'elle n'est pas suffisante ni indispensable pour qu'une protéine soit correctement exportée. Les auteurs indiquent cependant que des lacunes persistent sur la compréhension de ce mécanisme d'export. Leurs conclusions concernant le rôle joué par l'acétylation dans l'export sont mitigées bien qu'ils aient démontré qu'un clivage après la leucine est systématique pour les protéines exportées.

Un des modèles suggérés pour l'export repose sur trois signaux (en ce qui concerne les protéines membranaires) situés dans les séquences des protéines exportées. Un peptide signal N-terminal nécessaire pour l'entrée dans la voie sécrétoire du réticulum endoplasmique (RE), une séquence HCT pour le passage à travers la membrane parasitophore, et enfin, pour les protéines membranaires, un domaine transmembranaire pour leur translocation à la membrane des vésicules de Maurer. Ceci est suggéré par les données de Przyborski et coll. grâce à des expériences de mutations dirigées, de délétions ménagées de séquences et de traitement à la bréfeldine A [87]. La bréfeldine A est connue pour inhiber la sécrétion à travers le réticulum endoplasmique et le traitement des parasites transgéniques avec la bréfeldine A a conduit à l'ablation de l'export et à la rétention des protéines marquées avec la GFP (« Green Fluorescent protein ») dans le réticulum endoplasmique. La délétion de la séquence HCT conduit à l'export de la protéine vers la vacuole parasitophore mais pas au-delà, suggérant un peptide signal permettant l'export à travers la membrane plasmique du parasite comme dans une voie sécrétoire RE/Golgi classique. La mutation du résidu R du motif PEXEL a conduit à la rétention des protéines chimères dans le cytosol du parasite. La mutation du résidu Q/E/D a conduit à la rétention de la chimère dans la vacuole parasitophore. Ce dernier phénotype est aussi observé par d'autres équipes [43,86] mais aucune explication n'est proposée.

Lingelbach et Przyborski ont revu les différents mécanismes suggérés pour l'export des protéines [93]. Un des mécanismes proposés suggère la présence d'un complexe protéique dans la

membrane de la vacuole parasitophore qui sert de translocon putative pour l'export des protéines à travers cette membrane. L'équipe de Crabb a repris cette hypothèse et a proposé le modèle illustré par la figure 3. Ce modèle repose sur la présence de ce translocon appelé PTEX (« Plasmodium Translocon of EXported proteins ») [92]. Ce complexe protéique PTEX a été partiellement caractérisé par Bullen et coll. [94] mais sa fonction de translocon reste putative.

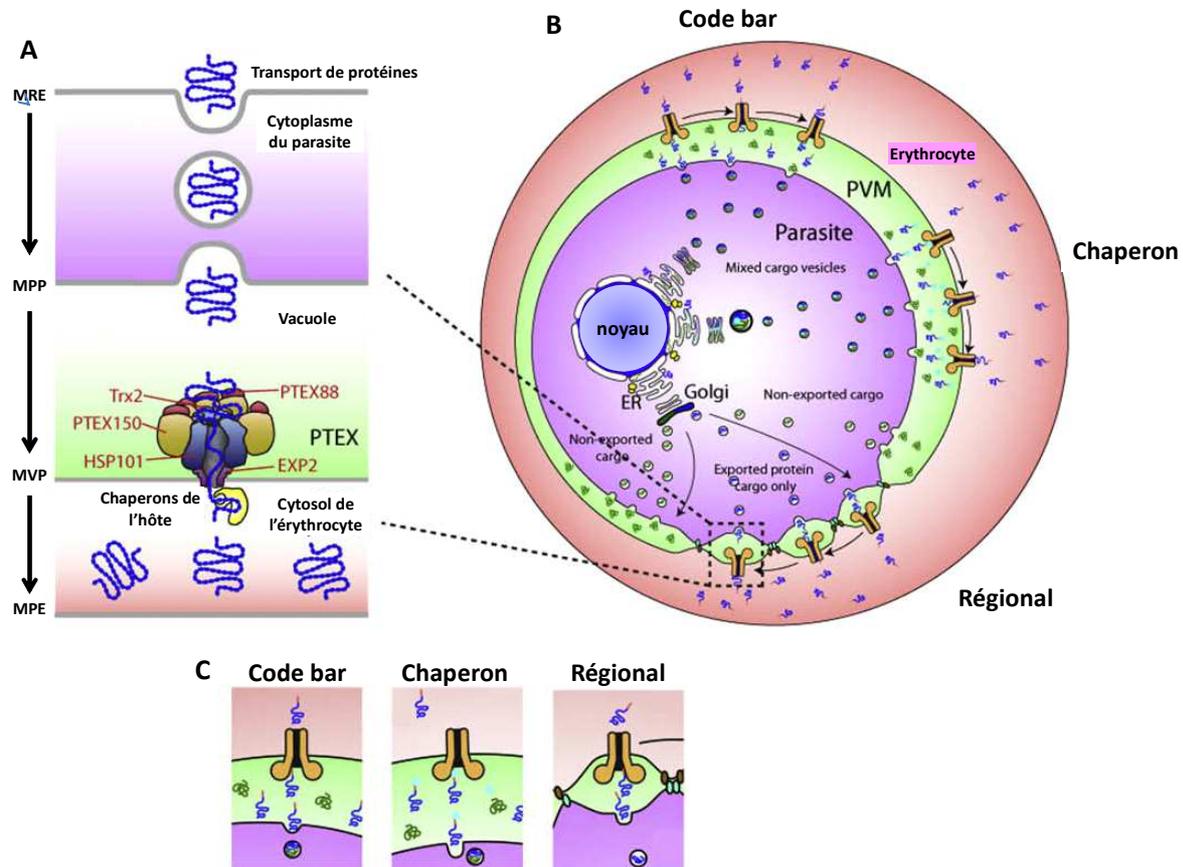


Figure 3 : Schéma de modèles proposés pour l'export de protéines à motif PEXEL dans les érythrocytes infectés par *P. falciparum*. A) Les protéines parasitaires à motif PEXEL sont maturées dans le réticulum endoplasmique (RE) et migrent ensuite à la membrane plasmique dans des vésicules. Une fois dans la vacuole parasitophore, ces protéines sont probablement dépliées et extrudées à travers un pore traversant la membrane de la vacuole parasitophore (MVP) vers le cytosol de l'érythrocyte via le translocon du complexe PTEX. Ce schéma montre la disposition putative des 5 protéines connues du complexe PTEX. Du côté érythrocytaire de la MVP, l'export est probablement effectué par les protéines chaperons de l'érythrocyte. B et C) Trois modèles sont proposés pour expliquer les stades intermédiaires de l'export des protéines du RE vers le PTEX ; ces modèles sont : code bar, chaperon et régional. Dans les deux premiers modèles, les protéines à motifs PEXEL (bleue) quittent le RE/Golgi dans des vésicules accompagnées de protéines non exportées (vertes) et sont déposées dans la vacuole. Dans le modèle code bar, seulement quelques acides aminés N-terminaux sont nécessaires pour la reconnaissance par le PTEX et la translocation vers le cytosol de l'érythrocyte. Dans le modèle chaperon, la reconnaissance par le PTEX s'effectue grâce à une protéine chaperon (cyan) attachée à la protéine à motif PEXEL. Dans le modèle régional, les protéines à motif PEXEL sont triées dans le RE/Golgi et emballées dans des vésicules spécialisées qui transportent et libèrent ces protéines dans des compartiments spécialisés du complexe PTEX qui les exportent vers le cytosol de l'érythrocyte. MRE = Membrane du RE, MPP = Membrane Plasmique du Parasite, MPE = Membrane Plasmique de l'Erythrocyte. Adaptée de [92].

### 3.3. Stratégie d'analyse : Application de la stratégie Ingel N-TOP

Afin d'étudier le rôle de chaque acide aminé du motif PEXEL, des parasites *P. falciparum* transgéniques ont été préparés. Ces parasites expriment une protéine chimérique contenant le signal HCT. Il s'agit des 83 premiers acides aminés de la protéine STEVOR fusionnés à l'aide d'un linker de 4 acides aminés du côté C-terminal avec un marqueur fluorescent, la GFP (« Green Fluorescent Protein »). Le motif PEXEL contenu dans ces construits est <sup>48</sup>RLLAQ<sup>52</sup> (Figure 4-A). Ensuite plusieurs mutations dans ce motif sont effectuées en remplaçant chaque acide aminé par une alanine, et l'alanine en position 51 par un acide aspartique. Nous avons ainsi les mutants R48A, L49A, L50A, A51D et Q52A (Figure 4-B). Dans la suite de ce chapitre, le construit chimérique avec un motif PEXEL non muté sera désigné par « WT ». Des études de microscopie confocale à balayage laser ont été effectuées par les parasitologues afin de déterminer la localisation cellulaire de chacun des mutants.

A) Séquence de la protéine chimérique sans mutation dans le motif PEXEL						B) Mutations du PEXEL	
-2	MKMYNLKMLL	FNFLINVLVL	PHYENPRYQN	NHYNIRLIPN	NTYRITIKSR	48	52
51	LLAQ	TQIHNP	HYHNDPELKE	IIDKLNEDAI	KKARGTMSKG	EELFTGVVPI	WT
101	LVELDGDVNG	HKFSVSGEGE	GDAYGKLTLL	KFICTTGKLP	VPWPTLVTFE		R48A
151	AYGLQCFARY	PDHMKQHDFE	KSAMPEGYVQ	ERTIFFKDDG	NYKTRAEVKF		L49A
201	EGDTLVNRIE	LKGIDFKEDG	NILGHKLEYN	YNSHNVYIMA	DKQKNGIKVN		L50A
251	FKIRHNIEDG	SVQLADHYQQ	NTPIGDGPVL	LPDNHYLSTQ	SALSKDPNEK		A51D
301	RDHMLLEFV	TAAGITHGMD	ELYK				Q52A

Figure 4 : A) séquence en acides aminés du construit chimérique avec un motif PEXEL (rouge) non muté et désigné par WT. La numérotation commence à partir de la méthionine désignée par une flèche rouge puisque les 2 premiers acides aminés (MK) ne sont pas traduits comme il sera montré plus tard dans cette étude. Les acides aminés en vert appartiennent à la séquence de la GFP et ceux en orange au linker. B) Les différentes mutations du motif PEXEL

Les différents parasites mutants ont ensuite été lysés et les protéines chimériques ont été immunoprécipitées à l'aide d'un anticorps anti-GFP. L'éluion est effectuée à l'aide d'un tampon de charge commercial pour électrophorèse sur gel SDS PAGE. Les protéines sont ensuite déposées et migrées sur un gel monodimensionnel. La coloration au bleu montre que plusieurs protéines ont co-précipité avec les chimères. Les bandes contenant les protéines chimériques sont identifiées à l'aide d'un Western Blot. La bande d'intérêt avait une couleur très peu intense indiquant sa faible abondance. Ces bandes ont été excisées pour les analyser à l'aide de la stratégie Ingel N-TOP présentée dans le chapitre 1 de la partie 2. Après dérivation chimique des amines N-terminales par le TMPP-Ac-OSu dans le gel et digestion enzymatique, les peptides générés sont analysés par LC-MS/MS sur un appareil de haute résolution (50000) de type Q-TOF (MaXis, Bruker Daltonics) couplé à un système de chromatographie nanoUHPLC (NanoAcquity, Waters). Le spectromètre de masse est équipé d'un système « lockmass » qui permet la correction de la dérive de l'étalonnage de l'instrument à tout instant.

Les résultats des analyses LC-MS/MS sont ensuite traités et une recherche dans une banque de données est effectuée à l'aide du moteur de recherche Mascot. La banque utilisée est une banque composite « Target-Decoy » contenant les séquences NCBI des protéines de *P. falciparum* ainsi que les séquences des chimères étudiées et les séquences des protéines contaminantes courantes (trypsine, kératines). Les paramètres de recherche sont les suivants :

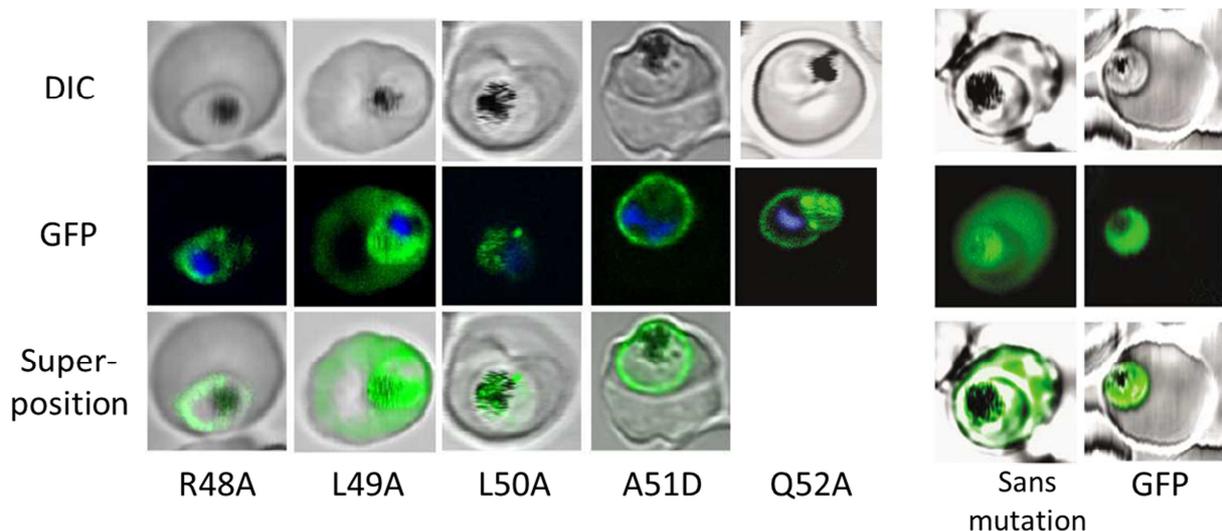
- Tolérance sur la masse de l'ion précurseur et sur la masse des fragments : 5 ppm ;

- Enzyme : semitrypsine (seule une extrémité du peptide doit correspondre à un site de clivage spécifique de la trypsine) ;
- Nombre de sites de clivage manqués autorisés : 1 ;
- Modifications fixes : carbamidométhylation des cystéines ;
- Modifications variables : Oxydation des méthionines, acétylation des amines N-terminales des peptides, TMPP sur les amines N-terminales des peptides.

Les spectres ayant permis une identification d'un peptide N-terminal (acétylé ou marqué au TMPP) sont inspectés manuellement.

### 3.4. Résultats et discussion

L'analyse de localisation cellulaire réalisée par les biologistes sur des parasites exprimant la protéine chimérique avec un motif PEXEL non muté montre un signal fluorescent dans le cytosol de l'érythrocyte. Ceci indique que la protéine chimère a été exportée à travers la membrane cytoplasmique du parasite et la membrane de la vacuole parasitophore. Un contrôle exprimant seulement la GFP montre un signal fluorescent confiné dans le cytosol du parasite. Quant aux mutants, ceux-ci présentent des phénotypes différents comme le montre la figure 5. En effet, les mutants R48A et L50A présentent un signal fluorescent confiné dans le cytoplasme du parasite indiquant que les protéines chimères n'ont pas été exportées. Le mutant L49A présente une distribution de la fluorescence comparable au parasite exprimant le construit sans mutation montrant que la protéine est exportée vers le cytosol de la cellule hôte. Les mutants A51D et Q52A montrent quant à eux un signal fluorescent qui correspond à la présence de la protéine dans la vacuole parasitophore.



**Figure 5 : Localisation cellulaire des mutants des différents construits par microscopie confocale à fluorescence. DIC = contraste d'interférence différentielle.**

Ces résultats montrent d'une part que l'acide aminé L<sup>49</sup> ne joue pas un rôle important dans l'export tandis que les acides aminés R<sup>48</sup>, L<sup>50</sup>, A<sup>51</sup> et Q<sup>52</sup> sont indispensables pour le bon export de la protéine à travers les deux membranes. Or le motif PEXEL est défini d'après la littérature par la séquence RXLXE/Q/D. D'après cette séquence, la mutation de l'alanine 51 en acide aspartique ne devrait pas altérer l'export. De plus, une question se pose, pourquoi les protéines A51D et Q52A

franchissent la membrane cytoplasmique et sont exportées vers la vacuole parasitophore tandis que les protéines R48A et L50A restent confinées dans le cytosol du parasite ? La littérature suggère que les protéines exportées subissent un clivage protéolytique dans la séquence PEXEL après le résidu L<sup>50</sup> suivi d'une acétylation de la nouvelle extrémité N-terminale. Afin d'étudier et de caractériser ces clivages protéolytiques nous avons appliqués notre stratégie Ingel N-TOP.

L'application de la stratégie Ingel N-TOP permet de cibler les bandes contenant la protéine d'intérêt et d'effectuer le marquage des amines N-terminales dans le gel. Ces bandes d'intérêt identifiées par Western Blot avec un anticorps anti-GFP ont été soumises à une dérivation chimique au TMPP-Ac-OSu suivie d'une digestion enzymatique. Les peptides ainsi générés sont analysés par LC-MS/MS. Les résultats pour chaque construit sont détaillés ci-dessous :

- Construit WT : Identification du peptide N-terminal acétylé A<sup>51</sup>-K<sup>67</sup> (Figure 6) et de peptides internes confirmant la séquence de la protéine. Aucun peptide qui correspond à la séquence en amont de l'alanine 51 n'est identifié.
- Mutant L49A : Identification du peptide N-terminal acétylé A<sup>51</sup>-K<sup>67</sup> (Figure 6) et de peptides internes. Aucun peptide qui correspond à la séquence en amont de l'alanine 51 n'est identifié.

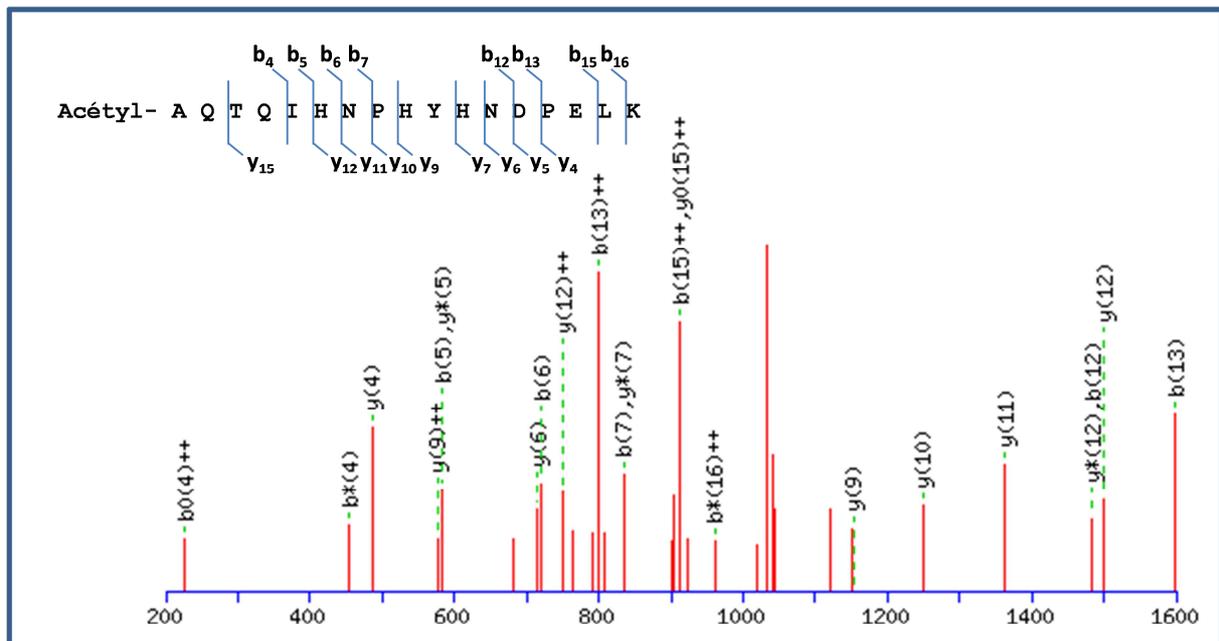


Figure 6 : Exemple d'un spectre MS/MS permettant l'identification du peptide N-terminal acétylé en A<sup>53</sup> comme dans le cas du WT et du mutant L49A.

- Mutant R48A : Le peptide N-terminal M<sup>1</sup>-K<sup>5</sup> est identifié marqué au TMPP (Figure 7). Le motif PEXEL est séquencé entièrement.
- Mutant L50A : Le peptide N-terminal M<sup>1</sup>-K<sup>5</sup> est identifié marqué au TMPP (Figure 7). Le motif PEXEL est entièrement séquencé. De plus deux autres peptides marqués au TMPP sont identifiés, le peptide V<sup>17</sup>-R<sup>27</sup> et le peptide V<sup>19</sup>-R<sup>27</sup>.

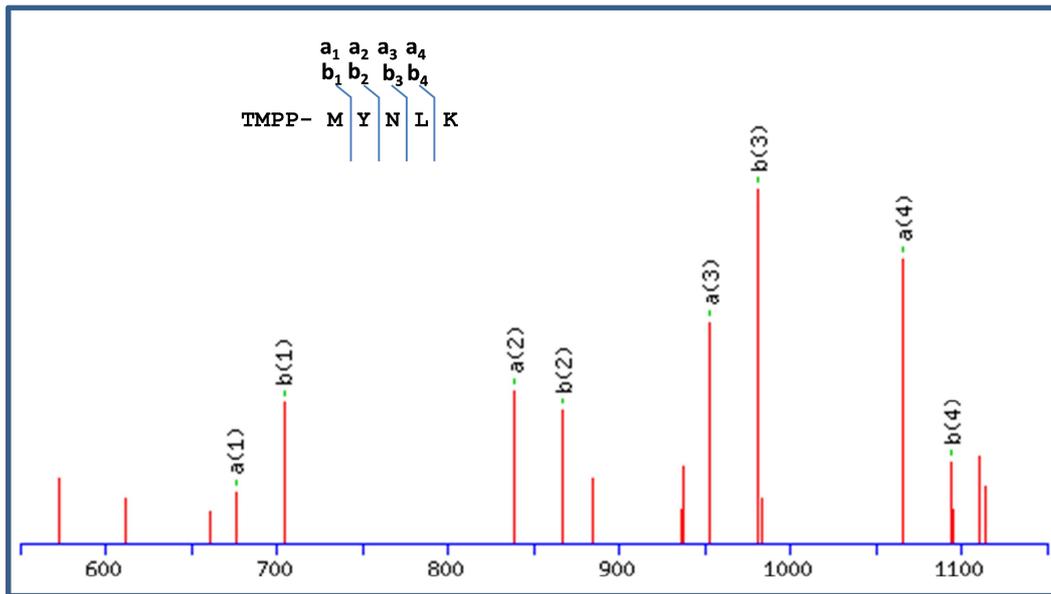


Figure 7 : Exemple d'un spectre MS/MS permettant l'identification du peptide N-terminal MYNLK comme dans le cas des mutants R48A et L50A. Il est à noter la présence sur le spectre des fragments  $a_1$  et  $b_1$  montre clairement que le marquage est spécifique sur l'amine N-terminale et non sur le résidu tyrosine.

- Mutant Q52A : Identification du peptide N-terminal acétylé  $A^{51}Q-K^{67}$  et de peptides internes (Figure 8). Aucun peptide qui correspond à la séquence en amont de l'alanine 51 n'est identifié.

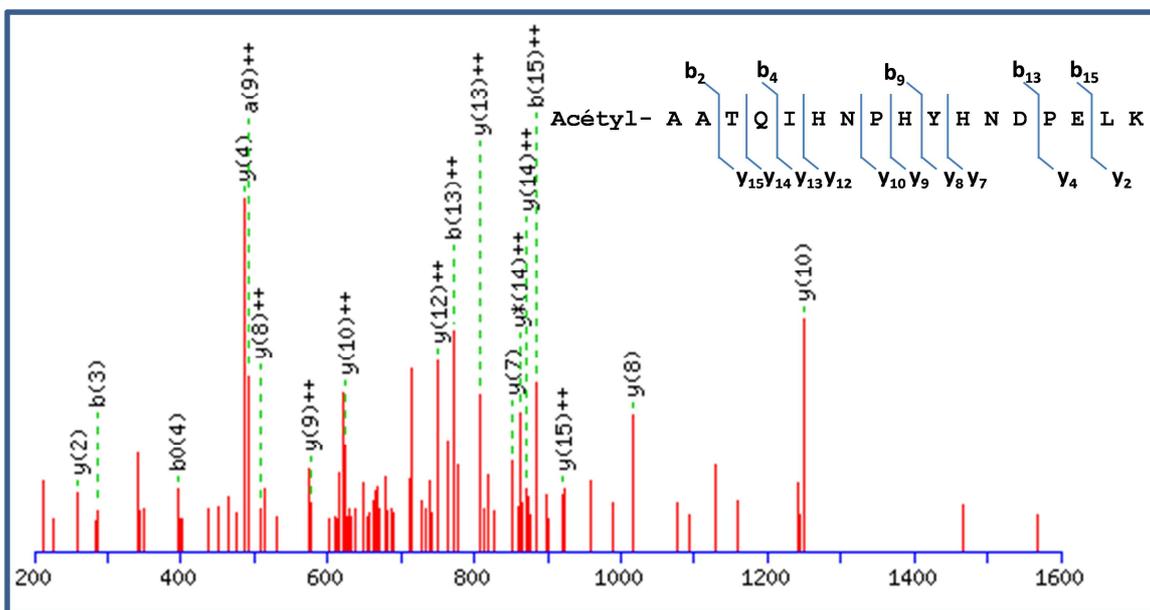


Figure 8 : Spectre MS/MS permettant l'identification du peptide N-terminal acétylé du mutant Q52A.

- Mutant A51D : Pour ce mutant, aucun peptide acétylé ou marqué au TMPP n'a été identifié dans les premières analyses. Cependant un spectre de bonne qualité a permis l'identification du peptide semitryptique  $D^{51}-K^{67}$ . Or seul un peptide acétylé ou marqué au TMPP permet d'identifier sans ambiguïté un peptide N-terminal. Si effectivement ce peptide semitryptique identifié correspond à un peptide N-terminal, il devrait donc être marqué au TMPP. Or ce peptide trichargé possède une masse moléculaire de 2085,95 Da. S'il est marqué au TMPP, il

aura une masse moléculaire de 2658,13 Da et il sera probablement 4 fois chargé. Les peptides 4 fois chargés génèrent généralement des spectres peu informatifs dans nos conditions expérimentales. En effet, en paramétrant le moteur de recherche Mascot de façon à rechercher les peptides 4 fois chargés, un spectre a permis l'identification du peptide D<sup>51</sup>-K<sup>67</sup> marqué au TMPP confirmant ainsi que c'est le peptide N-terminal (Figure 9). Cependant, ce spectre est peu informatif, il ne contient que peu de fragments peu intenses. Et donc, pour lever toute ambiguïté, un peptide de synthèse ayant la même séquence, dérivé au TMPP et contenant un acide aminé marqué aux isotopes stables a été commandé. Ce peptide a servi pour optimiser les transitions (fragmentation) afin de suivre notre peptide par SRM (« Single Reaction Monitoring ») à l'aide d'un analyseur de type triple quadripôle. La stratégie SRM permet de rechercher des peptides faiblement abondants dans un mélange à l'aide de critères de temps de rétention et de fragmentations spécifiques (transitions). L'application de cette stratégie SRM nous a permis de lever toute ambiguïté concernant l'identité du peptide identifié par notre stratégie Ingel-NTOP et confirmant que le spectre MS/MS de la figure 9 est le spectre de fragmentation du peptide N-terminal du mutant A51D.

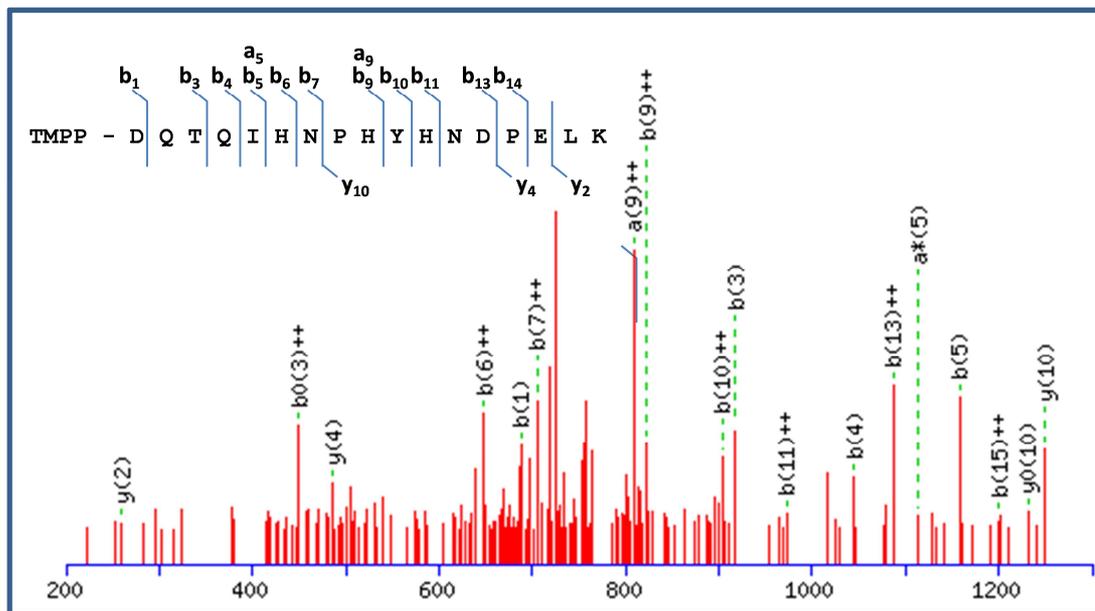


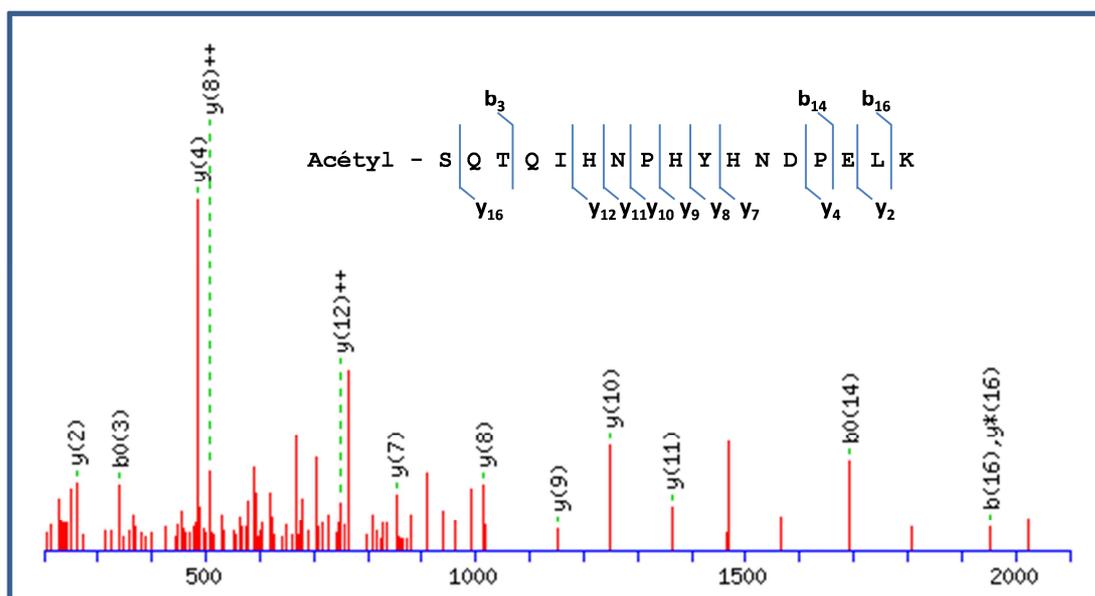
Figure 9 : Spectre du peptide N-terminal du mutant A51D marqué au TMPP. L'ion précurseur est 4 fois chargé.

En corrélant ces résultats avec ceux de la microscopie confocale, nous pouvons constater que le clivage après le résidu L50 (le résidu L conservé du motif PEXEL) et l'acétylation de l'acide aminé suivant sont nécessaires. Cependant on remarque d'une part que le résidu suivant cette leucine 50 est important et ne peut pas être substitué par n'importe quel acide aminé comme c'est stipulé dans la séquence consensus RXLXQ/E/D du motif PEXEL. Le fait d'avoir muté ce résidu en D a conduit à un phénotype particulier où le clivage a lieu après la leucine 50 mais avec une absence d'acétylation et un export altéré vers la vacuole parasitophore. D'autre part, la mutation du résidu Q<sup>52</sup> en alanine a aussi conduit à un phénotype différent par rapport au mutant R48A et L50A. Ce construit subit un clivage après la leucine 50 et une acétylation du résidu suivant (A<sup>51</sup>), comme dans le cas des construits exportés vers le cytosol de la cellule hôte, mais il reste confiné dans la vacuole parasitophore. Ces résultats sont illustrés dans la figure 10.

	Motif PEXEL				
Séquence de la protéine avant export	<sup>1</sup> MY-KS <sup>47</sup>	<sup>48</sup> RLLAQ <sup>52</sup>	<sup>53</sup> T-A <sup>83</sup>	Linker	GFP
Séquences exportées correctement dans l'érythrocyte: Cas du PEXEL sauvage et du mutant L49A	Acétyl-	<sup>51</sup> A-Q <sup>52</sup>	<sup>53</sup> T-A <sup>83</sup>	Linker	GFP
Séquence exportée vers la vacuole parasitophore: Cas du mutant A51D		<sup>51</sup> DQ <sup>52</sup>	<sup>53</sup> T-A <sup>83</sup>	Linker	GFP
Séquence exportée vers la vacuole parasitophore: Cas du mutant Q52A	Acétyl-	<sup>51</sup> A-A <sup>52</sup>	<sup>53</sup> T-A <sup>83</sup>	Linker	GFP
Séquences non exportées: Cas des mutants R48A et L50A	<sup>1</sup> MY-KS <sup>47</sup>	<sup>48</sup> R/A-L-L/A-AQ <sup>52</sup>	<sup>53</sup> T-A <sup>83</sup>	Linker	GFP

**Figure 10 : Caractérisation de l'extrémité N-terminale des différents construits à motif PEXEL.**

Ces résultats confirment tout d'abord que les résidus conservés du motif PEXEL (R, L et Q/E/D) sont nécessaires pour le bon export des protéines. Le fait que seuls les mutants R48A et L50A ne subissent pas un clivage protéolytique dans le motif PEXEL montre que ces résidus sont responsables de la reconnaissance du site de clivage par la protéase. Le résidu suivant la leucine 50 joue un rôle déterminant dans la reconnaissance de la N-acétyl-transférase (NAT) impliquée dans cette acétylation. Lorsque ce site est muté, l'acétylation n'a plus lieu et la protéine bien que clivée, est retenue dans la vacuole parasitophore. Ceci montre pour la première fois que l'acétylation des protéines à motif PEXEL maturés est nécessaire pour l'export à travers la membrane de la vacuole parasitophore (MVP). Comme cet acide aminé est non conservé, nous avons effectué des mutations de ce résidu vers tous les autres acides aminés dans deux protéines à motif PEXEL différentes (KAHRP et STEVOR). Nous avons trouvé que seul le remplacement de cette alanine par une valine ou une sérine permet l'export à travers la MVP. La mutation du résidu suivant n'empêchant pas l'acétylation, il semble que ce résidu est seul responsable de la reconnaissance par la NAT. Afin de vérifier si le clivage et l'acétylation ont lieu si l'alanine 51 est remplacé par une sérine, nous avons appliqué notre stratégie Ingel N-TOP sur le construit chimérique portant la mutation A51S. Effectivement, ce construit qui est exporté vers l'érythrocyte subit le même clivage que le construit non muté (WT) après le résidu L<sup>50</sup> et la sérine 51, qui est le nouveau résidu N-terminal, est acétylée. Ceci a pu être déterminé grâce à l'identification du peptide N-terminal du construit A51S (Figure 11).



**Figure 11 : Spectre MS/MS permettant l'identification du peptide N-terminal acétylé du construit A51S.**

Des mutations des autres résidus du motif PEXEL par les 20 acides aminés ont aussi été effectuées. Les résultats montrent que pour l'export vers la cellule hôte, le résidu L50 ne peut pas être remplacé tandis que le résidu R48 peut uniquement être remplacé par une lysine. La mutation du résidu Q52 vers un acide glutamique ou aspartique n'altère pas l'export comme stipulé par la séquence consensus du motif PEXEL. La retenue du mutant Q52A dans la vacuole parasitophore bien que mûri correctement et acétylé suggère que ce résidu avec le résidu N-terminal acétylé sont nécessaires pour la reconnaissance par le PTEX (ou la protéine chaperon selon les modèles suggérés) de la protéine afin qu'elle soit exportée à travers la MVP vers l'érythrocyte.

Concernant les extrémités N-terminales des mutants R48A et L50A, l'identification du peptide TMPP-<sup>1</sup>M-K<sup>5</sup> nous permet d'identifier l'extrémité N-terminale de la protéine avant export. En effet, les biologistes n'étaient pas sûrs de l'extrémité N-terminale de la séquence traduite. Cette identification permet donc de montrer expérimentalement que la séquence N-terminale commence par MYNLK- et non par MKMYNLM-.

### 3.5. Conclusion

Grâce à notre stratégie Ingel N-TOP, nous avons pu caractériser les extrémités N-terminales des différents mutants et ainsi mettre en évidence les clivages protéolytiques que subissent chacun d'eux. Ces résultats de caractérisation des séquences, combinés à la microscopie confocale, ont permis d'élucider le rôle de chaque résidu du motif PEXEL. Nous avons ainsi pu déterminer les acides aminés impliqués dans toutes les étapes de la reconnaissance moléculaire intervenant dans l'export, à savoir, le clivage du motif PEXEL par une protéase spécialisée, l'acétylation de la nouvelle extrémité N-terminale ainsi générée et la reconnaissance par le translocon pour l'export à travers la membrane de la vacuole parasitophore. Ces résultats feront l'objet d'une publication qui est en cours de préparation.

## 4. Etude du protéome N-terminal du parasite *P. falciparum* dans son stade érythrocytaire.

### 4.1. Contexte de l'étude

L'étude précédente a permis de montrer le rôle que peuvent jouer les protéases dans le cycle de vie du *P. falciparum*. Etant donné l'importance des protéases dans la régulation de différents processus biologiques tel que l'apoptose [95], la progression du cycle cellulaire [96], la migration cellulaire [97] etc, nous avons entrepris une étude du protéome N-terminal de ce parasite dans son stade érythrocytaire. L'objectif est d'étudier le protéome de chaque compartiment de l'érythrocyte infecté (parasite, vacuole parasitophore, érythrocyte) (Figure 12). Au cours de ce travail de thèse, nous avons accompli la première partie de cette étude qui consiste à identifier les protéines du parasite et à caractériser leurs extrémités N-terminales.

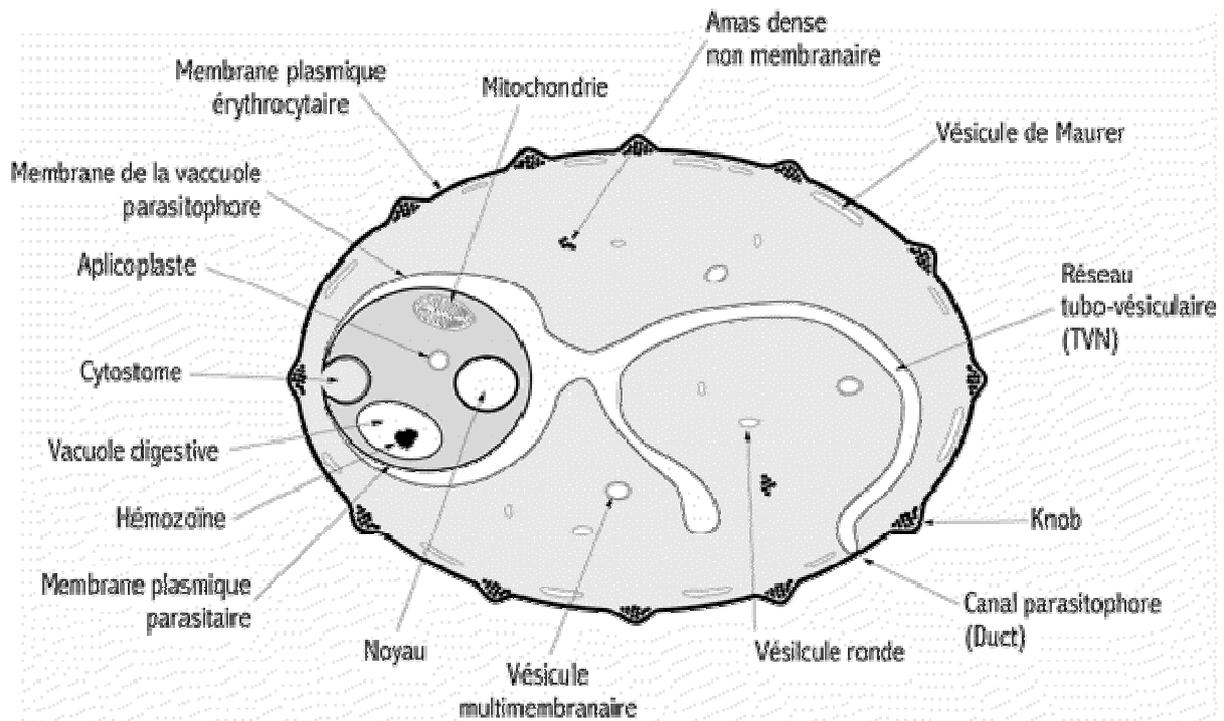


Figure 12 : Schéma d'un érythrocyte infecté par *P. falciparum* [98].

## 4.2. Stratégie analytique

### 4.2.1. Culture et purification des protéines de chaque compartiment

Des parasites *P. falciparum* de la lignée 3D7 sont cultivés dans des érythrocytes humains O+ comme décrit dans [99] dans un milieu RPMI supplémenté avec 5 % de sérum humain O+ et les nutriments nécessaires pour le développement du parasite. Les cultures sont synchronisées par lyse au sorbitol [100]. Au stade trophozoïte, les érythrocytes infectés sont purifiés sur une colonne MACS (« Magnetic-Activated Cell Sorting ») comme décrit dans [101]. Afin d'isoler les protéines de chaque compartiment, les membranes cytoplasmiques des érythrocytes sont dégradées à l'aide de la streptolysine O [102]. Les protéines du compartiment érythrocytaire sont ainsi collectées. Le culot constitué des vacuoles parasitophores contenant les parasites est lavé plusieurs fois avec un tampon PBS afin d'éliminer les débris érythrocytaires. La membrane de la vacuole parasitophore est lysée à l'aide de la saponine 0,07 % dans du PBS. Le surnageant contenant les protéines des vacuoles parasitophores est ainsi collecté. Enfin, après de nouveaux lavages des parasites avec un tampon PBS, ils sont lysés avec un tampon RIPA. A chaque étape, des inhibiteurs de protéases sont ajoutés.

Les protéines ainsi collectées de chaque compartiment sont précipitées avec du chloroforme/méthanol froid et conservé à -80°C.

### 4.2.2. Marquage N-TOP des protéines

Le précipitat des protéines du parasite est resolubilisé dans le tampon de la réaction N-TOP (Tris-HCl 50 mM, 6 M urée, 2 M thiourée, 1 % SDS pH 8,2). Les protéines sont dosées avec la méthode de Bradford puis les ponts disulfures sont réduits avec 10 mM de TCEP et les cystéines sont alkylées avec 30 mM d'iodoacétamide. 200 équivalents molaires de TMPP-Ac-OSu sont ensuite ajoutés pour marquer les amines N-terminales des protéines et le mélange est maintenu sous

agitation pendant 1 heure. La réaction est ensuite inhibée avec 100 mM d'hydroxylamine avant de séparer les protéines sur un gel SDS-PAGE. Cette étape de gel permet d'éliminer l'excès de réactif et de ses produits de dégradation et de décomplexifier l'échantillon. Le gel est découpé systématiquement en bandes de 2 mm. Ces bandes sont ensuite décolorées et déshydratées et les protéines sont digérées dans le gel. Afin d'augmenter le nombre de peptides N-terminaux et le nombre d'identifications de protéines, nous avons réalisé trois digestions enzymatiques parallèles à l'aide de trois enzymes différentes (Trypsine, Chymotrypsine et AspN).

#### 4.2.3. Analyse LC-MS/MS et identification des peptides et des protéines

Les peptides générés sont analysés à l'aide d'un système nanoLC-MS qui consiste en un spectromètre de masse hybride de type Q-TOF (MicroTofQ, Bruker) couplé à un système de chromatographie nanoUHPLC (NanoAquity, Waters). Le spectromètre de masse est équipé d'un système « lockmass » qui permet la correction de la dérive de l'étalonnage de l'instrument en continu.

Pour identifier la présence de protéines humaines contaminantes mais aussi des protéines recrutées par le parasite et importées dans son cytoplasme il faut inclure les séquences des protéines humaines dans la banque de séquences utilisée pour l'interrogation des spectres MS/MS. Il n'existe pas une banque de séquences protéiques de haute qualité pour le parasite *P. falciparum*. Seulement 326 séquences sont validées et intégrées dans la banque Swissprot. Nous avons donc décidé de prendre les séquences des protéines de *P. falciparum* de la banque Uniprot qui regroupe les séquences de la banque Swissprot et de la banque TrEMBL (20257 entrées) et de la concaténer avec les séquences humaines de la banque Swissprot (45374 entrées). Les séquences de la banque TrEMBL sont majoritairement de nature prédictive et ne sont pas validées expérimentalement. La banque obtenue est concaténée avec une banque contenant le même nombre d'entrée de séquences « Decoy » afin d'appliquer la stratégie Target-Decoy pour l'évaluation du taux de fausses découvertes.

La recherche dans cette banque de données est effectuée en deux temps. Dans un premier temps, la recherche est effectuée afin d'identifier les protéines présentes dans l'échantillon. Ainsi dans le paramétrage du moteur de recherche nous avons tenu compte seulement des peptides dont les deux extrémités correspondent à un site de clivage de l'enzyme utilisée et les modifications TMPP et acétylation n'ont été autorisées qu'en N-terminal des protéines. Ensuite, une fois les protéines identifiées, nous avons généré une sous-banque contenant les séquences de ces protéines. Dans un deuxième temps, c'est dans cette sous-banque que nous avons effectué notre recherche des peptides N-terminaux issus de clivages protéolytiques naturels en paramétrant le moteur de recherche de façon à identifier les peptides dont seulement une extrémité est spécifique au clivage par l'enzyme utilisée et en autorisant la modification TMPP et acétylation sur ces peptides. Ceci permet de réduire l'espace de recherche avec ce dernier paramétrage qui augmente considérablement le nombre de peptides de digestion *in silico* et donc de spectres *in silico* à comparer augmentant ainsi la probabilité qu'un spectre expérimental corresponde à un spectre théorique par hasard conduisant ainsi à un faux positif. Les spectres ayant conduit à l'identification d'un peptide N-terminal ont été ensuite inspectés manuellement.

### 4.3. Résultats et discussion

Dans cette analyse protéomique, nous avons identifié 957 protéines uniques dont 847 sont des protéines du parasite et 110 sont des protéines humaines. Nous avons aussi déterminé, grâce à notre stratégie N-TOP, 550 extrémités N-terminales uniques à partir de l'identification de 804 peptides N-terminaux. Le marquage au TMPP permet d'identifier les amines N-terminales libres mais n'empêche pas l'identification d'une amine N-terminale acétylée. Parmi ces extrémités N-terminales identifiées, on distingue :

- 230 protéines dont l'extrémité N-terminale est issue d'un clivage et dont l'amine N-terminale est libre ;
- 96 protéines dont la méthionine N-terminale est clivée et dont l'amine N-terminale est libre ;
- 52 protéines n'ayant pas subies de clivage et dont l'amine N-terminale est libre ;
- 35 protéines dont l'extrémité N-terminale est issue d'un clivage et acétylé ;
- 95 protéines dont la méthionine N-terminale est clivée et l'amine N-terminale est acétylée ;
- 42 protéines n'ayant pas subies de clivage et dont l'amine N-terminale est acétylée.

L'utilisation de plusieurs enzymes n'a pas permis d'augmenter de manière significative le nombre de protéines identifiées et la trypsine reste l'enzyme qui a permis d'obtenir le nombre le plus élevé d'identifications (figure 13-A). Cependant, l'utilisation de l'AspN et de la Chymotrypsine ont permis d'augmenter considérablement le nombre de nouvelles identifications d'extrémités N-terminales. Ainsi, 185 extrémités N-terminales sont identifiées exclusivement grâce à la digestion trypsique, 97 grâce à la digestion chymotrypsique et 82 grâce à la digestion AspN (Figure 13-B)

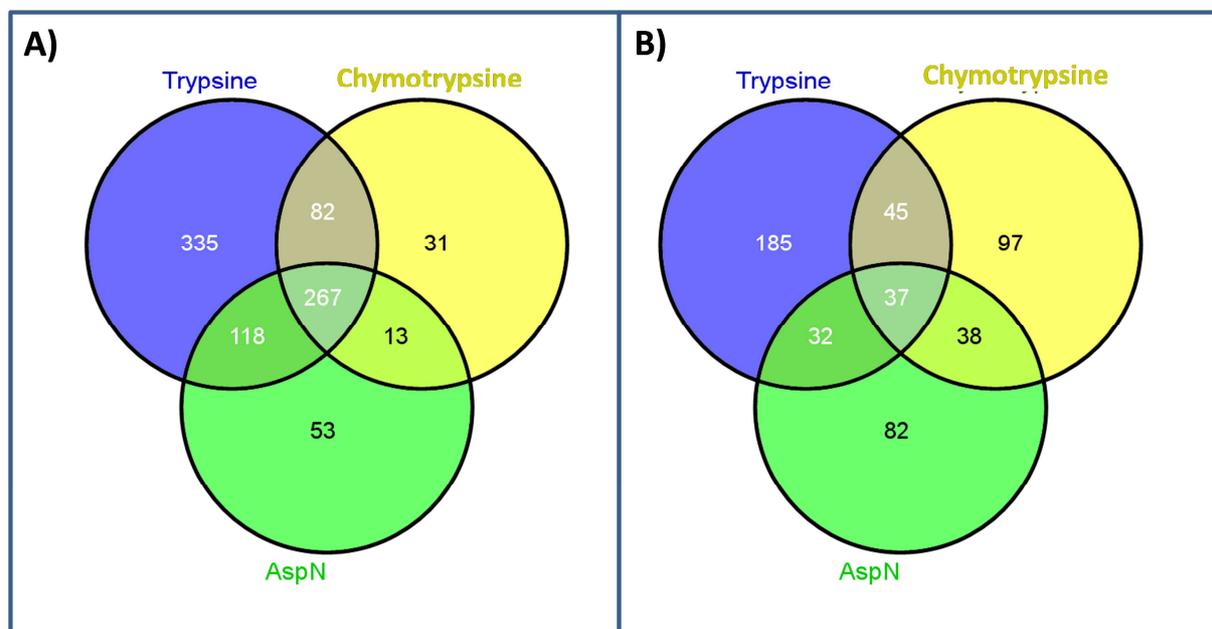


Figure 13 : Diagrammes de Venn montrant le nombre d'identifications de protéines (A) et d'extrémités N-terminales (B) obtenues par chaque digestion enzymatique.

L'analyse bioinformatique de ces résultats a permis de déterminer les voies métaboliques dans lesquelles sont impliquées les protéines identifiées et de les classer selon ces voies métaboliques. Le tableau 1 indique le nombre de protéines identifiées et classées dans chaque voie.

<b>Acides aminés</b>	<b>Nombre de protéines</b>
Métabolisme de l'asparagine et de l'aspartate	8
Métabolisme de l'arginine et de la proline	8
Métabolisme de la méthionine et des polyamines	21
Métabolisme du glutamate	14
Métabolisme de la Phénylalanine et la Tyrosine	10
<b>Total</b>	<b>61</b>
<b>Cofacteurs</b>	<b>Nombre de protéines</b>
Métabolisme de la vitamine B6 (Pyridoxal phosphate)	23
Métabolisme de l'ubiquinone	8
Métabolisme des thiamines	19
Biosynthèse des folates	14
<b>Total</b>	<b>64</b>
<b>Transport extracellulaire</b>	<b>Nombre de protéines</b>
Protéome de la vacuole Parasitophore	27
Protéines des vésicules de Maurer	109
<b>Total</b>	<b>136</b>
<b>Transport intracellulaire</b>	<b>Nombre de protéines</b>
ER/Golgi, translocation et control qualité	13
Cycle des rab-protéines	43
Complexes impliqués dans la voie sécrétoire endocytaire	34
Transport vésiculaire classique par l'intermédiaire COPI	39
<b>Total</b>	<b>129</b>
<b>Organites</b>	<b>Nombre de protéines</b>
Import des protéines vers les mitochondries	27
Flux électronique mitochondrial	23
Cycle du TCA Mitochondrial	20
Gène nucléaire portant des séquences signal de l'apicoplaste	284
Gènes nucléaires portant des séquences signal mitochondriales	280
Biogenèse de cytochrome oxidase	23
Voies de l'export des peptides à des mitochondries	5
<b>Total</b>	<b>662</b>
<b>Traduction</b>	<b>Nombre de protéines</b>
Initiation de la traduction	43
Biosynthèse des Protéines	55
Translocation post-traductionnelle chez eucaryotes	27
Reconnaissance moléculaire des protéines membranaire et des protéines sécrétées	10
Activation de eiF5A	3
<b>Total</b>	<b>138</b>

**Tableau 1 : Nombre de protéines identifiées par voie métabolique.**

Cette analyse bioinformatique est encore en cours de réalisation. Son objectif est de comparer nos données d'identification des sites de clivage protéolytique avec ceux de la littérature afin de dégager des pistes permettant la compréhension du rôle des protéines clivées et/ou acétylées que nous avons identifiées et d'orienter la suite du projet.

#### 4.4. Conclusion

Nous avons identifié à l'aide de notre stratégie analytique 957 protéines à partir de la fraction parasitaire de l'érythrocyte infecté. Nous avons pu aussi caractériser 550 extrémités N-terminales de ces protéines grâce à notre stratégie N-TOP. Notre stratégie s'est révélée, encore une fois, permettant de caractériser les extrémités N-terminales des protéines sans pénaliser l'identification des peptides internes et des peptides acétylés. Ces résultats préliminaires sont encore en cours d'analyse bioinformatique par notre collaborateur Dr. H. Ginsburg. Cette étude est la première étape d'un projet plus large visant à étudier les protéomes des différents compartiments de l'érythrocyte infecté par le parasite *P. falciparum*. Les autres fractions (vacuole parasitophore et fraction érythrocytaire) issues des autres compartiments de l'érythrocyte infecté seront aussi analysées ce qui devra permettre de réaliser une cartographie des protéines parasitaires et humaines de cet érythrocyte.

### 5. Etude de la voie de translocation de la PfCRT vers la membrane de la vacuole digestive de *Plasmodium falciparum*

Dans cette étude, nous nous sommes intéressés au mécanisme de translocation de la protéine PfCRT (Plasmodium *falciparum* Chloroquine Resistance Transporter). Cette protéine transporteur membranaire se trouve à la membrane de la vacuole digestive et est responsable, avec deux autres protéines de la membrane de cette vacuole, de la résistance chez *P. falciparum* à la chloroquine et aux autres drogues antipaludiques [103]. Ces deux autres protéines sont la PfMDR1 (« Pf Multi-Drug Resistance transporter 1 ») [104,105] et la PfMRP (Pf Multi-Drug Resistance-associated Protein) [106]. Les voies et mécanismes de translocation des protéines à destination de cette vacuole digestive sont méconnus. Nous avons donc étudié le mécanisme de translocation de la PfCRT à l'aide de mutations contrôlées et de la microscopie confocale de fluorescence à balayage laser et nous avons caractérisé cette protéine par spectrométrie de masse ce qui nous a permis de mettre en évidence l'implication de la phosphorylation d'un résidu thréonine dans la translocation de la PfCRT. Cette étude a fait l'objet d'une publication dans le journal Traffic en 2010 [107].

#### 5.1. Contexte de l'étude

Malgré la mise en place et le développement de plusieurs classes successives de drogues antipaludiques, *P. falciparum* reste une cause majeure de morbidité et de souffrance dans les pays endémiques [48]. La première drogue antipaludique efficace était la quinine extraite de l'écorce des *Cinchona* d'Amérique Latine [108]. Celle-ci a été remplacée par la chloroquine, un dérivé de quinoline, une alternative synthétique moins onéreuse, plus efficace et plus facile à produire. Si cette drogue a été employée pendant plus de 50 ans, l'émergence de parasites résistants au fil des années dans toutes les régions endémiques l'ont rendu inefficace. Les autres drogues antipaludiques non dérivées de quinoline ont souffert d'une très courte durée de vie et les résistances ont émergé au bout d'un an d'utilisation [108]. Il n'existe actuellement aucun médicament ou vaccin efficace

contre cette maladie et les campagnes visant à contrôler le paludisme se concentrent actuellement sur la lutte contre le vecteur du parasite [109].

La vacuole digestive est un organite cellulaire unique au genre *Plasmodium* [110]. Cette vacuole digestive sert d'organite de dégradation dans lequel l'hémoglobine importée par le parasite par endocytose est dégradée dans un processus impliquant des cystéines protéases (les plasmepsines) et des acides aspartiques protéases (les falcipainés) [111]. L'import et la dégradation de l'hémoglobine permet au parasite d'assurer ses besoins nutritionnels [112] et de contrer le risque de stabilité osmotique qui découle d'une perméabilité accrue de la membrane cytoplasmique de l'érythrocyte infecté aux ions [113]. L'hème libéré par cette dégradation est toxique pour le parasite, il est minéralisé en hémozoïne inerte dans cette vacuole digestive [114]. Nombre de drogues antipaludiques telles que la chloroquine et la quinine ciblent cette vacuole digestive et plus précisément le mécanisme de détoxification de l'hème conduisant à la formation de complexes toxiques qui finissent par tuer le parasite en perforant ses membranes intracellulaires [115].

La PfCRT est une protéine de 424 acides aminés contenant 10 domaines transmembranaires prédits et localisée à la membrane de la vacuole digestive [103]. Son rôle en tant que déterminant majeur de la résistance à la chloroquine a été identifié grâce à des analyses de croisement génétique entre des clones de *P. falciparum* chloroquino-résistants (CQR) et chloroquino-sensibles (CQS) par l'équipe de Welems [103]. Leurs expériences d'échange d'allèles de la PfCRT ont confirmé le rôle pivot du polymorphisme de PfCRT dans cette résistance. Les parasites CQR accumulent des quantités beaucoup plus faibles de chloroquine dans leur vacuole digestive que les parasites CQS [115] de façon à réduire la concentration de cette drogue dans la vacuole et de la maintenir en dessous des seuils nécessaires pour inhiber la biominéralisation de l'hème [116]. Cette observation a conduit à l'hypothèse que la PfCRT contrôle l'accès de la drogue à sa cible [117-120]. Dans ce contexte nous sommes intéressés à la caractérisation de cette protéine PfCRT dans le but d'étudier la voie de signalisation et le mécanisme de sa translocation vers la vacuole digestive. Ceci permettra d'améliorer notre compréhension du rôle de cette protéine dans la résistance du parasite aux drogues dans le but d'élaborer de nouvelles tactiques antipaludiques [108].

## 5.2. Etats des connaissances au début de l'étude

Le mécanisme qui régit le tri et la translocation des protéines vers la vacuole digestive est peu connu. Des études récentes suggèrent que la plasmepsine II passe tout d'abord par la voie sécrétoire du parasite en tant que preprotéine transmembranaire de type II pour être ensuite transportée à la membrane cytoplasmique du parasite. Elle est ensuite emmenée dans une vésicule cytosomale avec son substrat (l'hémoglobine) vers la vacuole digestive [121] où une convertase libère la protéine mature soluble [122,123]. Les séquences signal dirigeant les plasmepsines à la vacuole digestive semblent se trouver dans le pro-domaine [124]. Les falcipainés prennent eux aussi une voie similaire et il a été démontré que leur pro-domaine héberge une séquence signal bipartite pour la translocation vers la vacuole digestive [125,126]. Ces informations concernent les séquences signal intervenant dans la translocation des enzymes (solubles) vers la lumière de la vacuole digestive, mais il n'est pas encore clair comment les transporteurs de la vacuole digestive, et plus particulièrement les transporteurs de drogues, sont dirigés vers la membrane de cette vacuole.

Dans d'autres systèmes, la translocation des protéines membranaires lysosomales est régulée par des phosphorylations ou des lipidations [127]. Les protéines résidentielles ainsi que les protéines cargo sont dirigées à travers le réseau trans-Golgi directement au compartiment lysosomal ou endosomal ou indirectement à travers la membrane cytoplasmique puis à travers la voie endocytotique [128].

Nous avons caractérisé la protéine PfCRT et nos données de spectrométrie de masse montrent qu'une thréonine fonctionnelle en position 416 est indispensable à la translocation de cette protéine.

### 5.3. Stratégie analytique mise en œuvre

Tous les détails expérimentaux sont dans la partie « matériel et méthodes » de l'article dont fait l'objet cette étude [107]. Brièvement, des parasites *P. falciparum* de la lignée 3D7 sont cultivés et transfectés avec plusieurs construits et mutants de la PfCRT fusionnés avec la GFP (« Green Fluorescent protein »), CAD-GFP (« Conditional Agregation Domain - GFP »), ou la pHuorin. Des mutations ménagées et des délétions de séquence sont effectuées pour déterminer les domaines ou les résidus indispensables à la bonne translocation de la protéine chimère PfCRT-GFP. La localisation de la PfCRT-GFP est déterminée à l'aide de la microscopie confocale à balayage laser.

Pour les études par nanoLC-MS/MS, la PfCRT est immunoprécipitée à l'aide d'un anticorps anti-GFP. Elle est ensuite déposée sur un gel SDS-PAGE et la détection est effectuée en Western-Blot à l'aide d'un anticorps anti-GFP. La bande de gel est excisée, après une étape de réduction et d'alkylation, la protéine est digérée dans le gel à l'aide de la trypsine et les peptides générés sont analysés par nanoLC-MS/MS. Des précautions sont prises pour éviter l'élution précoce des peptides phosphorylés dans le volume mort en débutant le gradient à faible pourcentage en acétonitrile (2%).

Comme il existe plusieurs sites potentiels de phosphorylation, une deuxième analyse identique est effectuée après une digestion à l'aide de l'AspN. Les données de MS/MS générées sont soumis au moteur de recherche Mascot et sont comparées avec les spectres théoriques des peptides de digestion *in silico* des séquences de la banque NCBI restreinte à la taxonomie *P. falciparum* additionnées de la séquence de PfCRT-GFP. Les phosphorylations des résidus sérine, thréonine et tyrosine sont recherchés.

### 5.4. Résultats et discussion

Les analyses de microscopie confocale ont montré que la protéine PfCRT ne contient pas une séquence signal pour la voie endosomal/cytosomal. Les analyses des construits avec délétion des séquences C-terminales ont montré que les résidus 1 à 400 ne sont pas suffisant pour la translocation de la PfCRT à la vacuole digestive. Le construit chimérique PfCRT<sup>1-57</sup>-GFP (contenant seulement les 57 premiers acide aminés) montre un signal fluorescent à la périphérie du parasite et un petit signal à la périphérie de la vacuole digestive. Le construit PfCRT<sup>1-89</sup>-GFP (qui contient le premier domaine transmembranaire), PfCRT<sup>1-124</sup>-GFP, PfCRT<sup>1-160</sup>-GFP, PfCRT<sup>1-179</sup>-GFP, PfCRT<sup>1-240</sup>-GFP et PfCRT<sup>1-310</sup>-GFP ont tous montré un signal fluorescence dans le réticulum endoplasmique (RE). Le construit PfCRT<sup>1-400</sup>-PfCRT montre un signal fluorescent autour de la membrane cytoplasmique (Figure 14). Ces résultats indiquent que les résidus C-terminaux 400 à 427 sont indispensables à la bonne translocation de la PfCRT. La délétion de tous les domaines transmembranaires internes de la

PfCRT en gardant le premier domaine du côté N-terminal et le dernier domaine du côté C-terminal (construit PfCRT<sup>Δin</sup>) conduit à la rétention de la protéine dans le RE ce qui indique que les acides aminés 400 à 424 sont nécessaires mais pas suffisants à la translocation vers la vacuole digestive du parasite.

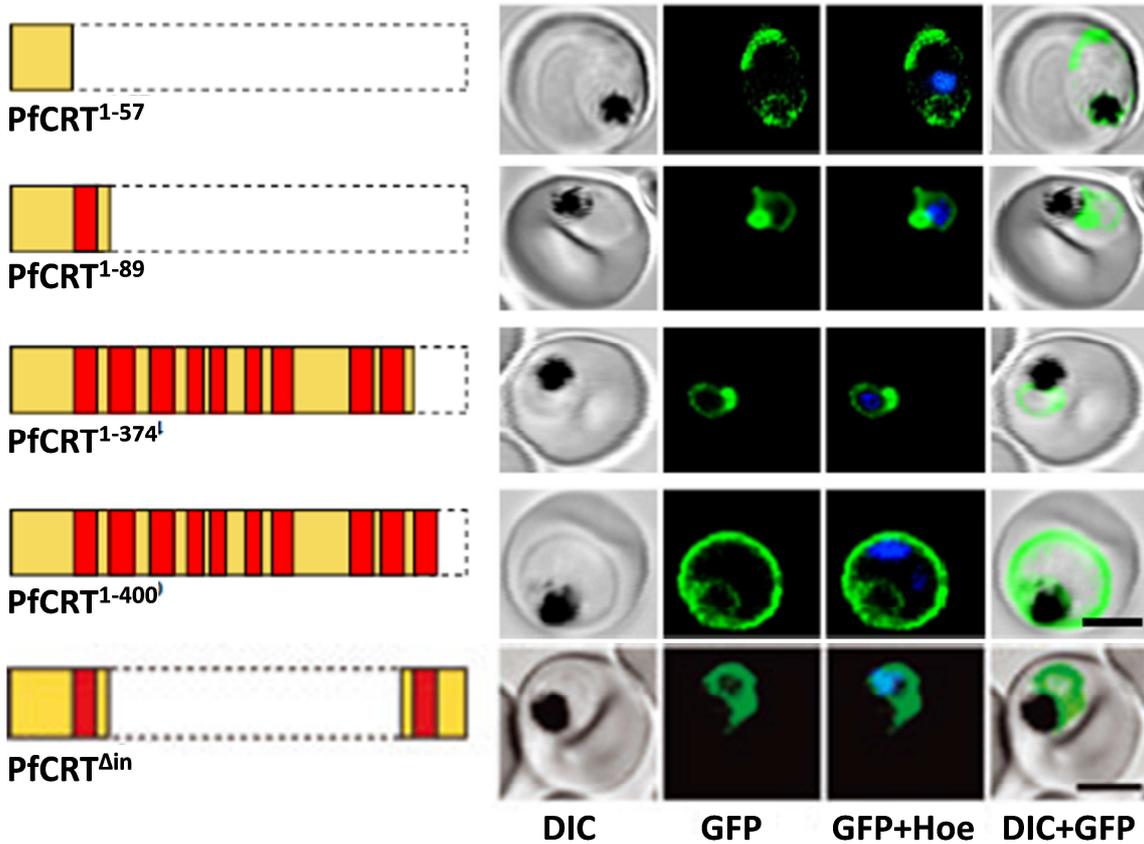


Figure 14 : Images de microscopie confocale montrant la localisation de différents construits chimériques de la PfCRT-GFP. Les domaines transmembranaires sont schématisés en rouge. DIC = contraste d'interférence différentielle, GFP = Signal de fluorescence de la GFP, GFP + Hoe = GFP + coloration nucléaire au Hoechst, DIC+GFP = superposition de DIC et de GFP.

Ensuite des mutations dans le domaine 400-424 en remplaçant les acides aminés 5 à 5 par des alanines ont été effectuées. Ceci a permis de constater que le remplacement des résidus <sup>413</sup>GELTN<sup>417</sup> par des alanines altère la translocation avec un signal fluorescent intense à la membrane cytoplasmique, tandis que le remplacement des autres résidus n'a pas d'effet (Figure 15 –A). Ceci indique que c'est dans cette séquence <sup>413</sup>GELTN<sup>417</sup> que se trouve un signal nécessaire à la translocation vers la vacuole parasitophore. Chaque acide aminé de cette séquence a donc été muté et les résultats montrent que le remplacement de la thréonine 416 par une alanine altère la translocation (figure 15 –B).

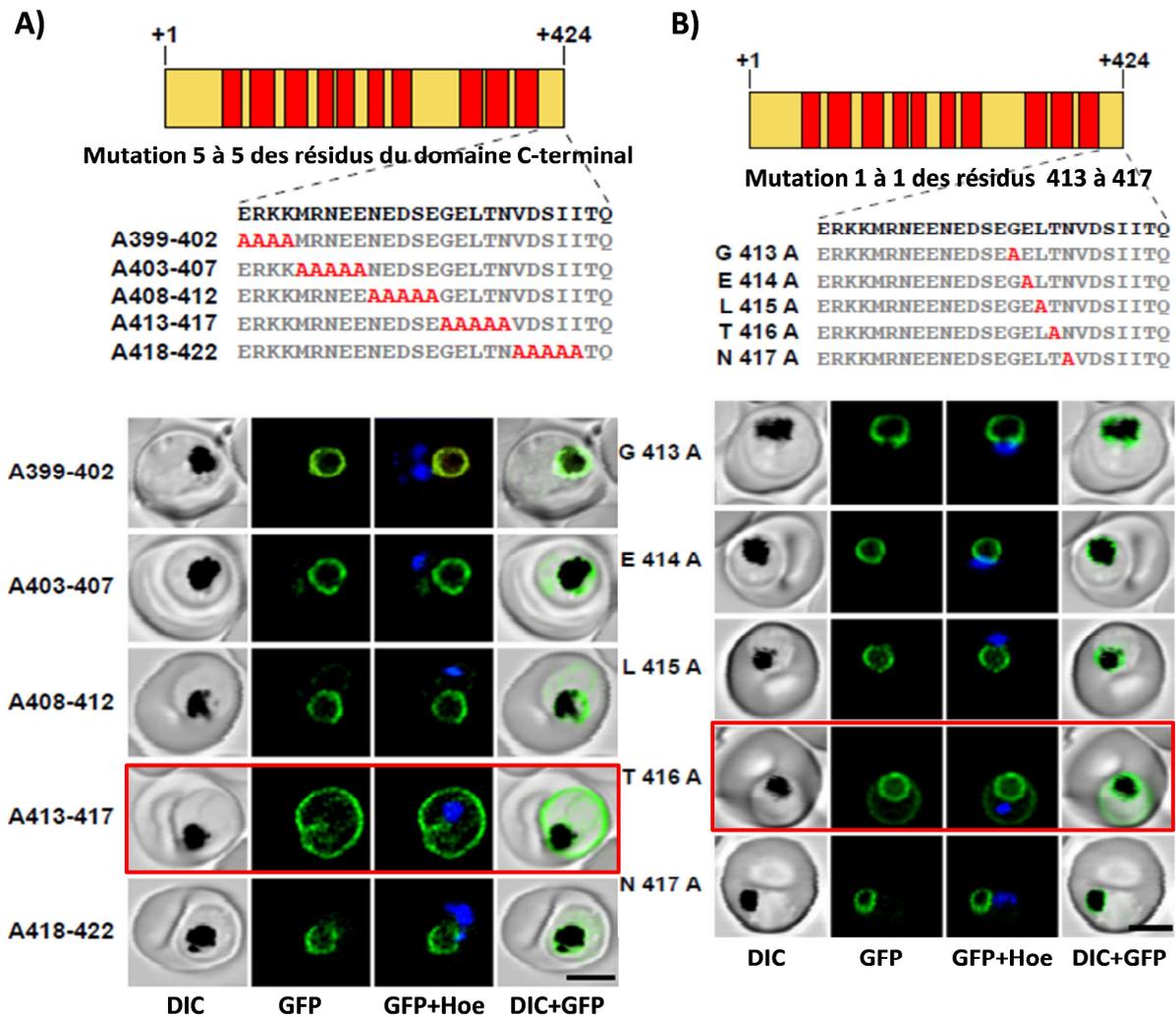


Figure 15 : A) localisation cellulaire des construits de la PfCRT-GFP avec mutations, cinq à cinq, des acides aminés du domaine C-terminal en alanine. B) localisation cellulaire des 5 construits après mutation un à un des résidus de la séquence <sup>413</sup>GELTN<sup>417</sup> en alanine. Le rectangle rouge indique une translocation altérée.

L'analyse bioinformatique du domaine C-terminal à l'aide de l'algorithme de prédiction PROSITE montre que la thréonine 416 peut être phosphorylée. Pour caractériser de la PfCRT par spectrométrie de masse, nous avons tout d'abord analysé le digest trypsique de la PfCRT-GFP en recherchant les éventuelles phosphorylations sur les résidus sérine, thréonine et tyrosine. A notre surprise, nous avons identifié le peptide trypsique C-terminal portant une phosphorylation sur la sérine 411 et non sur la thréonine 416 (Figure 16-A). Il s'agit du peptide <sup>405</sup>NEEDEDpSEGELTNVDSIIITQPR<sup>426</sup>. Ce peptide contient plusieurs sérines et thréonines susceptibles d'être phosphorylées, nous avons donc réanalysé manuellement le spectre à la recherche du même peptide portant une phosphorylation sur un autre résidu et qui aurai coélué avec le peptide phosphorylé sur la sérine. Cependant, nous avons pu attribuer tous les pics au peptide phosphorylé sur la sérine 411. Nous avons ensuite effectué une digestion de la PfCRT en remplaçant la trypsine avec l'endoprotéinase AspN. Ceci nous a permis d'identifier par LC-MS/MS deux peptides isobares ; l'un portant une phosphorylation sur la sérine 411 (Figure 16-B) et l'autre sur la thréonine 416 (Figure 16-C). Ces deux peptides sont séparés d'une minute dans nos conditions. Nous avons aussi identifié une phosphorylation sur la sérine 33.

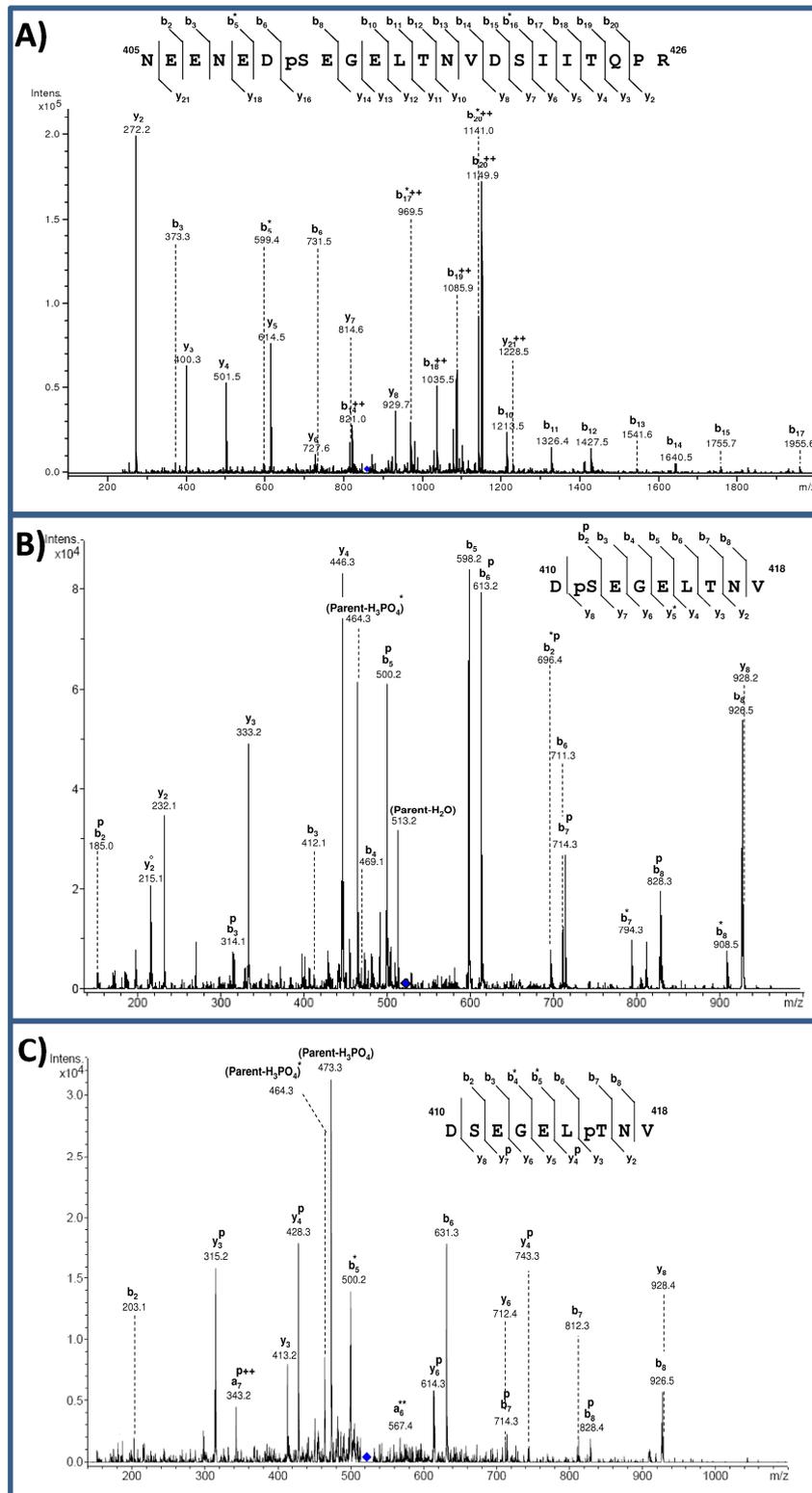


Figure 16: Spectres MS/MS des peptides contenant la thréonine 416. A) Peptide tryptique avec phosphorylation sur la sérine 411, B) peptide AspN avec phosphorylation sur la sérine 411, C) peptide AspN avec phosphorylation sur la thréonine 416.

Des extractions successives ont été effectuées afin de déterminer si la PfCRT endogène, la PfCRT-GFP et la PfCRT<sup>T416A</sup>-GFP sont des protéines membranaires intégrales (c'est-à-dire passant à travers la membrane) ou des protéines ancrées dans la membrane ou associées à la membrane. Si

ces deux protéines sont membranaires, elles résisteront à l'extraction par le carbonate de sodium et ne seront extractibles qu'en présence d'un détergent tel que le Triton X-100. L'urée permet d'extraire les protéines ancrées dans la membrane via des interactions protéine-protéine en dissolvant les complexes protéiques. L'extraction successive à l'aide du carbonate de sodium, suivie d'urée puis de Triton X-100 montre que toutes les protéines (PfCRT endogène, PfCRT-GFP et PfCRT<sup>T416A</sup>-GFP) sont dans la troisième fraction indiquant qu'il s'agit de protéines membranaires intégrales.

Etant donné que les expériences de délétion ont montré que les construits de la PfCRT sont transloqués au RE, la membrane cytoplasmique ou la vacuole digestive, les biologistes ont émis l'hypothèse que la PfCRT passe tout d'abord par la voie sécrétoire RE/Golgi, migre vers la membrane cytoplasmique pour être transportée par la voie endocytotique vers la vacuole digestive. Les parasites ont donc été traités avec la bréfeldine A qui bloque le transport à travers la voie RE/Golgi. Les études de localisation cellulaire montrent un signal indiquant que la PfCRT-GFP est bloquée dans le RE.

Un domaine d'agrégation conditionnel (CAD) a ensuite été ajouté à la séquence de deux construits chimériques donnant ainsi le construit PfCRT-CAD-GFP et PfCRT<sup>T416A</sup>-CAD-GFP. Ce domaine CAD induit l'agrégation de la protéine bloquant ainsi son mouvement et son translocation. Cette agrégation est réversible à l'aide d'un ligand anti-agrégation appelé AP21998, elle permet de suivre la localisation cellulaire des construits chimériques à différents moments. Ces expériences ont montré, pour PfCRT-CAD-GFP un signal fluorescent allant du RE directement vers la vacuole digestive. Pour PfCRT<sup>T416A</sup>-CAD-GFP, ce signal partait du RE vers la membrane cytoplasmique. A aucun moment, nous n'avons pu observer un signal fluorescent de PfCRT-CAD-GFP près de la membrane cytoplasmique ce qui suggère que la PfCRT prend une route directe du RE vers la vacuole digestive.

Nos données suggèrent que la translocation de la PfCRT à la vacuole digestive est un procédé qui comporte plusieurs étapes et qui requiert plusieurs signaux parmi lesquels une thréonine phosphorylée en position 416. Si la vacuole digestive présente des ressemblances avec les compartiments lysosomaux et endosomaux des cellules eucaryotes, aucune séquence signal comparable à celles trouvées chez les protéines endosomales ou lysosomales n'est présente chez la PfCRT. Il semble donc que la translocation de la PfCRT requière au moins trois signaux. Un signal interne non clivable pour le passage par la voie RE/Golgi, les domaines transmembranaires et enfin une thréonine phosphorylée en position 416. Le fait que nous détectons un peptide majoritaire avec une sérine phosphorylé en position 411 et une thréonine 416 non modifiée nous conduit à spéculer que la phosphorylation de cette thréonine est transitoire permettant le tri de la protéine à destination de la vacuole digestive. Les expériences avec le domaine d'agrégation conditionnel CAD suggèrent que la PfCRT prend une route directe vers la vacuole digestive. Cependant, nous ne pouvons pas exclure, étant donné que la résolution temporelle de notre expérience est de 30 secondes, que la PfCRT s'associe à des protéines de la membrane cytoplasmique d'une manière transitoire et rapide.

## 5.5. Conclusion

Nous avons pu montrer par nanoLC-MS/MS que la thréonine 416 qui est indispensable à la translocation de la PfCRT est phosphorylée. Cette étude a permis d'améliorer notre compréhension du mécanisme de translocation de la PfCRT. Elle ajoute à la notion du parasite *P. falciparum*

possédant des mécanismes de translocation sophistiqués. Ce mécanisme nécessitant une thréonine fonctionnelle phosphorylée s'ajoute à d'autres mécanismes dans le répertoire du parasite tel que celui faisant intervenir un motif pentamérique clivé et acétylé (motif PEXEL) ou un motif bipartite pour la translocation des protéines vers l'apicoplaste. Le rôle de la phosphorylation des sérines 33 et 411 est en cours d'analyse dans notre équipe. Les résultats de cette étude ont fait l'objet d'un article dans Traffic en 2010 et d'une communication par affiche (poster) au 58<sup>ème</sup> congrès de la Société Américaine de Spectrométrie de Masse (ASMS) à Salt Lake City.



## **Publication**

### Trafficking of the Phosphoprotein PfCRT to the Digestive Vacuolar Membrane in Plasmodium *falciparum*

Yvonne Kuhn, Cecilia P. Sanchez, Daniel Ayoub, Theodora Saridaki, Alain Van Dorsselaer and Michael Lanzer



# Trafficking of the Phosphoprotein PfCRT to the Digestive Vacuolar Membrane in *Plasmodium falciparum*

Yvonne Kuhn<sup>1,†</sup>, Cecilia P. Sanchez<sup>1,†</sup>, Daniel Ayoub<sup>2</sup>, Theodora Saridaki<sup>1</sup>, Alain van Dorselaer<sup>2</sup> and Michael Lanzer<sup>1,\*</sup>

<sup>1</sup>Hygiene Institut, Abteilung Parasitologie, Universitätsklinikum Heidelberg, Im Neuenheimer Feld 324, 69120 Heidelberg, Germany

<sup>2</sup>Laboratoire de Spectrométrie de Masse Bio-Organique, IPHC-DSA, Université de Strasbourg, CNRS UMR 7178, ECPM, 25 rue Becquerel, 67087 Strasbourg, France

\*Corresponding author: Michael Lanzer, michael.lanzer@med.uni-heidelberg.de

† These authors contributed equally to this work.

**The digestive vacuole plays an important role in the pathophysiology of the human malaria parasite *Plasmodium falciparum*. It is a terminal degradation organelle involved in the proteolysis of the host erythrocyte's haemoglobin; it is the site of action of several antimalarial drugs and its membrane harbours transporters implicated in drug resistance. How the digestive vacuole recruits residential proteins is largely unknown. Here, we have investigated the mechanism underpinning trafficking of the chloroquine resistance transporter, PfCRT, to the digestive vacuolar membrane. Nested deletion analysis and site-directed mutagenesis identified threonine 416 as a functional residue for sorting PfCRT to its site of residence. Mass spectroscopy demonstrated that threonine 416 can be phosphorylated. Further phosphorylation was detected at serine 411. Our data establish PfCRT as a phosphoprotein and suggest that phosphorylation of threonine 416 is a possible deciding signal for the sorting of PfCRT to the digestive vacuolar membrane.**

**Key words:** digestive vacuole, malaria, PfCRT, protein trafficking, sorting signal

Received 5 June 2009, revised and accepted for publication 10 November 2009, uncorrected manuscript published online 12 November 2009, published online 11 December 2009

The human malaria parasite *Plasmodium falciparum* has captured the scientific interest since its discovery more than a century ago. Initially, research on *P. falciparum* has been motivated by the desire to develop novel strategies to control malaria, an infectious disease that causes an estimated 515 million clinical cases and 1.0 million deaths annually (1). In recent years, *P. falciparum* itself has moved into focus as an organism that maintains a number of fascinating biological processes. Of particular

interest are the mechanisms underpinning protein trafficking, owing to the fact that *P. falciparum* possesses an extended set of morphologically and functionally distinct organelles to sustain its complex parasitic life cycle, alternating between vertebrate and invertebrate hosts (2,3). To direct newly synthesized proteins to the various subcellular compartments in a specific and co-ordinated manner, the parasite has established a protein sorting and trafficking machinery that, in terms of complexity, outranks that found in many model organisms in which protein transport has been studied thus far. A better understanding of the mechanisms underpinning protein trafficking in *P. falciparum* may, therefore, provide novel insights into a fundamental biological principle and, at the same time, may reveal new targets for rational intervention.

An organelle unique to the genus *Plasmodium* is the digestive vacuole (4). The digestive vacuole serves as a terminal degradation organelle where the haemoglobin that the parasite takes up from its host erythrocyte by endocytosis is degraded in a process involving aspartic and cysteine proteases, termed plasmepsins and falcipains, respectively (4–6). The toxic heme released from the proteolysed haemoglobin mineralizes to inert hemozoin (the malaria pigment) in the digestive vacuole (7). Several antimalarial drugs, including chloroquine and quinine interfere with heme detoxification (8). Resistance to these quinoline drugs is associated with polymorphisms within transporters of the digestive vacuolar membrane, such as PfCRT and PfMDR1 (9,10).

The mechanism underpinning sorting of proteins to the digestive vacuole is only rudimentarily understood. A recent study has suggested that plasmepsin II first enters the parasite's secretory pathway as a membrane inserted type II preprotein, is then transported to the parasite plasma membrane and is finally carried in cytosomal vesicles together with its substrate haemoglobin to the digestive vacuole (11), where a convertase releases the mature and soluble protein (12,13). The sorting signals directing plasmepsins to the digestive vacuole seem to reside within the pro-domain, as shown for plasmepsin IV (14). Falcipains are thought to take a similar route, and it has been shown that the pro-domain of falcipain II and III harbour a bipartite digestive vacuolar sorting signal flanking the transmembrane domain (15,16). Although there is some information regarding the signals mediating trafficking of enzymes to the digestive vacuolar lumen, it is still unclear how transporters of the digestive vacuolar membrane, particularly drug transporters, reach their destination.

In other systems, trafficking of lysosomal membrane proteins is mediated by distinct motifs, such as the tyrosine- and dileucine-based sorting signals, which interact with components of clathrin coats or adaptor protein complexes (17,18). In addition, phosphorylation and lipid modification can regulate trafficking of lysosomal membrane proteins (17). Residential and cargo proteins are routed from the *trans* Golgi network directly to the endosomal/lysosomal compartment or indirectly via the plasma membrane and then following the endocytotic pathway (18).

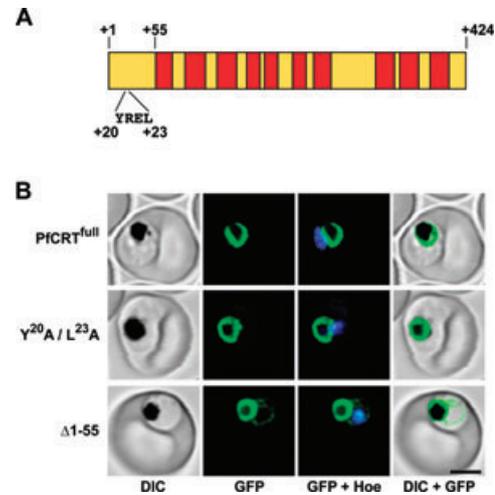
Here, we have studied trafficking of the chloroquine resistant determinant PfCRT to the membrane of the digestive vacuole. PfCRT is a protein of 424 amino acids, it contains 10 predicted transmembrane domains (10), and it is a putative member of the drug/metabolite transporter family (19,20). The *pfcr*t gene can carry mutational changes associated with chloroquine resistance (10,21,22), stereospecific responses to quinine and quinidine (23,24), and altered responsiveness to other antimalarial drugs, including amodiaquine and halofantrine (25). Biochemical studies have suggested that PfCRT may act as a pore (26) or a carrier (27–29) through which antimalarial drugs are released or expelled from the digestive vacuole of resistant *P. falciparum* parasites, where they would otherwise accumulate to toxic concentrations. Our data suggest that trafficking of PfCRT to the digestive vacuolar membrane requires a functional threonine at position 416.

## Results

### PfCRT lacks established endosomal/lysosomal sorting signals

Initially, we examined targeting of full-length PfCRT fused to the green fluorescent protein (GFP) (PfCRT<sup>full</sup>) (Figure 1A). Confocal laser scanning microscopy of parasites expressing PfCRT<sup>full</sup> revealed a fluorescence signal surrounding the malaria pigment (Figure 1B), consistent with a localization of the corresponding protein at the digestive vacuolar membrane. Quantification of the fluorescence signals revealed that 81 ± 3% of the total cellular fluorescence was associated with the digestive vacuolar compartment (Table 1). These data suggest that all signals necessary for correct targeting are contained within the PfCRT primary sequence.

Screening of the PfCRT polypeptide for established endosomal and lysosomal sorting signals identified several



**Figure 1: Effect of the N-terminal domain on trafficking of PfCRT.** A) Schematic representation of the PfCRT primary structure, including predicted transmembrane domains (red). A hypothetical tyrosine-based endosomal/lysosomal sorting motif is indicated (YREL). The positions of relevant residues are highlighted. B) Subcellular localization of different PfCRT/GFP chimera in *P. falciparum*-infected erythrocytes, detected by live cell confocal microscopy. PfCRT<sup>full</sup> corresponds to the full-length protein. Y<sup>20</sup>A/L<sup>23</sup>A indicates alanine replacements at positions 20 and 23 within full-length PfCRT. Δ1-55 lacks the first 55 amino acids of PfCRT. First column, differential interference contrast (DIC); second column, GFP; third column, GFP and nuclear staining with Hoechst; fourth column overlay of green and DIC channel. Representative images are shown in this as well as in the following figures. Scale bar = 4 μm.

motifs including a putative tyrosine-based motif (YREL, consensus motif YXXØ with Ø being a bulky, hydrophobic residue) within the N-terminal domain at position 20–23 (Figure 1A; Figure S1). To investigate whether this motif plays a role in trafficking, we generated a mutant (Y<sup>20</sup>A/L<sup>23</sup>A) in which the tyrosine and the leucine residue at positions 20 and 23, respectively, were replaced by an alanine in full-length PfCRT. As shown in other systems, tyrosine and leucine are essential for the function of a tyrosine-based lysosomal/endosomal sorting motif and cannot be replaced by other residues (30,31). However, replacing the two amino acids by alanines did not affect sorting of a full-length PfCRT-GFP chimeric protein (Figure 1B). Moreover, a chimeric protein lacking the first 55 amino acids of PfCRT (Δ1-55) was properly targeted to the digestive vacuolar membrane, although with a lesser

**Table 1:** Relative fluorescence associated with the digestive vacuolar compartment in *P. falciparum* strains expressing different PfCRT chimera. The relative fluorescence associated with the digestive vacuolar compartment is given in percent of the total cellular fluorescence, yielding F<sub>vac</sub>. The means ± SEM of (n) independent images are shown. <sup>a</sup>p < 0.01, in reference to PfCRT<sup>full</sup>; <sup>b</sup>p < 0.01 in reference to T<sup>416</sup>A

Protein	PfCRT <sup>full</sup>	PfCRT <sup>CAD</sup>	Δ1-55	PfCRT <sup>1-57</sup>	T <sup>416</sup> A	T <sup>416</sup> A <sup>CAD</sup>	PfCRT <sup>1-400</sup>	T <sup>416</sup> E	T <sup>416</sup> D	T <sup>416</sup> S
F <sub>vac</sub> (%)	81 ± 3 (27)	86 ± 3 (15)	67 ± 3 <sup>a</sup> (17)	19 ± 2 <sup>a,b</sup> (26)	12 ± 2 <sup>a</sup> (22)	9 ± 1 <sup>a</sup> (14)	8 ± 1 <sup>a</sup> (40)	11 ± 1 <sup>a</sup> (11)	37 ± 3 <sup>a,b</sup> (31)	17 ± 3 <sup>a,b</sup> (19)

efficiency (Figure 1B). Quantification of the fluorescence pattern revealed that  $67 \pm 3\%$  of the fluorescence signal was associated with the digestive vacuolar compartment and approximately 33% with the parasite's endoplasmic reticulum (ER) in cells expressing  $\Delta 1-55$  (Table 1). These findings suggest that neither the YREL sequence nor the first 55 amino acids play a crucial role in sorting PfCRT to the digestive vacuolar membrane.

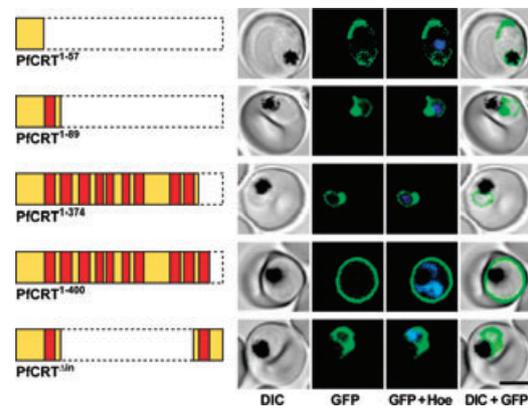
### **The C-terminal domain is necessary for efficient sorting of PfCRT to the digestive vacuolar membrane**

To better define the mechanism underpinning sorting of PfCRT to the digestive vacuolar membrane, we examined a series of nested C-terminal deletion constructs. A chimeric protein containing only the first 57 residues of PfCRT (PfCRT<sup>1-57</sup>) revealed an area of fluorescence at the parasite's periphery and some fluorescence at the periphery of the digestive vacuole ( $19 \pm 2\%$  of the total cellular fluorescence; Table 1) (Figure 2). In comparison, an extended protein containing the first 89 residues, including the first transmembrane domain (PfCRT<sup>1-89</sup>) showed a perinuclear fluorescence indicative of a localization at the ER (Figure 2). Comparable fluorescence patterns were observed for all further chimeric proteins extending up to residue 374 (PfCRT<sup>1-374</sup>; Figure 2). This included PfCRT<sup>1-124</sup>; PfCRT<sup>1-160</sup>; PfCRT<sup>1-179</sup>; PfCRT<sup>1-240</sup> and PfCRT<sup>1-310</sup> (data not shown). When the protein was extended to include amino acids 1–400 (PfCRT<sup>1-400</sup>), the subcellular localization changed. Now, a homogenous fluorescence was observed at the periphery of the parasite, consistent with a localization of the corresponding protein at the parasite plasma membrane (Figure 2). Only  $8 \pm 1\%$  of the total cellular fluorescence was associated with the digestive vacuolar compartment in parasites expressing PfCRT<sup>1-400</sup> (Table 1). These data indicate that the residues 1–400 are insufficient for effective sorting of PfCRT to the digestive vacuolar membrane and that additional information, possibly contained in the region defined by amino acids 400 to 424, is required.

Deleting all internal transmembrane domains of PfCRT, thereby appending the N-terminal domain including the first transmembrane domain with the C-terminal domain including the last transmembrane domain, resulted in a protein that remained in the parasite's ER (Figure 2; PfCRT <sup>$\Delta$ in</sup>). These findings suggest that amino acids 400–424, although necessary, are insufficient to direct PfCRT to its site of residence and that information from other parts of the protein has to contribute.

### **Threonine 416 is crucial for sorting of PfCRT**

To dissect the domain between residues 400–424, we performed an alanine scan across this region, replacing four or five successive amino acids by alanines within PfCRT<sup>full</sup>. This yielded PfCRT<sup>A399-402</sup>, PfCRT<sup>A403-407</sup>, PfCRT<sup>A408-412</sup>, PfCRT<sup>A413-417</sup> and PfCRT<sup>A418-422</sup> (Figure 3A). All chimeric proteins, with the exception of PfCRT<sup>A413-417</sup>, were targeted to the digestive vacuolar membrane, as



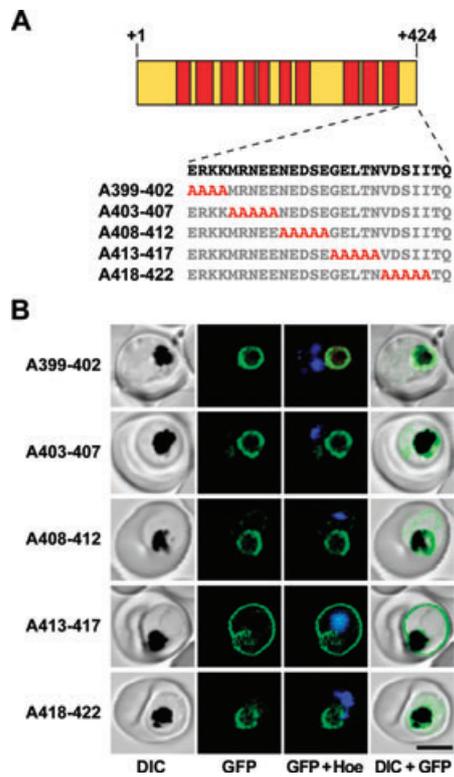
**Figure 2: Deletion analysis of PfCRT.** The confocal images show the subcellular localization of different PfCRT/GFP chimera in *P. falciparum*-infected erythrocytes. Schematic drawings depict the different deletion constructs. Predicted transmembrane domains are indicated in red. First column, differential interference contrast (DIC); second column, GFP; third column, GFP and nuclear staining with Hoechst; fourth column overlay of green and DIC channel. Scale bar = 4  $\mu$ m.

indicated by the fluorescence surrounding the malaria pigment (Figure 3B). In comparison, PfCRT<sup>A413-417</sup>, in which residues GELTN were replaced by alanines, displayed an altered fluorescence pattern, with a strong fluorescence signal being present at the parasite plasma membrane (Figure 3B).

To further delineate region 413–417, we sequentially replaced each of the five residues by an alanine (Figure 4A). Substituting G<sup>413</sup>, E<sup>414</sup>, L<sup>415</sup> or N<sup>417</sup> for an alanine did not affect sorting of PfCRT. The corresponding proteins revealed fluorescence patterns indicative of the digestive vacuolar membrane (Figure 4B). Only the T<sup>416</sup>A replacement altered sorting, with fluorescence being predominately present at the parasite plasma membrane (Figure 4B) and only  $12 \pm 2\%$  of the total subcellular fluorescence being associated with the digestive vacuolar compartment (Table 1).

### **Threonine 416 is phosphorylated**

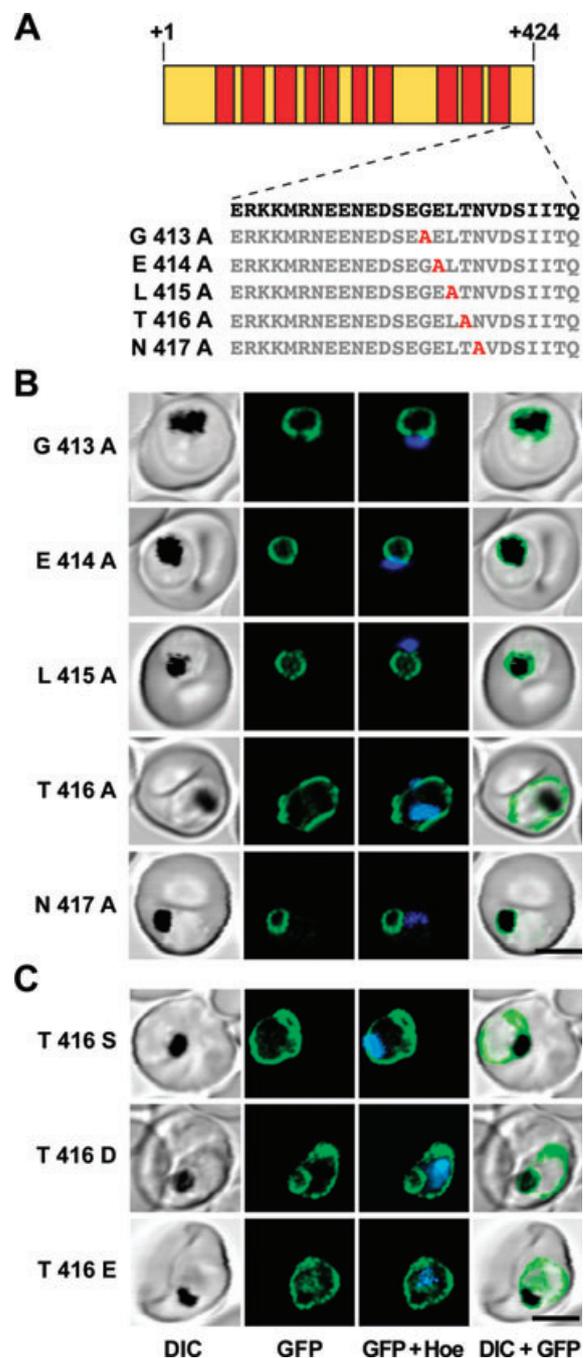
Analysis of the C-terminal domain of PfCRT by PROSITE predicted a possible phosphorylation of the threonine 416 by the threonine/serine protein kinase CK2. To explore the possibility that phosphorylation of threonine 416 plays a role in sorting of PfCRT, we replaced T<sup>416</sup> in the full-length PfCRT by serine, another polar amino acid that, like threonine, contains a free hydroxyl group that can be phosphorylated by CK2. In addition, we replaced T<sup>416</sup> by the negatively charged amino acids aspartic acid and glutamic acid, which, as shown in other systems, can mimic a phosphorylated threonine or serine (32). The mutants T<sup>416</sup>S and T<sup>416</sup>D displayed a complex fluorescence pattern, with fluorescence being present at the parasite plasma membrane and the digestive vacuolar compartment (Figure 4C). Quantification of the



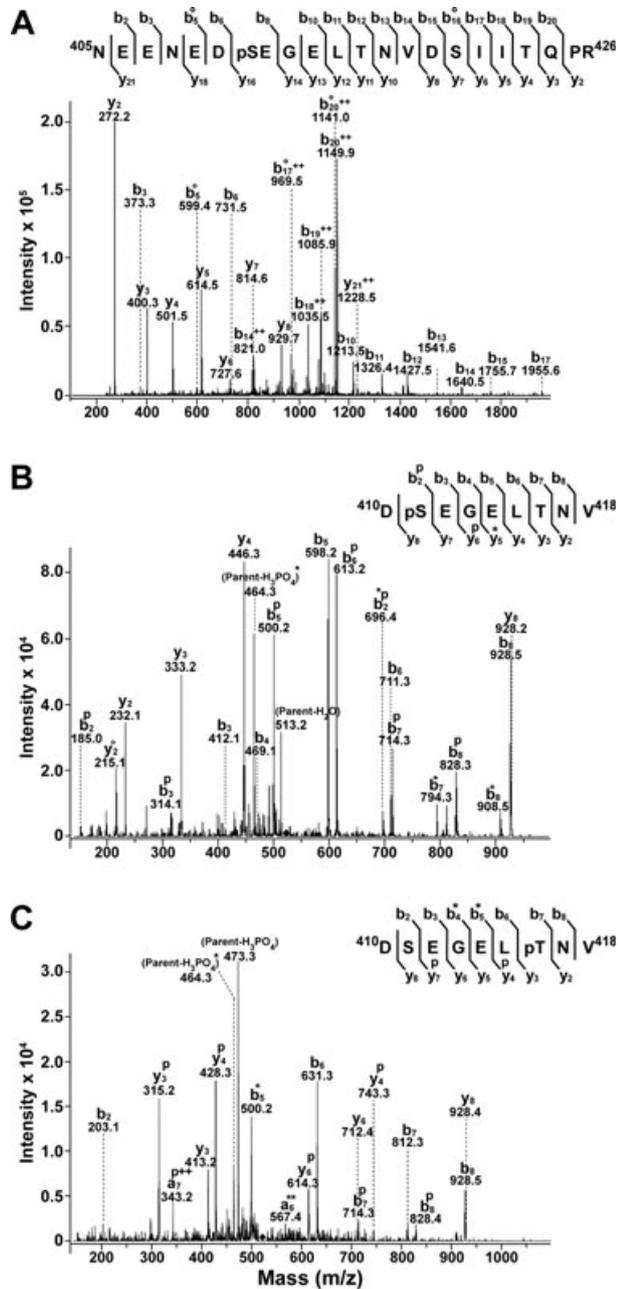
**Figure 3: Alanine scan across the C-terminal domain of PfCRT.** A) Schematic drawing of PfCRT. The sequence of the C-terminal domain of PfCRT is indicated and alanine replacements are highlighted in red. The position of the alanine replacements defines the name of the corresponding construct. B) Confocal images show the subcellular localization of the PfCRT/GFP chimera. First column, DIC; second column, GFP; third column, GFP and nuclear staining with Hoechst; fourth column overlay of green and DIC channel. Scale bar = 4  $\mu$ m.

fluorescence signals revealed a modest, but significant, and a major increase in digestive vacuolar fluorescence in the T<sup>416</sup>S and the T<sup>416</sup>D mutants, respectively, (17  $\pm$  3% and 37  $\pm$  3%), as compared with T<sup>416</sup>A (Table 1;  $p < 0.01$  in both cases), suggesting that serine and aspartic acid can partly substitute for threonine 416. No significant effects were observed for T<sup>416</sup>E, suggesting that glutamic acid cannot reconstitute the transport phenotype (Table 1).

We next investigated purified PfCRT<sup>full</sup> by nano-liquid chromatography coupled to tandem mass spectrometry (nanoLC-MS/MS) for possible modifications of threonine 416. To this end, PfCRT<sup>full</sup> was purified from transgenic *P. falciparum* by immunoprecipitation, using an anti-GFP antibody, followed by SDS-PAGE (Figure S2). PfCRT<sup>full</sup> was subsequently excised from the gels, reduced, alkylated, in gel digested with trypsin and subsequently subjected to nanoLC-MS/MS to characterize the tryptic peptides. A high quality fragmentation spectrum with a mascot ion score of 77 was obtained for the tryptic peptide containing the threonine 416. However, no modifications of threonine 416 were observed. Instead, we detected a



**Figure 4: Mutational analysis of a putative sorting signal within the C-terminal domain of PfCRT.** A) Schematic drawing of PfCRT, including transmembrane domains (red). The C-terminal residues are highlighted and alanine replacements are indicated in red. B) Confocal images of the subcellular localization of different PfCRT/GFP chimera with single alanine replacements. C) Mutational analysis of the threonine at position 416 within PfCRT. Confocal images of different PfCRT/GFP chimera in which the threonine (T) at position 416 was replaced by serine (T 416 S), aspartic acid (T 416 D) and glutamic acid (T 416 E). First column, DIC; second column, GFP; third column, GFP and nuclear staining with Hoechst; fourth column overlay of green and DIC channel. Scale bar = 4  $\mu$ m.



phosphorylation of serine 411 (Figure 5A). We repeated the experiment, but this time we digested the purified PfCRT fusion protein with AspN. Again, we observed phosphorylation of the serine 411 (Mascot ion score 45.2) (Figure 5B). In addition, we detected phosphorylation of threonine 416, with a 1 min retention time difference and a lower intensity chromatographic peak and a lower intensity MS/MS spectrum (Mascot ion score 39.2) (Figure 5C).

### Mapping the topology of PfCRT

To be assessable to phosphorylation, the T<sup>416</sup> should reside within the cytoplasm to interact with kinases. Topological models of PfCRT have predicted the C-terminal

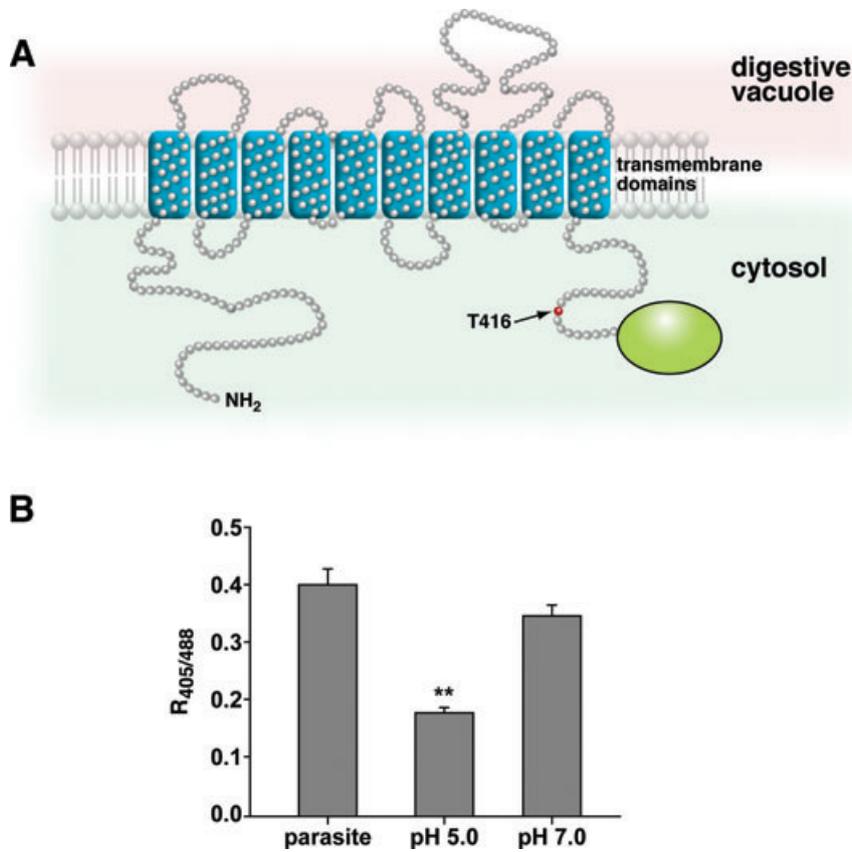
**Figure 5: Mass spectrometric analyses of the C-terminal domain of PfCRT.** A) CID mass spectrum of the tryptic junction peptide consisting of the C-terminus of PfCRT and the N-terminus of GFP. A series of 'y' ions excluded phosphorylation of serine 420 and threonine 423. Fragments b<sub>10</sub> and b<sub>11</sub> identified phosphorylation of serine 411. A representative example of two independent determinations totalling four spectra is shown. B) CID mass spectrum of the AspN peptide D<sup>410</sup>-V<sup>418</sup> of PfCRT. Fragments b<sub>2</sub> to b<sub>6</sub> are consistent with phosphorylation of serine 411. C) CID mass spectrum of the AspN peptide D<sup>410</sup>-V<sup>418</sup> of PfCRT showing phosphorylation of threonine 416. Fragments y<sub>3</sub> to y<sub>7</sub> and the mass differences between fragments b<sub>6</sub> and b<sub>7</sub> confirm phosphorylation of threonine 416. A total of 11 spectra were assigned to the D<sup>410</sup>-V<sup>418</sup> peptide of PfCRT, out of which four and one spectra unambiguously revealed phosphorylation of serine 411 and threonine 416, respectively. (+), double charged fragments; (°), fragments that lost an ammonia residue; (\*), fragments that lost a water molecule; (p) fragments with a neutral loss of 97.98 Da, which corresponds to H<sub>3</sub>PO<sub>4</sub> molecule.

domain to face the parasite's cytoplasm (19,20); however, experimental validation has yet to be presented. To map the orientation of the C-terminus, we fused full-length PfCRT to pHluorin (Figure 6A), a pH sensitive, ratiometric derivative of GFP (33). We reasoned that, depending on the topology of PfCRT within the membrane, pHluorin may face either the cytoplasm or the lumen of the digestive vacuole and should, therefore, record a neutral or an acidic pH, respectively.

pHluorin fluorescence was determined in live parasites expressing the PfCRT/pHluorin fusion protein at the digestive vacuolar membrane, using a confocal laser scanning microscopy. The fluorescence ratio R<sub>405/488</sub> was calculated and calibrated *in vivo* using two buffers with defined pH values of 7.0 and 5.0. A pH of approximately 7.3 was observed, consistent with a cytoplasmic localization of pHluorin and, hence, the C-terminal domain of PfCRT (Figure 6B).

### Extraction profiles of PfCRT chimeric proteins

We next investigated the extraction and solubility profile of PfCRT<sup>full</sup> and T<sup>416</sup>A, in comparison with that of the endogenous PfCRT. If the two PfCRT chimeric proteins are integral membrane proteins, as anticipated, then they should resist extraction from the membrane by sodium carbonate and urea and should only be extractable in the presence of a detergent, such as Triton X-100 (34). In comparison, peripheral membrane proteins can be removed from the membrane using sodium carbonate (35). Urea dissolves protein complexes and can typically extract proteins that are anchored into membranes via protein-protein interactions (36). To investigate the membrane association of PfCRT chimeric proteins, we lysed the cells by hypotonic shock and prepared total membrane fractions from erythrocytes infected with parasites expressing PfCRT<sup>full</sup> and T<sup>416</sup>A. The membrane fractions were subsequently extracted with sodium carbonate, urea and Triton X-100.



**Figure 6: Mapping the topology of PfCRT within the digestive vacuolar membrane.** A) Predicted topology of PfCRT within the digestive vacuolar membrane. The T<sup>416</sup> is indicated in red and highlighted by an arrow. pHluorin fused to the C-terminus of PfCRT is indicated in green. B) Parasites harbouring PfCRT/pHluorin at the digestive vacuolar membrane were incubated in physiological Ringer’s solution (pH 7.4) and the fluorescence ratio R<sub>405/488</sub> was determined in a confocal live cell set-up (bar labeled parasite). The fluorescence ratio was calibrated *in situ* using two buffers with a pH of 5.0 and 7.0, respectively. The mean ± SEM of at least 21 independent single cell determinations is shown. \*\*p < 0.001 using the Student’s t-test.

As shown in Figure 7, the endogenous PfCRT and the chimeric proteins PfCRT<sup>full</sup> and T<sup>416</sup>A, all had solubility properties consistent with integral transmembrane proteins. They were insoluble in sodium carbonate and urea buffer and partly soluble in the presence of Triton X-100 (Figure 7). In comparison, PfCRT<sup>1-57</sup>, containing the first 57 amino acids of PfCRT fused to GFP (see Figure 2), was partly extractable by sodium carbonate and, to a large extent, by urea and Triton X-100 (Figure 7).

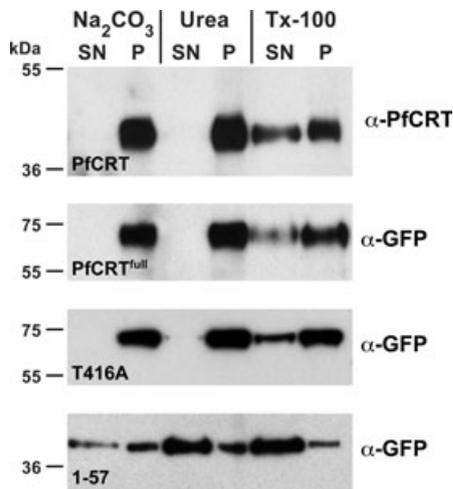
**PfCRT seems to takes a direct trafficking route**

Given that nested deletion constructs of PfCRT were targeted to either the ER/Golgi complex, the parasite plasma membrane and/or the digestive vacuolar membrane, we wondered whether PfCRT enters the secretory pathway and then traffics via the plasma membrane and subsequently the endocytotic pathway to its final destination. Previous studies have shown that transport through the ER/Golgi mediated pathway in *P. falciparum* is sensitive to brefeldin A (37). To investigate whether PfCRT enters the secretory pathway, we blocked anterograde protein trafficking by treating highly synchronized young trophozoites expressing PfCRT<sup>full</sup> with brefeldin A (5 µg/mL) for 3 h. A strong perinuclear fluorescence indicative of a localization at the ER was observed in the brefeldin A treated parasites, which was absent in the untreated controls (Figure 8). The arrest of protein transport was reversible after removal of brefeldin A from the culture medium and parasites were able to re-invade (data not shown). This

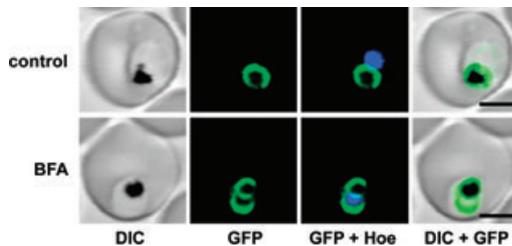
finding suggests that PfCRT enters the secretory pathway, consistent with previous reports (38).

To better analyze the trafficking route of PfCRT, we generated conditional trafficking mutants by inserting the conditional aggregation domain (CAD) (39) between PfCRT and GFP and between PfCRT<sup>T416A</sup> and GFP (Figure 9A), resulting in PfCRT<sup>CAD</sup> and T<sup>416</sup>A<sup>CAD</sup>, respectively. As previously shown, exported proteins containing the CAD aggregate in the ER, resulting in the arrest of protein transport, in a manner reversible by the anti-aggregation ligand AP21998 (39–41). In the absence of the anti-aggregation ligand, both PfCRT<sup>CAD</sup> and T<sup>416</sup>A<sup>CAD</sup> revealed a single dotted fluorescence signal in the vicinity of the nucleus (Figure 9B,D), consistent with a localization of the protein in the parasite’s ER. When the transgenic parasites were grown in the presence of a saturating concentration of the anti-aggregation ligand (1.0 µM), the subcellular distribution of the fluorescence pattern changed, with the fluorescence now being predominantly observed at the digestive vacuolar membrane, for PfCRT<sup>CAD</sup>, and at the parasite plasma membrane, for T<sup>416</sup>A<sup>CAD</sup> (Figure 9B,D; Table 1). These data indicate that trafficking of PfCRT and T<sup>416</sup>A<sup>CAD</sup> can be controlled by the CAD domain and the anti-aggregation ligand.

To follow transport of PfCRT and T<sup>416</sup>A<sup>CAD</sup> to their sites of residence over time, transgenic parasites expressing these two proteins were treated with cycloheximide



**Figure 7: Solubility profile of endogenous PfCRT and PfCRT fusion proteins.** A) Trophozoite stage-infected erythrocytes expressing PfCRT<sup>full</sup>, the single amino acid replacement mutant T<sup>416</sup>A or PfCRT<sup>1-57</sup> were lysed and the membranes were separated from the supernatant. The total membrane fractions were subsequently extracted using sodium carbonate, urea and Triton X-100. Extracted material was separated by centrifugation in supernatant (SN) and pellet (P) fractions. Extracts from  $2 \times 10^7$  parasites were examined per lane by western analysis using antibodies against GFP ( $\alpha$ -GFP) recognizing the GFP-fusion proteins PfCRT<sup>full</sup>, T<sup>416</sup>A and PfCRT<sup>1-57</sup>. As a control, the extraction profile of the endogenous PfCRT was analyzed in parallel, using a polyclonal guinea pig antiserum raised against PfCRT. Protein size standards are indicated in kDa. A representative example of three independent sets of experiments is shown.



**Figure 8: Effect of brefeldin A on the trafficking of PfCRT.** Parasites expressing PfCRT<sup>full</sup> were treated with 5  $\mu$ g/mL brefeldin A (BFA) for 3 h before confocal images were taken. The images were compared with an untreated control that was examined in parallel. First column, DIC; second column, GFP; third column, GFP and nuclear staining with Hoechst; fourth column overlay of green and DIC channel. Scale bar = 4  $\mu$ m.

(50  $\mu$ g/mL) to block *de novo* protein synthesis (40). Fixed cells were then analyzed at different time-points following the addition of the anti-aggregation ligand (Figure 9C,E). In the case of PfCRT<sup>CAD</sup>, fluorescence extended from its initial focus at the ER towards the digestive vacuolar membrane within less than 5 min after the addition of the ligand (Figure 9C). At the 60 min time-point, all fluorescence was localized at the digestive vacuolar membrane (Figure 9C).

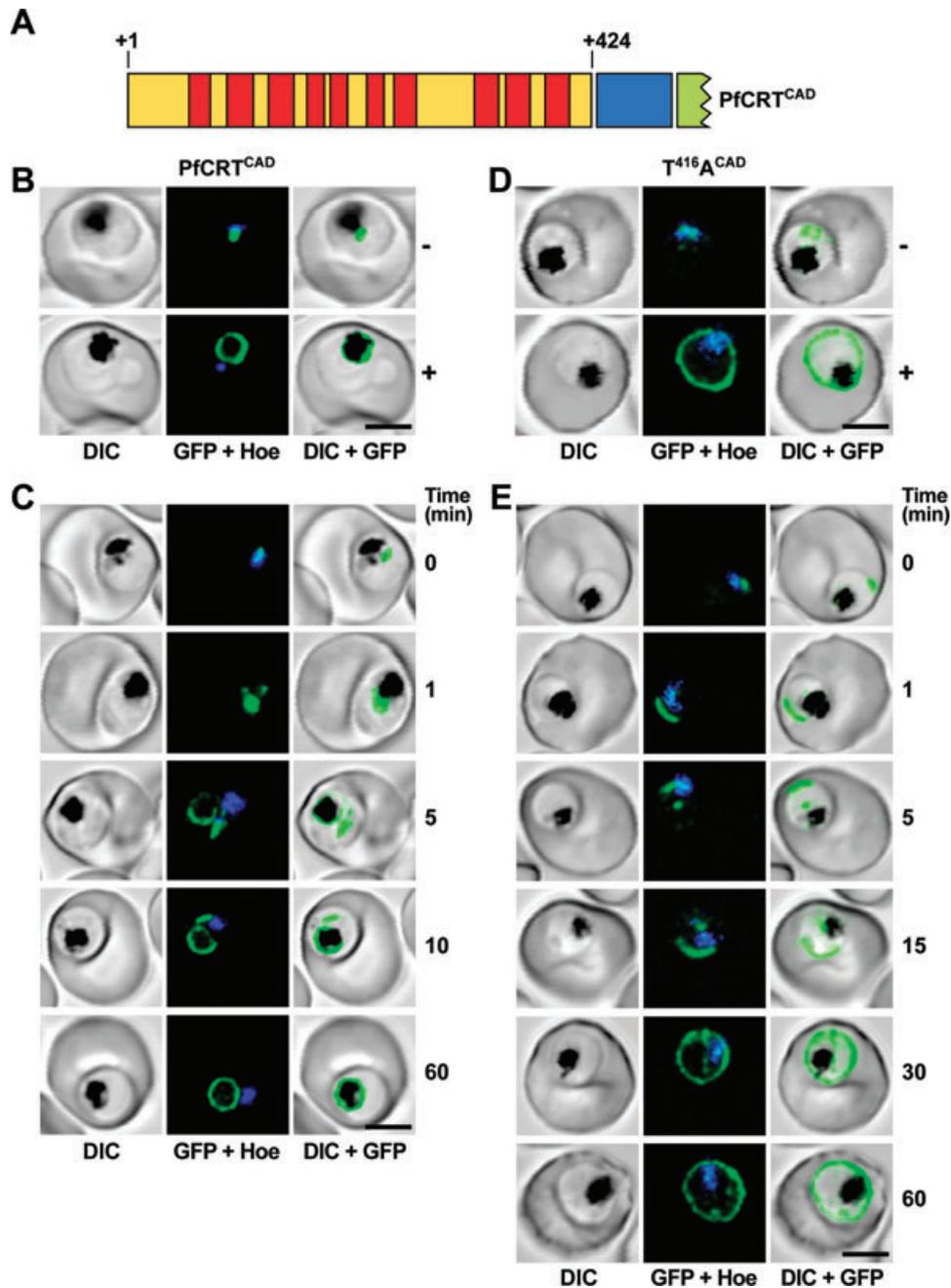
We repeated the experiment using live cell imaging of single cells. A time series revealed that, within 2.5 min after the addition of the ligand, PfCRT<sup>CAD</sup>-derived fluorescence spread towards the digestive vacuolar membrane (Figure S3). At no time-point did we observe fluorescence of PfCRT<sup>CAD</sup> at the parasite plasma membrane. In comparison, in T<sup>416</sup>A<sup>CAD</sup> expressing parasites, fluorescence spread from the ER to the parasite plasma membrane upon the addition of the anti-aggregation ligand (Figure 9E).

## Discussion

Our data suggest that trafficking of PfCRT to the digestive vacuolar membrane is a multi-step process that requires several signals including a functional threonine 416. The digestive vacuole of *P. falciparum* shares functional similarities with lysosomal and endosomal compartments of other eukaryotic cells. This includes an acidic milieu, residential proteases, and a role in protein degradation (5,6). One may surmise that the similarities extend to the mechanisms by which these organelles recruit residential proteins. However, none of the structural sorting signals described for lysosomal and endosomal membrane proteins in other systems (17,18) are functionally present in PfCRT. A sequence in the N-terminal domain at position 20–23 that resembles a tyrosine-based lysosomal sorting motif played no major functional role, as shown by amino acid replacement mutagenesis and deletional analysis (Figure 1).

The aspartic and cysteine proteases targeted to the lumen of the digestive vacuole for the degradation of the host erythrocyte haemoglobin also lack conserved lysosomal and endosomal sorting motifs (11,15). Instead, their sorting to the digestive vacuole seems to be mediated by a bipartite signal that flanks the transmembrane domain of the preprotein, as shown for the falcipain cysteine proteases (15). This bipartite motif seems to direct the protein, as a membrane-associated prepeptide, from the ER/Golgi complex to the parasite plasma membrane and then following the endocytotic pathway to the digestive vacuole (11,15), where a convertase releases the mature and soluble enzyme (12,13).

Trafficking of PfCRT seems to be just as complex, involving at least three intermediate steps and both topogenic and sorting signals. The finding that brefeldin A arrested transport of PfCRT suggests that PfCRT passes through the ER/Golgi complex (Figure 8) (38). Typical of a polytopic protein with no cleavable N-terminal signal sequence (42,43), the first transmembrane domain of PfCRT seems to function as an internal signal sequence and as the first signal anchor sequence, mediating translocation of the nascent protein into the ER (compare PfCRT<sup>1-57</sup> with PfCRT<sup>1-89</sup>; Figure 2). By analogy, the next transmembrane domain of PfCRT might serve as the first stop transfer anchor sequence.



**Figure 9: Analysis of the trafficking pathway of PfCRT<sup>full</sup> and PfCRT<sup>T416A</sup> using a conditional protein transport system.** A) Schematic representation of the PfCRT<sup>CAD</sup> chimeric protein containing the CAD (39) inserted between full-length PfCRT and GFP. B and D) Subcellular localization of PfCRT<sup>CAD</sup> and T<sup>416A</sup>PfCRT<sup>CAD</sup> in *P. falciparum*-infected erythrocytes grown in the presence (+) and absence (-) of the anti-aggregation ligand AP21998. C and E) Time-courses in minutes of PfCRT<sup>CAD</sup> and T<sup>416A</sup>PfCRT<sup>CAD</sup> trafficking upon the addition of the anti-aggregation ligand. Representative confocal images of fixed cells are shown. First column, DIC; second column, GFP; third column, overlay of green and DIC channel. Scale bar = 4 μm.

As shown in other systems, there are several other factors that affect membrane insertion of polytopic proteins. These include interactions between the hydrophobic stretches, the distance between them and the distribution of positive charges – with positively charged amino acids preferentially facing the cytoplasmic side of the

membrane (positive-inside rule) (44,45). Distortion of the topogenic process may explain why a chimeric protein that lacked the eight internal transmembrane domains of PfCRT remained in the ER (Figure 2; PfCRT<sup>Δin</sup>). Topogenic reasons may also be responsible for the reduced trafficking efficiency of a chimeric protein lacking the first 55

amino acids of PfCRT (Figure 1;  $\Delta 1-55$ ). The N-terminal domain is highly positively charged and it may, therefore, function as a topogenic signal for the correct insertion of PfCRT into the membrane.

We have no experimentally verified explanation why the N-terminal domain of PfCRT, when fused to GFP, behaved like a membrane or membrane-associated protein. Although lacking hydrophobic stretches, PfCRT<sup>1-57</sup> precipitated with the total membrane fraction and it required urea or Triton X-100 to extract PfCRT<sup>1-57</sup> from the membrane (Figure 7) (sodium carbonate extracted only a fraction of PfCRT<sup>1-57</sup>). Consistent with a proposed association with membranes, live cell fluorescence microscopy revealed that PfCRT<sup>1-57</sup> is localized at the periphery of the parasite and around the digestive vacuole (Figure 2). It is possible that PfCRT<sup>1-57</sup>, because of its highly positive charge, binds to negatively charged membrane components. Alternatively, PfCRT<sup>1-57</sup> may be anchored to the membrane through a covalently attached fatty acid. Whether this property reflects a physiological function or it simply results from aberrant expression and localization remains to be seen.

The region between residues 375 to 400, which includes the last transmembrane domain of PfCRT, seems to be important for the departure of the protein from the ER/Golgi complex (compare PfCRT<sup>1-400</sup> with PfCRT<sup>1-374</sup> that lacks residues 375–424; Figure 2). How residues 375–400 mediate this phenotype was not investigated in this study. In other systems, specific ER exit signals have been described that frequently consist of an acidic patch, such as DxE (46,47), or a di-hydrophobic motif (48), which recruit COPII proteins to depart from the ER in anterograde-directed vesicles (49). However, there are no obvious acidic patches or di-hydrophobic motifs in PfCRT in the domain defined by residues 375–400. Alternatively, these residues may indirectly affect trafficking by influencing the secondary and tertiary structure or the membrane topology of PfCRT. It should further be noted that GFP-tagged transmembrane proteins tend to be retained in the ER.

Essential for efficient sorting of PfCRT to the digestive vacuolar membrane was a threonine at position 416, as revealed by deletional analysis (Figure 2), alanine scans (Figure 3) and single amino acid replacements (Figure 4). Chimeric proteins in which the threonine was deleted or replaced by alanine revealed an aberrant transport phenotype in that the corresponding proteins were predominantly sorted to the parasite plasma membrane – instead of the digestive vacuolar membrane (Figure 4), although some minor fluorescence was also observed in the digestive vacuolar compartment (Table 1). Fluorescence associated with the digestive vacuole is commonly seen in transgenic *P. falciparum* lines expressing fluorescent chimeric proteins and is explained by the function of this organelle in protein degradation (50–52).

Both PfCRT<sup>full</sup> and the single amino acid replacement mutant T<sup>416</sup>A exhibited the extraction and solubility profile of an integral membrane protein. Like the endogenous PfCRT, which was analyzed in parallel, they resisted extraction by sodium carbonate and urea and were only partly extractable in buffer containing Triton X-100 (Figure 7). Thus, the different trafficking phenotypes displayed by PfCRT<sup>full</sup> and T<sup>416</sup>A cannot readily be explained by topogenic effects and, instead, suggest a unique functional role of threonine 416 in the sorting process of PfCRT to the digestive vacuolar membrane.

Threonine is an amino acid that can be posttranslationally modified, which, in turn, can regulate signal recognition and trafficking of membrane proteins, as shown in other systems (17). *In silico* data suggested that threonine 416 of PfCRT is a putative substrate of a serine/threonine kinase (data not shown). Direct experimental evidence for a possible modification of this residue came from mass spectroscopy. NanoLC-MS/MS of purified PfCRT<sup>full</sup> unequivocally detected phosphorylation of threonine 416 (Figure 5C). On the basis of our findings, it is tempting to speculate that phosphorylation of threonine 416 may function as a deciding signal in the sorting of PfCRT to the digestive vacuolar membrane.

This conclusion is supported by the following additional evidence: Serine, an amino acid that is structurally related to threonine and which, like threonine, possesses a free hydroxyl group that can be phosphorylated, modestly rescued the transport phenotype (Figure 4C and Table 1). An even better recovery of the trafficking phenotype was brought about by aspartic acid, an amino acid that can mimic a phosphorylated residue (Figure 4C and Table 1). Thus, the presence of a negative charge at position 416 seems to be important. Phosphorylation of threonine 416 may be transient and only be present while the protein traffics. The amount of phosphorylated versus non-phosphorylated threonine 416 could not be determined in this study owing to the fact that our nanoLC-MS/MS analysis of PfCRT was qualitative and not quantitative.

In addition to threonine 416, we detected phosphorylation of serine 411 (Figure 5A,B) and serine 33 (data not shown). These modifications did not affect trafficking of PfCRT, as shown by alanine replacement mutagenesis (Figure 3; A<sup>408-412</sup>) and deletional analysis (Figure 1B;  $\Delta 1-55$ ), respectively. The role of these and, possibly, other modifications of PfCRT are unclear, and are being subject to further investigation in our laboratory. It is further worth pointing out that the threonine 416 and the serine 411 are conserved in all *Plasmodium* spp. but not in the orthologues of other apicomplexa.

As shown in other systems, sorting signals of lysosomal proteins can be located at the cytoplasmic and/or luminal side of the membrane (17). They recruit trafficking factors that then direct the protein to its site of residence (17,18,53). To map the topology of PfCRT

within the digestive vacuolar membrane we used a live cell fluorescence approach in which we imaged the fluorescence signal of a C-terminal PfCRT/pHluorin fusion protein (Figure 6). pHluorin is a ratiometric pH sensitive GFP derivative, allowing intracellular pH values to be measured in transgenic cells expressing this protein (33). Proof of principle has recently been established in *P. falciparum* using pHluorin to determine the resting cytoplasmic and digestive vacuolar pH in this parasite (14). As to PfCRT, depending on the topology of the PfCRT/pHluorin fusion protein, pHluorin should reside either in the cytoplasm of the parasite or in the digestive vacuolar lumen and thus, should record a neutral or an acidic pH, respectively. Given that pHluorin measured a neutral pH, this finding suggested that the C-terminal domain of PfCRT, which includes threonine 416, is oriented towards the cytoplasm, consistent with *in silico* models of the membrane topology of PfCRT (19,20) (Figure 6). Thus, threonine 416 would be accessible to cytosolic kinases.

Although threonine 416 is a necessary signal it does not seem to be sufficient. A minimal construct termed PfCRT<sup>Δin</sup> consisting of residues 1–89 fused to residues 368–424 remained in the parasite's ER (Figure 2). It could well be that additional sorting signals are necessary or that the detailed topology of PfCRT is needed to ensure that this necessary signal at residue 416 can become effective.

Does PfCRT, like falcipains and plasmepsins (11,15) or lysosomal membrane proteins in other systems (17), travel via the plasma membrane and subsequently the endocytotic pathway to the digestive vacuole? One may jump to this conclusion based on the observation that chimeric proteins that lack a functional threonine at position 416 are predominantly targeted to the parasite plasma membrane (Figures 3 and 4). However, the results obtained using a conditional trafficking system (39) may suggest an alternative explanation.

In the absence of the anti-aggregation ligand, transport of PfCRT<sup>CAD</sup> was arrested in the ER (Figure 9B), as one would expect based on previous protein transport studies using the CAD in *P. falciparum* and in other organisms (39–41). Upon the addition of the ligand, PfCRT<sup>CAD</sup> fluorescence quickly spread from its initial focus towards the digestive vacuolar membrane (Figure 9C,D). At no time-point did we observe fluorescence at the parasite plasma membrane – not in fixed or live cells. A direct trafficking route of PfCRT from the ER/Golgi complex to the digestive vacuolar membrane would also be consistent with immunofluorescence and immuno-EM studies that exclusively localized PfCRT to the digestive vacuolar membrane (10,54). However, given the temporal resolution of our analysis of approximately 30 seconds, we cannot exclude the possibility of transiently associated protein rapidly trafficking via the parasite plasma membrane – which might have escaped detection.

If PfCRT takes a direct trafficking route, then PfCRT derivatives directed to the plasma membrane were mistargeted. This conclusion is further supported by T<sup>416</sup>A<sup>CAD</sup>. Upon the addition of the anti-aggregation ligand, T<sup>416</sup>A<sup>CAD</sup> trafficked to the parasite plasma membrane, most likely via the secretory pathway (the single dots of fluorescence present within the cytoplasm may represent T<sup>416</sup>A<sup>CAD</sup> carrying transport vesicles).

In summary, our data add to the notion that *P. falciparum* possesses a sophisticated protein trafficking system. By complementing an evolutionary conserved mechanism with components of its own design, the parasite seems to meet the demand for protein delivery to an extended set of subcellular compartments. In this context, previous studies have identified a pentameric host cell targeting motif for the export of proteins to destinations within the host erythrocyte (55–58) and a bipartite motif for sorting proteins to the parasite's apicoplast, an organelle remnant of a secondary endosymbiosis (51). Our finding that trafficking of PfCRT to the digestive vacuolar membrane requires a functional, and possibly phosphorylated, threonine extends our knowledge of sorting processes in this important protozoan parasite.

## Materials and Methods

### Parasite culture and transfection

The *P. falciparum* clone 3D7 was cultured as previously described (59), using RPMI medium supplemented with 5% human serum, 0.25% Albumax II (Gibco, Germany), 200 μM hypoxanthine and 20 μg/mL gentamycin. Cultures were synchronized by D-sorbitol lysis (60). Parasites were transfected, using 100 μg plasmid DNA, and transfectants were selected, using 5 nM WR99210, as previously described (50,61). Transfectants were detected in blood smears 14–21 days post-transfection.

### Vector construction

The *pfcr*t coding region was amplified from cDNA derived from the *P. falciparum* strain 3D7, using the primers CRT<sup>fw</sup> and CRT<sup>rev</sup> and cloned into the XhoI and AvrII restricted pARL1a<sup>+</sup> transfection vector (62) in a manner that allowed *pfcr*t expression to be controlled by its own promoter. The pHluorin and GFP coding sequences were amplified using the primers pH<sup>fw</sup> and pH<sup>rev</sup> and cloned into AvrII and XmaI restricted pARL1a<sup>+</sup> vector containing the *pfcr*t coding sequence to create PfCRT<sup>full</sup> and PfCRT/pHluorin, respectively. The C-terminal deletion constructs of *pfcr*t were generated using the forward primer CRT<sup>fw</sup> and the following reverse primers: CRT-1, CRT-2, CRT-3, CRT-4. The N-terminal deletion construct, Δ1-55, was generated using the primer pair CRT-5<sup>fw</sup> and CRT<sup>rev</sup>. PfCRT<sup>CAD</sup> was generated as follows: The CAD was released from the pC<sub>4</sub>S<sub>1</sub>-F<sub>M</sub> 4-FCS-hGH-vector (ARIAD Pharmaceuticals Inc.) using the restriction endonucleases XbaI and SpeI. The corresponding fragment was subsequently cloned into AvrII linearized PfCRT<sup>full</sup> and PfCRT<sup>416A</sup> between the *pfcr*t and the GFP coding region.

### Site-directed mutagenesis

Alanine scans and site-directed mutagenesis were performed using megaprimers (63) or the overlap extension method (64). For the mutagenesis of tyrosine and leucine at positions 20 and 23 within PfCRT, two megaprimers were amplified from PfCRT<sup>full</sup>, using the primer pairs CRT<sup>fw</sup>/CRT-6<sup>rev</sup> and CRT-7<sup>fw</sup>/CRT<sup>rev</sup>. This resulted in two complementary overlapping megafragments that were used as a template in the

subsequent overlap extension PCR with the primers  $CRT^{fw}$  and  $CRT^{rev}$ . The resulting PCR fragment was digested using the restriction endonucleases *XhoI* and *AvrII* and cloned into the pARL1a<sup>+</sup> vector. The C-terminal alanine scan within PfCRT was conducted by generating megaprimers in a first PCR reaction using  $CRT^{rev}$  as the reverse primer and one of the following forward primers: A399-402<sup>fw</sup>, A403-40<sup>fw</sup>, A408-412<sup>fw</sup>, A409-412<sup>fw</sup>, G413A<sup>fw</sup>, E414A<sup>fw</sup>, L415A<sup>fw</sup>, T416A<sup>fw</sup>, N417A<sup>fw</sup>, T416D<sup>fw</sup>, T416E<sup>fw</sup> and T416S<sup>fw</sup>. In the second PCR, PfCRT<sup>full</sup> template DNA, forward primer  $CRT^{fw}$  and one of the megaprimers were used to amplify the *pfCRT* coding region containing the pertaining mutation. A418–422 was generated using the primers  $CRT^{fw}$  and A418-422<sup>rev</sup>. All PfCRT fragments were digested with *XhoI* and *AvrII* and cloned into the pARL1a<sup>+</sup> vector.

### Primers

The sequences of the primers are compiled in Table S1.

### Live cell imaging

Live *P. falciparum*-infected erythrocytes were incubated in physiological Ringer's solution (122.5 mM NaCl, 5.4 mM KCl, 1.2 mM CaCl<sub>2</sub>, 0.8 mM MgCl<sub>2</sub>, 11.0 mM D-Glucose, 1.0 mM NaH<sub>2</sub>PO<sub>4</sub>, 25.0 mM HEPES) and viewed using an LSM510 confocal laser scanning microscope (Carl Zeiss). For GFP fluorescence detection, live cells were excited at a wavelength of 488 nm using an argon laser (Argon, laser power 40%, transmission 2%, C-Apochromat 63x/1.2 water immersion) and emission was captured through a 505–550 band pass filter. Nuclear staining was carried out using Hoechst33342 (5 μM) and images were captured using a 385–470 band pass filter upon excitation at a wavelength of 364 nm. Throughout the study, we examined parasites at the trophozoite stage. Representative examples of at least 20 independent observations are shown.

### Image processing

Images were deconvoluted using Huygens Essential (Scientific Volume Imaging). Fluorescence signals associated with the digestive vacuolar compartment and the entire cell were quantified from single, unprocessed confocal images, using IMAGEJ 1.42q, and subsequently divided by each other to yield the relative digestive vacuolar fluorescence in percent.

### pHluorin measurements

Live *P. falciparum*-infected erythrocytes expressing PfCRT/pHluorin at the digestive vacuolar membrane were incubated in physiological Ringer's solution and the pHluorin fluorescence intensities at the excitation wavelengths of 405 nm (3% transmission) and 488 nm (3% transmission) were determined, using a Zeiss LSM510 Meta confocal microscope. Background fluorescence was subtracted and the ratio of the fluorescence intensities  $R_{405/488}$  calculated. pH calibration was performed *in situ*, using the high potassium method (65) and buffers with a pH of 5.0 and 7.0, respectively, containing nigericin (10 μM) (Sigma).

### Brefeldin A treatment

A *P. falciparum* culture was synchronized twice using the sorbitol method (60). Young trophozoites were treated with 5 μg/mL brefeldin A (Sigma) for 3 h, as previously described (11).

### Membrane preparation and protein analysis

The solubility and extraction profile of *P. falciparum* membrane proteins were assessed as described (66,67). Briefly, infected erythrocytes at the trophozoite stage were lysed by hypotonic shock and a soluble and a membrane fraction was prepared. The membrane fraction was divided into three aliquots. One aliquot was treated with 6 M urea in 7.5 mM Tris–Cl pH 8.0 for 30 min at room temperature; another aliquot was treated with 1.0% Triton X-100 and 0.2 M SB201 in 10 mM Tris–Cl pH 8.0 for 30 min at room temperature; and the third aliquot was treated with 100 mM sodium

carbonate pH 11.5 for 30 min on ice. Samples corresponding to  $2 \times 10^7$  parasites were separated in soluble and insoluble components, size fractionated on SDS–PAGE gels and transferred to PVDF membranes. Membranes were analyzed using a monoclonal mouse α-GFP antibody (1:1000; Roche) or a polyclonal guinea pig PfCRT antisera (68). The appropriate horseradish conjugated antibodies were used at a concentration of 1:10 000 (Dianova).

### Conditional protein transport system

The dynamics of PfCRT trafficking was examined using fixed and live synchronized *P. falciparum*-infected erythrocytes (trophozoite stage) expressing PfCRT<sup>CAD</sup> and T<sup>416A</sup><sup>CAD</sup>. For the analysis of fixed cells, a saturating concentration of the anti-aggregation ligand AP21998 (1 μM) and cycloheximide (50 μg/mL) (Sigma) were added to the medium and the cells were returned to culture. At different time-points upon the addition of the ligand, samples were withdrawn, washed in phosphate-buffered solution (PBS) supplemented with the ligand and cycloheximide, and then fixed in 4% paraformaldehyde and 0.0075% glutaraldehyde (in PBS) for 30 min at room temperature, as described previously (69). Fixed cells were washed in PBS and stored at 4°C. Fixed cells were settled on a coverslip mounted on a Zeiss LSM510 confocal microscope. For live cell imaging, cells resuspended in physiological Ringer's solution were settled in a perfusion chamber heated to a constant temperature of 37°C. Representative live cells were selected and time series of individual cells were recorded after the addition of the anti-aggregation ligand.

### Immunoprecipitation and purification of PfCRT

Following synchronization with 1% sorbitol (60), infected erythrocytes at the trophozoite stage were purified from 600 mL parasite culture (5% haematocrit and 8% parasitaemia) using the Super Macs magnet (Miltenyi Biotec) and the D column. Infected erythrocytes were subsequently treated with 0.1% saponin in PBS. Immunoprecipitation was performed as described (40). Briefly, proteins were extracted using the RIPA buffer (1% NP-40, 1% sodium deoxycholate, 0.1% SDS, 150 mM NaCl, 10 mM Na-Phosphate buffer pH 7, 1 mM EDTA). The lysate was then diluted with 9 volumes of NETT buffer (10 mM Na-Phosphate buffer pH 7, 150 mM NaCl, 0.1% NP-40, and 1 mM EDTA). Prior to immunoprecipitation, lysates were cleared using goat IgGs. For immunoprecipitation, the ProFound™ Co-Immunoprecipitation Kit (Pierce) was used according to the manufacturer's instructions, using goat anti-GFP antiserum (Rockland). Immunoprecipitated material was washed in buffers with increasing NaCl concentrations ranging from 250 to 350 mM and one additional wash with 10 mM Na-phosphate buffer pH 7.0 before being size fractionated by SDS–PAGE (10% acrylamide, 6 M Urea). The gel was stained in Imperial Protein stain (Pearce) and slices of 2 mm width were excised around the size of the PfCRT fusion protein. All buffers used were supplemented with a cocktail of protease and phosphatase inhibitors (50 μg mL<sup>-1</sup> aprotinin, 20 μg mL<sup>-1</sup> leupeptin, 1 mM PMSF, 50 mM NaF, 0.2 mM NaVO<sub>3</sub>). For western analysis, the following antisera were used: mouse anti-GFP (Roche Diagnostics) (1/400), guinea pig anti-PfCRT (68); horseradish peroxidase antibodies (Jackson ImmunoResearch) (1/10 000; detection using the BM Chemiluminescence Blotting Substrate, Roche Diagnostics). The mass spectroscopic analysis shown in Figure 5A,B/C were performed on PfCRT-GFP fusion proteins purified from the *P. falciparum* strains 3D7 and Dd2, respectively.

### In gel digestion and sample preparation for mass spectrometry

The gel slices were predigested using a MassPREP automated system (Waters). The gel slices were washed twice in 25 mM ammonium hydrogen carbonate and acetonitrile. The cysteine residues were subsequently reduced in 10 mM dithiothreitol at 57°C and then alkylated in 55 mM iodoacetamide. After dehydration with acetonitrile, the proteins were cleaved in gel using 12.5 ng/μl of modified porcine trypsin (Promega) or 2.5 ng/μl of AspN (Roche Diagnostics) in 25 mM NH<sub>4</sub>HCO<sub>3</sub> at 37°C overnight. The tryptic and the AspN peptides were then extracted twice, using 0.4% CF<sub>3</sub>COOH in CH<sub>3</sub>CN/H<sub>2</sub>O (60/40) for the first extraction and

100% CH<sub>3</sub>CN for the second extraction. The volumes were then reduced to 5 µl by speed vacuuming to eliminate acetonitrile and to concentrate the peptides.

### NanoLC-MS/MS

Samples were analyzed using an Agilent 1200 series HPLC-Chip/MS system (Agilent Technologies) coupled to an HCT Plus ion trap mass spectrometer (Bruker Daltonics). The chip contained a Zorbax 300SB-C18 (150 mm × 75 µm, with a 5-µm particle size) column and a Zorbax 300SB-C18 (40 nL, 5 µm) enrichment column (Agilent Technologies). The following solvents were used: 2% acetonitrile, 0.1% formic acid in water (solvent A); and 2% water, 0.1% formic acid in acetonitrile (solvent B). Two-microlitres of sample was loaded onto the enrichment column at a flow rate of 3.75 µl/min using H<sub>2</sub>O/CH<sub>3</sub>CN (98:2) and 0.1% HCOOH as a solvent. Elution was performed at a flow rate of 300 nL/min with a 8–40% linear gradient (solvent B) over 45 min followed by a 70% stage (solvent B) over 3 min before the reconditioning of the column at 92% of solvent A. The HCT Plus ion trap was operated in the positive ion mode. The peptides were fragmented by collision-induced dissociation (CID). The voltage applied to the capillary cap was optimized at –1750 V. The MS scanning was performed in the standard enhanced resolution mode at a scan rate of 8.100 *m/z* per second. The mass range was 250–2000 *m/z*. The ion charge control was 100 000 and the maximum accumulation time was 200 ms. A total of three scans were averaged to obtain an MS spectrum. For tandem MS experiments, the system was operated with automatic switching between MS and MS/MS modes. The three most abundant peptides with an isolation width of 4 *m/z*, preferring doubly charged ions (absolute threshold of 10 000 and a relative of 5%) were selected on each MS spectrum for further isolation and fragmentation. Ions were excluded after 1 spectrum and released after 0.2 min. The MS/MS scanning was performed in the ultrascan resolution mode at a scan rate of 26 000 *m/z* per second. The mass range was 50–2800 *m/z*. The ion charge control was 300 000. A total of five scans were averaged to obtain an MS/MS spectrum. The complete system was fully controlled by ChemStation (Agilent Technology) and EsquireControl softwares (Bruker Daltonics).

### Protein identification

Mass data collected during analysis were processed, converted into .mgf files using DataAnalysis 4 Build 192 software (Bruker Daltonics). mgf files were submitted to Mascot search engine (Matrix Sciences) installed in its 2.2.0 version on a local server. Searches were performed against an in house generated composite target-decoy database containing NCBI protein sequences of *P. falciparum* (25 868 entries downloaded on 20 July 2009), the sequences of PfCRT-GFP and the contaminants trypsin and human keratins. Searches were performed with a tolerance on mass measurements of 0.5 Da for both precursor and fragment ions. Carbamidomethylation of cystein residues, oxidation of methionine residues, acetylation of protein N-terminal residues and phosphorylation of serine, threonine and tyrosine residues were searched as variable modifications. Up to one and four miss cleavages were allowed for tryptic and AspN digests, respectively. The spectra of the phosphorylated peptides of PfCRT were manually reinspected and annotated.

### Acknowledgments

This work was supported by a grant from the European Commission (BioMalPar cluster of excellence). We are further grateful to the 'Fondation pour la Recherche Medicale (FRM)' for financing a mass spectrometer. We also acknowledge the financial support of the Region Alsace. M. L. is a member of the Cellnetworks Cluster. We thank DiAngelis for the pHluorin construct. WR99210 was a gift from D. Jacobus, Jacobus Pharmaceuticals, Princeton, NJ. The CAD and the ligand AP 21998 were kindly provided by ARIAD Pharmaceuticals Inc., Cambridge, MA ([www.ariad.com/regulationkits](http://www.ariad.com/regulationkits)). We would further like to thank Petra Rohrbach, Matthias Seedorf and Wilfried Stein for discussions and critically reading this article. We thank Anurag Dave for the bioinformatic analysis of PfCRT.

## Supporting Information

Additional Supporting Information may be found in the online version of this article:

### Table S1: Primers and oligonucleotides used in this study.

**Figure S1: Primary sequence of PfCRT including hypothetical sorting motifs.** Putative tyrosine-based sorting motifs and acidic clusters are highlighted in red and blue, respectively. Residues within the acidic cluster that can be modified by phosphorylation are highlighted in green. Putative transmembrane domains of PfCRT are indicated in red. CYT, predicted cytosolic orientation; FV, predicted food vacuolar orientation.

**Figure S2: Western analysis of immunoprecipitated PfCRT<sup>full</sup>.** PfCRT<sup>full</sup> was purified from transgenic parasites by immunoprecipitation using an antisera against GFP. Immunoprecipitated protein was detected using an anti-PfCRT and an anti-GFP antiserum. The untransfected *P. falciparum* strain was examined as a control. A size standard is indicated in kDa. Lysate, total protein extract; IP, immunoprecipitate.

**Figure S3: Dynamics of PfCRT<sup>CAD</sup> trafficking in a live cell set-up.** The anti-aggregation ligand was added and the dynamics of PfCRT<sup>CAD</sup> trafficking was monitored in single live cells over time. A representative example is shown. First column, differential interference contrast (DIC); second column, GFP; third column, overlay of green and DIC channel. Scale bar = 4µm.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

## References

1. Snow RW, Guerra CA, Noor AM, Myint HY, Hay SI. The global distribution of clinical episodes of *Plasmodium falciparum* malaria. *Nature* 2005;434:214–217.
2. Tilley L, McFadden G, Cowman A, Klonis N. Illuminating *Plasmodium falciparum*-infected red blood cells. *Trends Parasitol* 2007; 23:268–277.
3. Tonkin CJ, Pearce JA, McFadden GI, Cowman AF. Protein targeting to destinations of the secretory pathway in the malaria parasite *Plasmodium falciparum*. *Curr Opin Microbiol* 2006;9:381–387.
4. Goldberg DE, Slater AF, Cerami A, Henderson GB. Hemoglobin degradation in the malaria parasite *Plasmodium falciparum*: an ordered process in a unique organelle. *Proc Natl Acad Sci USA* 1990; 87:2931–2935.
5. Goldberg DE. Hemoglobin degradation. *Curr Top Microbiol Immunol* 2005;295:275–291.
6. Banerjee R, Goldberg DE. The *Plasmodium* food vacuole. In: Rosenthal, PJ, editor. *Antimalarial Chemotherapy: Mechanisms of Action, Resistance and New Directions in Drug Discovery*. Totowa, New Jersey: Humana Press; 2001, pp. 43–64.
7. Jani D, Nagarkatti R, Beatty W, Angel R, Slebodnick C, Andersen J, Kumar S, Rathore D. HDP-a novel heme detoxification protein from the malaria parasite. *PLoS Pathog* 2008;4:e1000053.
8. Fitch CD. Ferritoporphyrin IX, phospholipids, and the antimalarial actions of quinoline drugs. *Life Sci* 2004;74:1957–1972.
9. Reed MB, Saliba KJ, Caruana SR, Kirk K, Cowman AF. Pgh1 modulates sensitivity and resistance to multiple antimalarials in *Plasmodium falciparum*. *Nature* 2000;403:906–909.
10. Fidock DA, Nomura T, Talley AK, Cooper RA, Dzekunov SM, Ferdig MT, Ursos LM, Sidhu AB, Naude B, Deitsch KW, Su XZ, Wootton JC, Roepe PD, Wellems TE. Mutations in the *P. falciparum* digestive vacuole transmembrane protein PfCRT and evidence for their role in chloroquine resistance. *Mol Cell* 2000;6:861–871.
11. Klemba M, Beatty W, Gluzman I, Goldberg DE. Trafficking of plasmeprin II to the food vacuole of the malaria parasite *Plasmodium falciparum*. *J Cell Biol* 2004;164:47–56.

12. Banerjee R, Francis SE, Goldberg DE. Food vacuole plasmepsins are processed at a conserved site by an acidic convertase activity in *Plasmodium falciparum*. *Mol Biochem Parasitol* 2003;129:157–165.
13. Kim YM, Lee MH, Piao TG, Lee JW, Kim JH, Lee S, Choi KM, Jiang JH, Kim TU, Park H. Prodomain processing of recombinant plasmepsin II and IV, the aspartic proteases of *Plasmodium falciparum*, is auto- and trans-catalytic. *J Biochem* 2006;139:189–195.
14. Kuhn Y, Rohrbach P, Lanzer M. Quantitative pH measurements in *Plasmodium falciparum*-infected erythrocytes using pHluorin. *Cell Microbiol* 2007;9:1004–1013.
15. Subramanian S, Sijwali PS, Rosenthal PJ. Falcipain cysteine proteases require bipartite motifs for trafficking to the *Plasmodium falciparum* food vacuole. *J Biol Chem* 2007;282:24961–24969.
16. Dasaradhi PV, Korde R, Thompson JK, Tanwar C, Nag TC, Chauhan VS, Cowman AF, Mohammed A, Malhotra P. Food vacuole targeting and trafficking of falcipain-2, an important cysteine protease of human malaria parasite *Plasmodium falciparum*. *Mol Biochem Parasitol* 2007;156:12–23.
17. Braulke T, Bonifacino JS. Sorting of lysosomal proteins. *Biochim Biophys Acta* 2009;1793:605–614.
18. Bonifacino JS, Traub LM. Signals for sorting of transmembrane proteins to endosomes and lysosomes. *Annu Rev Biochem* 2003;72:395–447.
19. Martin RE, Kirk K. The malaria parasite's chloroquine resistance transporter is a member of the drug/metabolite transporter superfamily. *Mol Biol Evol* 2004;21:1938–1949.
20. Tran CV, Saier MH, Jr. The principal chloroquine resistance protein of *Plasmodium falciparum* is a member of the drug/metabolite transporter superfamily. *Microbiology* 2004;150:1–3.
21. Lakshmanan V, Bray PG, Verdier-Pinard D, Johnson DJ, Horrocks P, Muhle RA, Alakpa GE, Hughes RH, Ward SA, Krogstad DJ, Sidhu AB, Fidock DA. A critical role for PfCRT K76T in *Plasmodium falciparum* verapamil-reversible chloroquine resistance. *EMBO J* 2005;24:2294–2305.
22. Sidhu AB, Verdier-Pinard D, Fidock DA. Chloroquine resistance in *Plasmodium falciparum* malaria parasites conferred by pfcr mutations. *Science* 2002;298:210–213.
23. Cooper RA, Ferdig MT, Su XZ, Ursos LM, Mu J, Nomura T, Fujioka H, Fidock DA, Roepe PD, Wellems TE. Alternative mutations at position 76 of the vacuolar transmembrane protein PfCRT are associated with chloroquine resistance and unique stereospecific quinine and quinidine responses in *Plasmodium falciparum*. *Mol Pharmacol* 2002;61:35–42.
24. Cooper RA, Lane KD, Deng B, Mu J, Patel JJ, Wellems TE, Su X, Ferdig MT. Mutations in transmembrane domains 1, 4 and 9 of the *Plasmodium falciparum* chloroquine resistance transporter alter susceptibility to chloroquine, quinine and quinidine. *Mol Microbiol* 2007;63:270–282.
25. Johnson DJ, Fidock DA, Mungthin M, Lakshmanan V, Sidhu AB, Bray PG, Ward SA. Evidence for a central role for PfCRT in conferring *Plasmodium falciparum* resistance to diverse antimalarial agents. *Mol Cell* 2004;15:867–877.
26. Bray PG, Mungthin M, Hastings IM, Biagini GA, Saidu DK, Lakshmanan V, Johnson DJ, Hughes RH, Stocks PA, O'Neill PM, Fidock DA, Warhurst DC, Ward SA. PfCRT and the trans-vacuolar proton electrochemical gradient: regulating the access of chloroquine to ferroprotoporphyrin IX. *Mol Microbiol* 2006;62:238–251.
27. Sanchez CP, Rohrbach P, McLean JE, Fidock DA, Stein WD, Lanzer M. Differences in trans-stimulated chloroquine efflux kinetics are linked to PfCRT in *Plasmodium falciparum*. *Mol Microbiol* 2007;64:407–420.
28. Naude B, Brzostowski JA, Kimmel AR, Wellems TE. Dictyostelium discoideum expresses a malaria chloroquine resistance mechanism upon transfection with mutant, but not wild-type, *Plasmodium falciparum* transporter PfCRT. *J Biol Chem* 2005;280:25596–25603.
29. Sanchez CP, Stein WD, Lanzer M. Is PfCRT a channel or a carrier? Two competing models explaining chloroquine resistance in *Plasmodium falciparum*. *Trends Parasitol* 2007;23:332–339.
30. Harter C, Mellman I. Transport of the lysosomal membrane glycoprotein Igp120 (lgp-A) to lysosomes does not require appearance on the plasma membrane. *J Cell Biol* 1992;117:311–325.
31. Williams MA, Fukuda M. Accumulation of membrane glycoproteins in lysosomes requires a tyrosine residue at a particular position in the cytoplasmic tail. *J Cell Biol* 1990;111:955–966.
32. Jones BG, Thomas L, Molloy SS, Thulin CD, Fry MD, Walsh KA, Thomas G. Intracellular trafficking of furin is modulated by the phosphorylation state of a casein kinase II site in its cytoplasmic tail. *EMBO J* 1995;14:5869–5883.
33. Miesenbock G, De Angelis DA, Rothman JE. Visualizing secretion and synaptic transmission with pH-sensitive green fluorescent proteins. *Nature* 1998;394:192–195.
34. Brandt ME, Riley BS, Radolf JD, Norgard MV. Immunogenic integral membrane proteins of *Borrelia burgdorferi* are lipoproteins. *Infect Immun* 1990;58:983–991.
35. Fujiki Y, Hubbard AL, Fowler S, Lazarow PB. Isolation of intracellular membranes by means of sodium carbonate treatment: application to endoplasmic reticulum. *J Cell Biol* 1982;93:97–102.
36. Borel AC, Simon SM. Biogenesis of polytopic membrane proteins: membrane segments assemble within translocation channels prior to membrane integration. *Cell* 1996;85:379–389.
37. Benting J, Mattei D, Lingelbach K. Brefeldin A inhibits transport of the glycoprotein-binding protein from *Plasmodium falciparum* into the host erythrocyte. *Biochem J* 1994;300 (Pt 3): 821–826.
38. Lee MC, Moura PA, Miller EA, Fidock DA. *Plasmodium falciparum* Sec24 marks transitional ER that exports a model cargo via a diacidic motif. *Mol Microbiol* 2008;68:1535–1546.
39. Rivera VM, Wang X, Wardwell S, Courage NL, Volchuk A, Keenan T, Holt DA, Gilman M, Orci L, Cerasoli F, Jr. Rothman JE, Clackson T. Regulation of protein secretion through controlled aggregation in the endoplasmic reticulum. *Science* 2000;287:826–830.
40. Saridaki T, Sanchez CP, Pfahler J, Lanzer M. A conditional export system provides new insights into protein export in *Plasmodium falciparum*-infected erythrocytes. *Cell Microbiol* 2008;10:2483–2495.
41. DeRocher A, Gilbert B, Feagin JE, Parsons M. Dissection of brefeldin A-sensitive and -insensitive steps in apicoplast protein targeting. *J Cell Sci* 2005;118:565–574.
42. Higby M, Junne T, Spiess M. Topogenesis of membrane proteins at the endoplasmic reticulum. *Biochemistry* 2004;43:12716–12722.
43. Wessels HP, Spiess M. Insertion of a multispanning membrane protein occurs sequentially and requires only one signal sequence. *Cell* 1988;55:61–70.
44. von Heijne G. Membrane-protein topology. *Nat Rev Mol Cell Biol* 2006;7:909–918.
45. van Geest M, Lolkema JS. Membrane topology and insertion of membrane proteins: search for topogenic signals. *Microbiol Mol Biol Rev* 2000;64:13–33.
46. Nishimura N, Balch WE. A di-acidic signal required for selective export from the endoplasmic reticulum. *Science* 1997;277:556–558.
47. Sevier CS, Weisz OA, Davis M, Machamer CE. Efficient export of the vesicular stomatitis virus G protein from the endoplasmic reticulum requires a signal in the cytoplasmic tail that includes both tyrosine-based and di-acidic motifs. *Mol Biol Cell* 2000;11:13–22.
48. Kappeler F, Klopfenstein DR, Foguet M, Paccaud JP, Hauri HP. The recycling of ERGIC-53 in the early secretory pathway. ERGIC-53 carries a cytosolic endoplasmic reticulum-exit determinant interacting with COPII. *J Biol Chem* 1997;272:31801–31808.
49. Barlowe C. Signals for COPII-dependent export from the ER: what's the ticket out? *Trends Cell Biol* 2003;13:295–300.
50. Przyborski JM, Miller SK, Pfahler JM, Henrich PP, Rohrbach P, Crabb BS, Lanzer M. Trafficking of STEVOR to the Maurer's clefts in *Plasmodium falciparum*-infected erythrocytes. *EMBO J* 2005;24:2306–2317.
51. Waller RF, Reed MB, Cowman AF, McFadden GI. Protein trafficking to the plastid of *Plasmodium falciparum* is via the secretory pathway. *EMBO J* 2000;19:1794–1802.
52. Wickham ME, Rug M, Ralph SA, Klonis N, McFadden GI, Tilley L, Cowman AF. Trafficking and assembly of the cytoadherence complex in *Plasmodium falciparum*-infected human erythrocytes. *EMBO J* 2001;20:5636–5649.
53. Kirchhausen T. Three ways to make a vesicle. *Nat Rev Mol Cell Biol* 2000;1:187–198.
54. Waller KL, Muhle RA, Ursos LM, Horrocks P, Verdier-Pinard D, Sidhu AB, Fujioka H, Roepe PD, Fidock DA. Chloroquine resistance

- modulated *in vitro* by expression levels of the *Plasmodium falciparum* chloroquine resistance transporter. *J Biol Chem* 2003;278:33593–33601.
55. Hiller NL, Bhattacharjee S, van Ooij C, Liolios K, Harrison T, Lopez-Estrano C, Haldar K. A host-targeting signal in virulence proteins reveals a secretome in malarial infection. *Science* 2004;306:1934–1937.
  56. Marti M, Good RT, Rug M, Knuepfer E, Cowman AF. Targeting malaria virulence and remodeling proteins to the host erythrocyte. *Science* 2004;306:1930–1933.
  57. Boddey JA, Moritz RL, Simpson RJ, Cowman AF. Role of the Plasmodium export element in trafficking parasite proteins to the infected erythrocyte. *Traffic* 2009;10:285–299.
  58. Chang HH, Falick AM, Carlton PM, Sedat JW, DeRisi JL, Marletta MA. N-terminal processing of proteins exported by malaria parasites. *Mol Biochem Parasitol* 2008;160:107–115.
  59. Trager W, Jensen JB. Human malaria parasites in continuous culture. *Science* 1976;193:673–675.
  60. Lambros C, Vanderberg JP. Synchronization of *Plasmodium falciparum* erythrocytic stages in culture. *J Parasitol* 1979;65:418–420.
  61. Fidock DA, Wellems TE. Transformation with human dihydrofolate reductase renders malaria parasites insensitive to WR99210 but does not affect the intrinsic activity of proguanil. *Proc Natl Acad Sci USA* 1997;94:10931–10936.
  62. Crabb BS, Rug M, Gilberger TW, Thompson JK, Triglia T, Maier AG, Cowman AF. Transfection of the human malaria parasite *Plasmodium falciparum*. *Methods Mol Biol* 2004;270:263–276.
  63. Colosimo A, Xu Z, Novelli G, Dallapiccola B, Gruenert DC. Simple version of “megaprimer” PCR for site-directed mutagenesis. *Biotechniques* 1999;26:870–873.
  64. Ho SN, Hunt HD, Horton RM, Pullen JK, Pease LR. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* 1989;77:51–59.
  65. Thomas JA, Buchsbaum RN, Zimniak A, Racker E. Intracellular pH measurements in Ehrlich ascites tumor cells utilizing spectroscopic probes generated *in situ*. *Biochemistry* 1979;18:2210–2218.
  66. Braun-Breton C, Jendoubi M, Brunet E, Perrin L, Scaife J, Pereira da Silva L. *In vivo* time course of synthesis and processing of major schizont membrane polypeptides in *Plasmodium falciparum*. *Mol Biochem Parasitol* 1986;20:33–43.
  67. Saridaki T, Frohlich KS, Braun-Breton C, Lanzer M. Export of PfSBP1 to the *Plasmodium falciparum* Maurer’s clefts. *Traffic* 2009;10:137–152.
  68. Nessler S, Friedrich O, Bakouh N, Fink RH, Sanchez CP, Planelles G, Lanzer M. Evidence for activation of endogenous transporters in *Xenopus laevis* oocytes expressing the *Plasmodium falciparum* chloroquine resistance transporter, PfCRT. *J Biol Chem* 2004;279:39438–39446.
  69. Tonkin CJ, van Dooren GG, Spurck TP, Struck NS, Good RT, Handman E, Cowman AF, McFadden GI. Localization of organellar proteins in *Plasmodium falciparum* using a novel set of transfection vectors and a new immunofluorescence fixation method. *Mol Biochem Parasitol* 2004;137:13–21.

## Chapitre 3 Analyse protéomique du tissu adipeux du rat en réponse à une privation de nourriture plus ou moins prolongée

### 1. Résumé

Les mécanismes moléculaires permettant à l'organisme de s'adapter à la privation de nourriture sont peu connus. Pour mieux les comprendre, le jeûne prolongé chez le rat est un modèle de choix. Il est caractérisé par une mobilisation séquentielle des substrats énergétiques, hydrates de carbone (phase 1), puis lipides (phase 2) et enfin protéines (phase 3). Le tissu adipeux est un organe endocrinien important lors des situations de métabolismes altérés, en plus d'être le réservoir majeur d'énergie. Cependant, les réponses moléculaires intervenant dans les cellules adipeuses restent peu étudiées. Par une approche basée sur la séparation/détection/quantification des protéines par électrophorèse bidimensionnelle (2D-DIGE) combinée à des analyses en nanoLC-MS/MS, nous avons donc étudié les variations du protéome du tissu adipeux entre des animaux nourris et des animaux en phase 2 et en phase 3 du jeûne.

Les variations de l'abondance relative de près de 25 protéines ont été mises en évidence. Ces variations représentent les bases moléculaires des grandes orientations du métabolisme au cours des phases du jeûne prolongé. Ainsi, les régulations identifiées sont celles qui participent à la réduction du métabolisme glucidique, aux variations du métabolisme des lipides et à celles du métabolisme des protéines entre la phase 2 et la phase 3. Les défenses antioxydantes et la capacité de repliement correct des protéines nouvellement synthétisées sont aussi altérées, ce qui pourrait compromettre l'intégrité des cellules adipeuses en phase 2, mais surtout en phase 3.

L'approche 2D-DIGE s'est donc révélée concluante puisque les nombreuses variations retrouvées apportent des informations précises sur les réponses métaboliques à long terme dans le tissu adipeux.

### 2. Contexte de l'étude

La malnutrition et la faim sont considérées comme des menaces majeures pour la santé publique par l'Organisation Mondiale de la Santé (OMS). D'après la Food and Agriculture Organisation, le nombre de personnes sous-alimentées dans le monde avoisinait le milliard en 2010 [129]. En cas de dénutrition, la perte des protéines corporelles est un risque léthal. Ce risque est le même dans le cas de certaines maladies sévères telles que le SIDA ou le cancer, dont les phases

terminales sont souvent associées à l'émaciation drastique du patient. Une meilleure compréhension des mécanismes moléculaires associés à la déplétion énergétique à long terme pourrait permettre de mieux traiter les patients émaciés, mais il existe très peu de données dans la littérature concernant ce sujet.

A long terme, la masse et la composition corporelle sont peu variables. Selon la théorie du point de consigne, toute déviation de la masse corporelle induit des changements dans le comportement alimentaire qui visent à restaurer la masse initiale [130]. La théorie lipostatique propose que les réserves adipeuses sont jaugées et régulées par une boucle de rétrocontrôle négatif impliquant un relais humoral ou métabolique qui contrôlerait la prise alimentaire et le niveau des réserves endogènes [131]. Cependant, les mécanismes de régulation de la balance et des réserves énergétiques sont encore largement méconnus.

### 3. Le modèle du jeûne prolongé

Le jeûne prolongé chez l'animal de laboratoire constitue un modèle de choix pour étudier la réponse de l'organisme à la privation de nourriture. En effet, les événements métaboliques survenant en cas de jeûne prolongé sont très similaires entre les animaux de laboratoires, les oiseaux sauvages et les patients en stade critique d'anorexie [132-134]. Des phases métaboliques successives ont été décrites dans le jeûne prolongé [135,136]. Après une brève phase 1 (P1) d'adaptation marquée par l'épuisement des réserves de glycogène, la phase 2 (P2) est caractérisée par une forte mobilisation des réserves de lipides alors que les protéines sont épargnées. L'épargne des protéines est une adaptation majeure au jeûne prolongé retrouvée quelle que soit l'espèce considérée. En effet, une perte des protéines corporelles à hauteur de 50 % se révèle mortelle [137]. Or la phase 3 (P3), qui est déclenchée lorsque qu'un seuil bas de l'adiposité est atteint, correspond justement à l'augmentation considérable du catabolisme des protéines tandis que la mobilisation et l'utilisation des réserves de lipides diminuent. Bien que critique, il a été montré que même lorsque la phase 3 est atteinte, les animaux réalimentés rétablissent leur réserves ce qui montre que ce processus est réversible [138]. Etant donné les changements comportementaux qui ont lieu durant la phase 3, l'existence d'une voie de signalisation, qui n'est pas encore déterminée, induisant ces comportements avant l'épuisement total des réserves de graisse a été rapportée [139].

### 4. Position du problème

D'après la littérature, la mobilisation des réserves énergétiques lipidiques stockées dans le tissu adipeux blanc pendant le jeûne implique une régulation phase-dépendante des taux de ARNm codant les lipases essentielles et d'autres facteurs liés à la lipolyse [140]. Le tissu adipeux, tant en sa qualité de réservoir majeur de substrats énergétiques que par ses fonctions endocrines est au centre des adaptations à la privation de nourriture. Certaines adipokines, comme la leptine, jouent un rôle actif dans l'établissement des transitions métaboliques au cours du jeûne et de la réalimentation. Les niveaux circulants de leptine contrôlent en effet l'expression des neuropeptides hypothalamiques au cours du jeûne [141]. Par ailleurs, d'autres molécules adipocytaires sont susceptibles de jouer un rôle de lipostat, mais ceci, de même que leur implication dans la réponse à une déplétion énergétique, reste à démontrer [142]. Enfin, les régulations moléculaires induites par le jeûne à long terme dans le tissu adipeux restent encore à ce jour peu étudiées. Une étude globale des variations du protéome

du tissu adipeux en réponse à un jeûne prolongé pourrait donc apporter des réponses quant à son implication possible et aux régulations mises en place en situation de balance énergétique négative. Chez les mammifères, le tissu adipeux est réparti entre plusieurs localisations anatomiques ayant des propriétés différentes. Au cours du jeûne, la mobilisation des substrats lipidiques stockés dans les tissus adipeux intra-abdominaux est plus grande que celle des substrats mis en réserve dans les dépôts sous-cutanés. Les bases moléculaires de ces différences restent méconnues. Là encore, une étude globale des variations du protéome réalisée dans un tissu adipeux intra-abdominal et dans un dépôt sous-cutané, pourrait permettre de mieux comprendre les différences physiologiques qui existent entre ces diverses localisations anatomiques.

## 5. Objectif

L'objectif de cette étude était d'identifier les protéines qui sont différentiellement exprimées entre l'état nourri et les phases 2 et 3 du jeûne prolongé. Nous avons donc étudié, au cours du jeûne prolongé chez le rat de laboratoire, les variations du protéome d'un dépôt viscéral (EPI, dépôt épидidymaire) et d'un dépôt sous-cutané (SC).

## 6. Stratégie analytique

Après le traitement des animaux, le prélèvement des dépôts adipeux et l'extraction des protéines, cette étude (Figure 1) reposait sur l'utilisation de l'électrophorèse bidimensionnelle 2D DiGE couplée à des analyses en spectrométrie de masse (nanoLC-MS/MS).

L'étude protéomique différentielle de l'EPI et du SC a été envisagée selon le plan d'expérience suivant :

- Expérience 1 : Comparaison du protéome de l'EPI entre des animaux nourris, en P2 et en P3 ;
- Expérience 2 : Comparaison du protéome du SC entre des animaux nourris, en P2 et en P3 ;
- Expérience 3 : Comparaison du protéome de l'EPI et du SC chez des animaux nourris.

Le traitement des animaux a été réalisé en collaboration avec les équipes du département d'Ecologie, Physiologie et Ethologie (DEPE) de l'IPHC. L'extraction et la séparation des protéines ont été réalisées par M. Thierry Wasselin dans le cadre de sa thèse, de même que l'analyse d'image. Pour ma part, j'ai pris en charge les analyses LC-MS/MS des protéines contenues dans tous les spots de gel excisés.

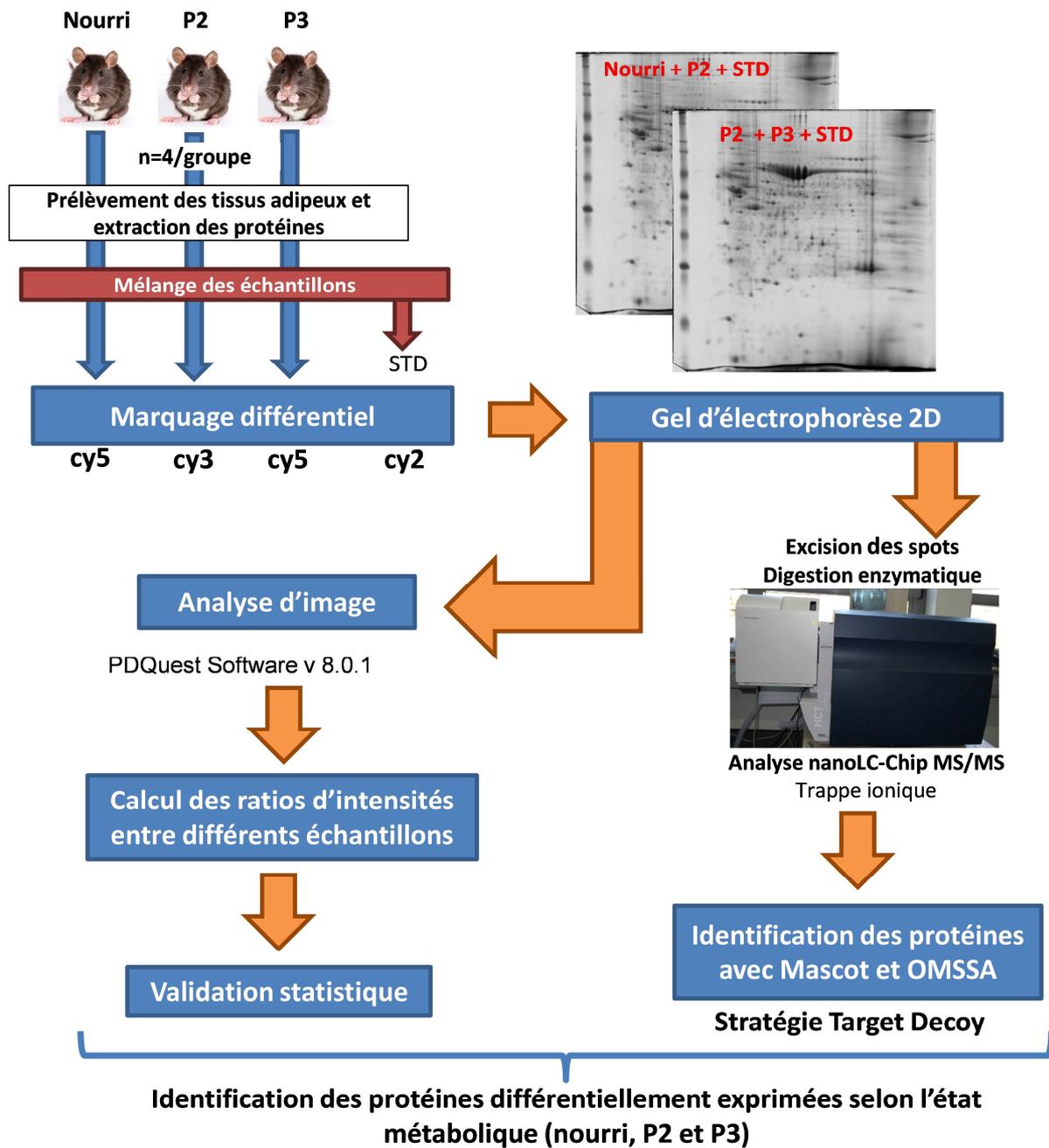


Figure 1 : Stratégie d'analyse différentielle par approche Gel 2D DiGE et nanoLC-MS/MS appliquée au tissu adipeux du rat soumis au jeûne prolongé.

## 7. Matériels et méthode

### 7.1. Traitement des animaux et prélèvements d'échantillons

Des rats mâles âgés de 8 semaines ont été maintenus dans des cages individuelles dans une salle à température contrôlée ( $25 \pm 2$  °C) et photopériode régulière (12 heures de lumière et 12 heures d'obscurité). Au cours de leur acclimatation, le régime alimentaire des rats était composé en % de masse de 50 % de carbohydrate, 5 % de matière grasse et 24 % de protéine. Les rats ont eu un accès *ad libitum* à l'eau de boisson tout au cours de l'expérience.

Après acclimatation, les rats (255 – 275 g) ont été aléatoirement répartis entre trois groupes (n = 4 / groupe). Quatre rats ont été immédiatement sacrifiés pour servir de contrôle en tant qu'animaux nourris. Les autres rats ont été soumis à un jeûne expérimental jusqu'à l'atteinte de la phase 2 (n=4) et de la phase 3 (n=4) du jeûne prolongé. Une pesée quotidienne des rats a permis d'identifier les phases métaboliques du jeûne. Le statut métabolique de tous les rats a été vérifié a posteriori par le dosage des métabolites et hormones plasmatiques.

Après dislocation cervicale et prélèvement du sang tronculaire, le tissu adipeux epididymaire (EPI) et le tissu adipeux sous cutané (SC) ont été disséqués, pesés, immédiatement congelés à l'aide d'azote liquide et conservés à – 80 °C jusqu'aux analyses. Le plasma a été préparé par centrifugation (3000 rpm, 4°C, 10 min.), puis congelé à -80 °C jusqu'aux analyses.

## 7.2. Extraction des protéines

Les échantillons de tissu adipeux congelés ont été broyés à l'aide d'un broyeur à froid et à sec de type Mikro-Dismembrator (Sartorius; une minute à 3000 tours/minutes). Environ dix milligrammes de chaque échantillon réduit en poudre ont été repris dans 400 µL d'une solution à base d'urée (8 M), thiourée (2 M), CHAPS 4 %, dithiothréitol (DTT) 1 %, Triton 10 %, TLCK (inhibiteur de trypsine) 0,05 % et inhibiteurs de protéases (cocktail commercial dilué 1 :50). Les échantillons ont été homogénéisés par agitation (vortex) et sonication (10 secondes, 135 Watts).

Les protéines ont été précipitées par ajout de 9 volumes d'acétone froid. Après une nuit à -20°C, les échantillons ont été centrifugés (14 minutes, +4 °C, 14000xg). Les surnageants ont été éliminés et les culots de protéines séchés par évaporation sous vide (Speedvac, Thermo Scientific). Chaque culot a été repris dans une solution à base d'urée (7M), thiourée (2M), Tris pH 8.5 (30 mM) et CHAPS 4 %. Le pH de chaque échantillon a été ajusté à pH 8,5 (tris 1,5 M pH 8,8). Après homogénéisation par agitation (vortex) et sonication (10 secondes, 135 watts), tout en maintenant les tubes dans la glace, les échantillons ont été centrifugés afin d'éliminer les éventuels débris cellulaires.

La concentration des protéines dans les échantillons a été mesurée selon la méthode de Bradford.

## 7.3. Marquage des protéines

Les cyanines Cy2, Cy3 et Cy5 (Cydyes, Amersham Biosciences, GE Healthcare) ont été reconstituées dans du diméthylformamide anhydre. Le marquage s'est fait par ajout de 400 picomoles de Cydyes à 50 µg de protéines (reprises dans 20 µL de solution à base d'urée (7 M), thiourée (2 M), CHAPS 4 %, et 30 mM Tris). Après centrifugation, les échantillons ont été incubés à l'obscurité pendant 30 minutes dans la glace. 1 µL de solution de lysine (10 mM) a alors été ajouté à chaque échantillon, puis les échantillons ont été incubés une seconde fois à l'obscurité pendant 10 minutes dans la glace.

Pour constituer un standard interne, des quantités équivalentes de protéines de chaque échantillon ont été mélangées. Le mélange a ensuite été repris dans 120 µL de solution à base d'urée (7 M), thiourée (2 M), CHAPS 4 %, et 30 mM Tris. Le marquage a ensuite été réalisé avec 2,4 ng de Cydyes selon le procédé énoncé précédemment.

## 7.4. Electrophorèse 2D

Deux échantillons marqués (soit 2x50 µg de protéines) ainsi que 50 µg de protéines du standard interne marqué également ont été mélangés avant la première étape d'isoélectrofocalisation (IEF). Le mélange de 150 µg de protéines a été repris dans un volume de 400 µL d'une solution d'urée (7 M), thiourée (2 M), 2 % CHAPS, 2 % DTT, 2 % Ampholytes et quelques gouttes de bleu de bromophénol. Le tout a été déposé sur une bande de gel à gradient de pH immobilisé (ReadyStrip IPG, Biorad) de 18cm pH 3-10 non linéaire. Chaque strip a été recouvert de 2mL d'huile végétale. La réhydratation passive a duré 2h30. La réhydratation active a ensuite été réalisée par l'application d'un courant de 50V pendant une nuit. Après réhydratation, un gradient de tension a été appliqué (0-200 V/1h linéaire ; 200-1000 V/4h linéaire ; 1000-5000 V/16h linéaire, 5000 V/7h). L'IEF a duré 28 heures et atteint un total de 85000 V/h (Protean ief cell, Biorad).

La réduction des ponts disulfures des protéines présentes dans les strips a été réalisée par équilibration du strip (30 minutes) dans une solution à 1% DTT, 6 M Urée, 50 mM Tris pH 8,8, 30 % glycerol, 2 % SDS (excédent molaire en dithiothréitol : >> 300). L'alkylation des cystéines réduites a été réalisée par équilibration du strip (30 minutes) dans une solution à 2,5 % iodoacétamide, 6M Urée, 50mM Tris pH 8,8, 30 % glycerol, 2 % SDS (excédent molaire en iodoacétamide : >>300). Le strip a ensuite été rincé à l'eau, puis dans une solution tampon (Tris 0.3%, Glycine 1.5%, SDS, 0,1 %, H<sub>2</sub>O). Les strips ont ensuite été déposés sur un gel en polyacrylamide réticulé à 10% de dimension 20 x 20 cm (0,15 M Tris pH 8,8, SDS 0,1 %, Temed 0,8 % et Persulfate NH<sub>4</sub> 0,1 %). Les conditions électrophorétiques pour cette deuxième dimension ont été fixées à 1 h à 5 mA par gel puis 8 h à 8 mA par gel (Protean II xi Cell, Biorad).

Un gel supplémentaire a été réalisé avec 1mg de protéines chargées (mélange en quantité équivalente des protéines non marquées de chaque échantillon). Ce gel, qui contenait une plus grande quantité de matériel, était destiné à l'analyse par spectrométrie de masse. Les deux dimensions ont été réalisées comme décrit précédemment. Le gel a ensuite été baigné dans une solution de fixation (50 :50 Ethanol/ Acide phosphorique) puis mis à colorer pendant quatre jours dans une solution de bleu de Coomassie.

## 7.5. Analyse des images de gels 2D

Les gels 2D-DIGE ont été rincés à l'eau milliQ, puis scannés (Ettan DIGE Imager, GE Healthcare). Trois images ont donc été générées pour chaque gel. Les images ont été traitées par le logiciel PDQuest (BioRad, V8 0.1). Ce logiciel permet la détection automatique des spots protéiques sur les images de gels, cependant cette étape doit être contrôlée manuellement. La correspondance des spots entre tous les gels doit être ensuite établie. Cette opération est réalisée manuellement. Le logiciel permet ensuite d'exporter les intensités de chacun des spots protéiques. L'intensité des spots protéiques issus de chaque échantillon a été normalisée par l'intensité des mêmes spots protéiques obtenues pour le standard interne.

Toutes les données normalisées ont été analysées à l'aide du logiciel « R » (v2.8.1 ; <http://www.R-project.org>). La distribution normale des variables mesurées a été vérifiée par le test de Shapiro-Wilk et l'égalité des variances par le test de Bartlett. Les variations entre les groupes ont été testées par ANOVA suivie du test post-hoc de Tukey. Le niveau de significativité a été établi à  $P > 0,01$  pour les tests de Shapiro-Wilk et de Bartlett et à  $P < 0,05$  pour les tests ANOVA et Tukey.

## 7.6. Analyse LC-MS/MS et identification des protéines

Les spots de gel sont tout d'abord décolorés, les ponts disulfures sont réduits (10 mM DTT, 45 mins, 60 °C) et les cystéines sont alkylées avec de l'iodoacétamide (30 mM). Les protéines sont ensuite digérées dans le gel avec la trypsine. Les peptides ont ensuite été extraits et analysés sur un système nanoLC Chip Cube (Agilent) couplé à un spectromètre de masse de type trappe d'ions (HCT Ultra, Bruker). La Chip contient une colonne 4,5 cm x 75 µm (C18 300 Å, 5 µM) et une précolonne (colonne de chargement) de 4 nL contenant la même phase. Les peptides sont chargés sur la colonne à 3,75 µL/min avec le solvant A (2 % acétonitrile dans 0,1 % acide formique). L'éluion est effectuée à 0,3 µL/min avec un gradient linéaire (8-40 %) du solvant B (98 % acétonitrile, 0,1 % acide formique, 2 % eau) dans le solvant A en 10 minutes, suivi d'un plateau à 70 % B pendant 2 minutes, puis 6 minutes de rééquilibrage à 92 % A. La durée totale de l'analyse est de 18 minutes. La trappe a été programmée pour alterner automatiquement les analyses MS et MS/MS. Les trois ions les plus abondants sur chaque spectre, en préférant les doublements chargés, sont sélectionnés et fragmentés en mode CID. Les ions sélectionnés sont ensuite exclus pendant 0,2 minutes.

L'identification des protéines à partir des données MS/MS a été obtenue à l'aide de deux moteurs de recherche différents, Mascot et OMSSA. La recherche a été effectuée dans une banque composite « Target-Decoy » (restriction taxonomique : *Rattus norvegicus*) contenant aussi les séquences des protéines contaminantes communes (trypsine et kératines). Une tolérance de 0,25 Da a été autorisée sur la masse du précurseur et des ions fragments. Un site de clivage manqué par la trypsine a été autorisé par peptide, ainsi que les modifications variables carbamidométylation des cystéines, acétylation des amines N-terminales des protéines et l'oxydation des méthionines.

Pour valider nos identifications nous avons appliqué la stratégie Target-Decoy en se fixant un taux de FDR maximal de 1 % à l'aide du logiciel Scaffold (Proteome Software).

## 8. Résultats et discussion

### 8.1. Validation du statut métabolique des animaux

La pesée des animaux a permis de montrer que les rats en P2 avaient une masse corporelle ~ 25 % inférieure à celle des rats nourris et que les rats en P3 avaient une masse corporelle ~ 8 % inférieure à celle des rats en P2. Ces diminutions de masse corporelle sont essentiellement dues à la diminution de la masse grasse, les lipides de réserves étant largement mobilisés pour satisfaire les besoins énergétiques de l'organisme.

	animaux nourris	animaux en P2	animaux en P3
<b>Masse corporelle (g)</b>	253.6 ± 5.8 <sup>a</sup>	190.3 ± 1.2 <sup>b</sup>	172.8 ± 0.4 <sup>c</sup>
<b>Perte de masse (%)</b>	-	24.8 ± 2.7 <sup>a</sup>	32.0 ± 1.0 <sup>b</sup>
<b>β-hydroxybutyrate (mM)</b>	0.32 ± 0.06 <sup>a</sup>	0.94 ± 0.04 <sup>b</sup>	0.18 ± 0.08 <sup>c</sup>
<b>Urée (mM)</b>	3.4 ± 0.2 <sup>a</sup>	4.4 ± 0.4 <sup>a</sup>	8.0 ± 0.5 <sup>b</sup>
<b>Cortisostérone (ng/mL)</b>	3.1 ± 0.6 <sup>a</sup>	273.5 ± 40.8 <sup>b</sup>	2102.6 ± 224.8 <sup>c</sup>

**Tableau 1: Profils des rats nourris et des rats en P2 et P3 du jeûne prolongé. Les valeurs significativement différentes (p<0.05) ne partagent pas la même lettre.**

La validation du statut métabolique des animaux a été obtenue par le dosage des niveaux circulants de β-hydroxybutyrate qui est un produit de dégradation des lipides, et de l'acide urique qui

est un produit de dégradation des protéines chez les oiseaux (Tableau 1). La diminution de la concentration du  $\beta$ -hydroxybutyrate entre les groupes P2 et P3 reflète donc la moindre utilisation des lipides en P3 par rapport à la P2 du jeûne. Les niveaux significativement plus élevés d'acide urique montre une augmentation de la mobilisation et de l'utilisation des protéines corporelles en P3 par rapport à la P2 du jeûne.

## 8.2. Protéines du tissu adipeux différenciellement exprimées en réponse au jeûne prolongé

Seuls les résultats de l'expérience à partir des échantillons du tissu adipeux épidydymaire (EPI) seront discutés ici, puisque les analyses pour le tissu sous cutané (SC) sont encore en cours.

Les analyses des images de gel 2D de l'expérience EPI ont permis de détecter près de 400 spots par gel. Parmi ceux-ci, l'analyse d'image a révélé 27 spots différentiels (Figure 2).

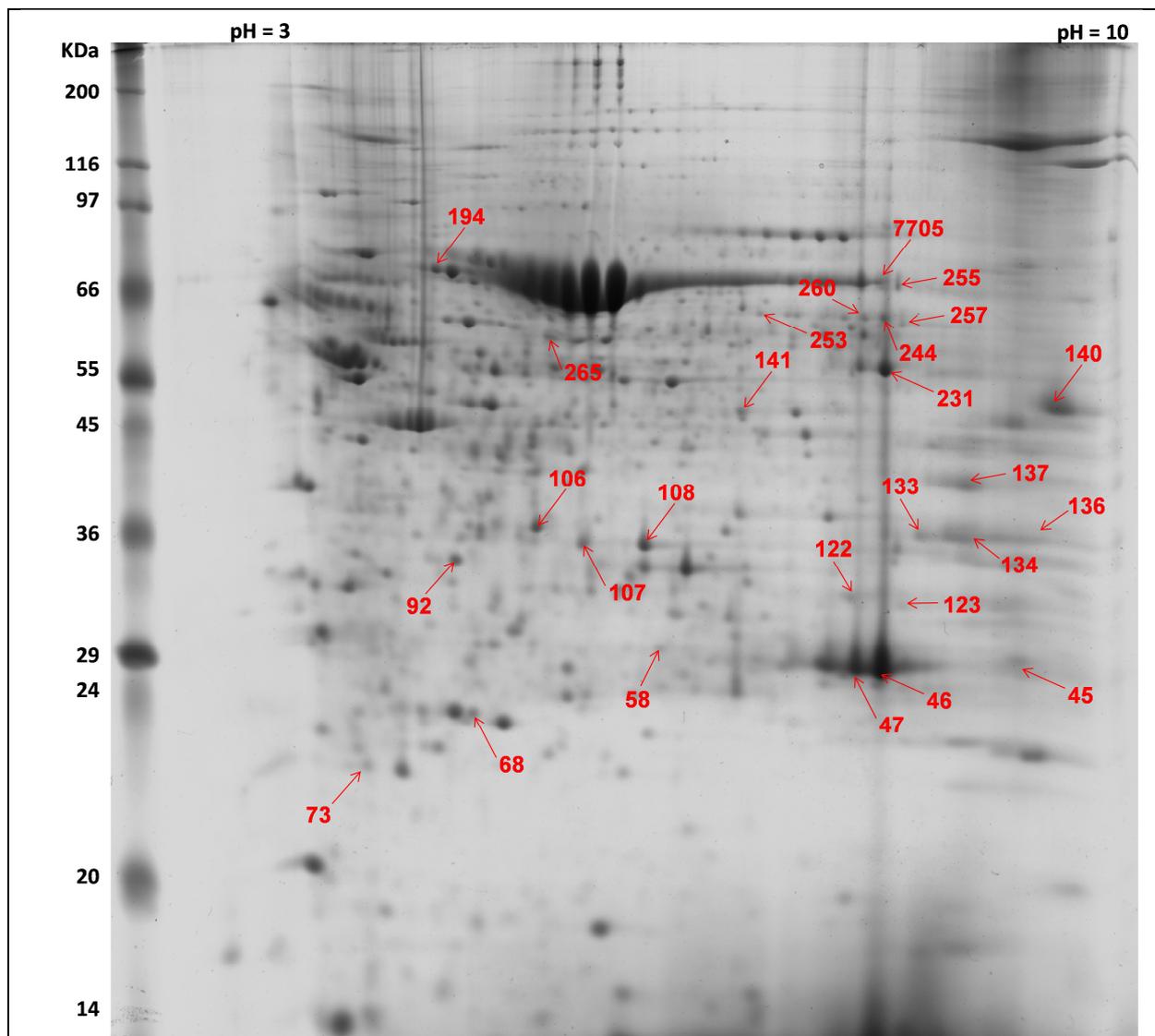


Figure 2 : Image représentative des spots excisés à partir des gels d'électrophorèse réalisés à partir d'échantillons de tissu adipeux de rat. Les nombres font référence aux numéros des spots que l'on retrouve dans le tableau 2 et indiquent les spots dont l'intensité était différentielle entre les groupes de rats.

Pour le tissu EPI, 288 spots ont été analysés par nanoLC-MS/MS. L'identification a été effectuée à l'aide de deux moteurs de recherche, Mascot et OMSSA. Cette utilisation combinée des deux moteurs de recherche permet de bénéficier de leurs résultats complémentaires et de renforcer la confiance dans les identifications. Elle s'inscrit dans les directives en cours d'élaboration pour la publication de données d'analyses protéomiques. Dans cette étude, nous avons établi les seuils de score de manière indépendante pour chacun des deux moteurs de recherche. Ces seuils de score ont été établis à l'aide de la stratégie « Target-Decoy » afin d'obtenir un taux de FDR pour les protéines inférieur à 1 %.

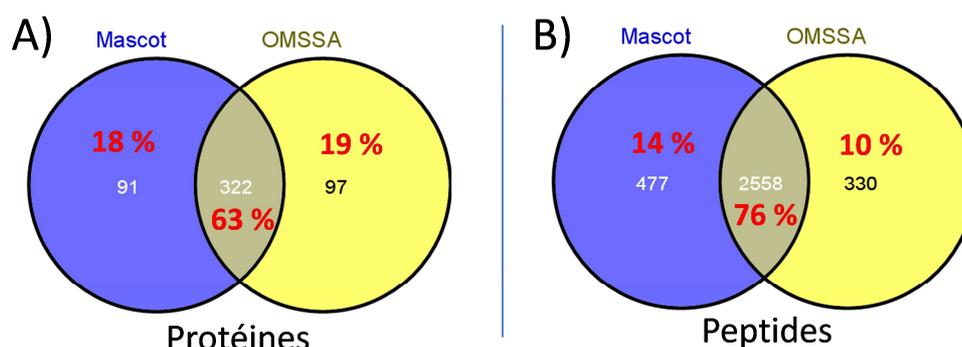
Scores seuils Mascot sur les identifications des peptides :

- Au moins deux peptides uniques identifiés dans le spot de gel participant à l'identification de la protéine : score >25 et > score identité – 15 ;
- Un seul peptide unique dans le spot de gel participe à l'identification : score > 25 et > score identité + 9.

Scores seuils OMSSA sur les identifications des peptides :

- Au moins deux peptides uniques identifiés dans le spot de gel participant à l'identification de la protéine :  $-\log(\text{E-value}) > 0,05$  ;
- Un seul peptide unique dans le spot de gel participe à l'identification :  $-\log(\text{E-value}) > 3,15$ .

Nos résultats ont permis l'identification de 510 protéines uniques avec un taux de FDR inférieur à 0,5 % pour les protéines et inférieur à 0,06 % pour les peptides. 63 % (322) des protéines uniques identifiées et 76 % (2558) des peptides l'ont été de façon commune par les deux algorithmes de recherche utilisés, ce qui renforce la confiance dans ces identifications. 18 % (91) des protéines et 14 % (477) des peptides ont été spécifiquement identifiées par l'algorithme de recherche Mascot, et 19 % (97) des protéines et 10 % (330) de peptides par OMSSA (Figure 3), ce qui illustre la complémentarité de ces deux algorithmes.



**Figure 3 : Nombre total des protéines (A) et des peptides (B) identifiés par chacun des moteurs de recherche Mascot et OMSSA.**

Pour les spots différentiels, l'analyse nanoLC-MS/MS a permis l'identification de 23 protéines uniques contenues dans les 27 spots différentiels (Tableau 2). Toutes ces protéines ont été identifiées par les deux moteurs de recherche. La cohérence entre les propriétés physico-chimiques théoriques (masse moléculaire et point isoélectrique) des protéines identifiées et leur position sur le gel d'électrophorèse 2D a été vérifiée.

Analyses des protéines différentielles						P2 vs. fed		P3 vs. Fed		P3 vs. P2	
# spot	Nom de la protéine	N° NCBI	PM	pi	Procédé	Régulation	Amplitude	Régulation	Amplitude	Régulation	Amplitude
<b>Métabolisme des carbohydate</b>											
92	Pyruvate dehydrogenase E1 component subunit beta, mitochondrial	gi   122065728	38964,4	5,29	Glycolyse	↘	5,11	↘	4,26	↗	1,20
106	L-lactate dehydrogenase B chain	gi   1170739	36594,5	5,7	Glycolyse	↘	1,87	↘	2,46	↘	1,32
107	Glycerol-3-phosphate dehydrogenase [NAD+], cytoplasmic	gi   118572652	37434,4	6,17	Néoglucogenèse	↘	2,02	↘	4,53	↘	2,25
108	Glycerol-3-phosphate dehydrogenase [NAD+], cytoplasmic	gi   118572652	37434,4	6,17	Néoglucogenèse	↘	3,46	↘	4,74	↘	1,37
133	glyceraldehyde 3-phosphate-dehydrogenase	gi   122065190	35810,1	8,17	Glycolyse	↘	2,48	↘	10,53	↘	4,25
134	L-lactate dehydrogenase A chain	gi   126051	36432,8	8,47	Glycolyse	↘	2,03	↘	3,88	↘	1,91
134	glyceraldehyde 3-phosphate-dehydrogenase	gi   122065190	35810,1	8,17	Glycolyse	↘	2,03	↘	3,88	↘	1,91
136	L-lactate dehydrogenase A chain	gi   126051	36432,8	8,47	Glycolyse	↘	3,26	↘	2,40	↗	1,36
136	Fructose-bisphosphate aldolase A	gi   113609	39334,5	8,4	Glycolyse	↘	3,26	↘	2,40	↗	1,36
137	Fructose-bisphosphate aldolase A	gi   113609	39334,5	8,4	Glycolyse	↘	3,06	↘	5,42	↘	1,77
255	Transketolase	gi   12018252	67626,7	7,22	Voie des pentoses phosphate	↘	3,90	↘	5,52	↘	1,42
257	M2 pyruvate kinase-like	gi   206205	57763,6	X7,57	Glycolyse	↘	2,09	↘	4,74	↘	2,27
7705	Transketolase	gi   12018252	67626,7	7,22	Voie des pentoses phosphate	↘	2,93	↘	4,96	↘	1,69
<b>Métabolisme des lipides</b>											
45	3,2-trans-enoyl-CoA isomerase, mitochondrial precursor	gi   118210	32237,8	9,55	oxydation des acides gras	↘	1,56	↘	1,81	↘	1,16
68	Apolipoprotein A-I	gi   146345369	30070,4	5,51	composant des HDL	↗	1,70	↗	1,13	↘	1,51
244	perilipin, isoform CRA_b	gi   1172433	55595,7	6,55	lipolyse	↘	1,97	↘	2,46	↘	1,25
253	perilipin, isoform CRA_b	gi   1172433	55595,7	6,55	lipolyse	↘	2,36	↘	2,66	↘	1,13
257	perilipin, isoform CRA_b	gi   1172433	55595,7	6,55	lipolyse	↘	2,09	↘	4,74	↘	2,27
260	perilipin, isoform CRA_b	gi   1172433	55595,7	6,55	lipolyse	↘	1,96	↘	2,55	↘	1,30
265	Fatty acid synthase	gi   204099	272633,1	5,96	lipogenèse	↘	1,92	↘	3,03	↘	1,58
<b>Métabolisme des protéines</b>											
107	Lys-63-specific deubiquitinase BRCC36	gi   229621690	32996,5	5,83	Proteolyse / réparation de l'ADN	↘	2,02	↘	4,53	↘	2,25
140	Elongation factor 1-alpha 1	gi   1220484	50096,2	9,1	traduction des protéines	↘	1,89	↘	2,72	↘	1,44
141	Elongation factor Tu, mitochondrial	gi   149067905	49504,8	6,2	traduction des protéines	↘	1,31	↘	1,74	↘	1,33
<b>Response au stress oxydatif</b>											
45	Carbonic anhydrase 3	gi   5921194	29413,8	6,97	réponse au stress oxydatif	↘	1,56	↘	1,81	↘	1,16
45	glutathione S-transferase Yc1 subunit	gi   1170084	25302,4	8,83	réponse au stress oxydatif	↘	1,56	↘	1,81	↘	1,16
46	Carbonic anhydrase 3	gi   5921194	29413,8	6,97	réponse au stress oxydatif	↘	2,17	↘	2,93	↘	1,35
47	Carbonic anhydrase 3	gi   5921194	29413,8	6,97	réponse au stress oxydatif	↘	2,38	↘	4,05	↘	1,70
73	Peroxiredoxin 2	gi   149037813	21780,1	5,34	réponse au stress oxydatif	↗	3,03	↗	3,51	↗	1,16
122	Carbonic anhydrase 3	gi   5921194	29413,8	6,97	réponse au stress oxydatif	↘	2,86	↘	3,38	↘	1,18
123	Carbonic anhydrase 3	gi   5921194	29413,8	6,97	réponse au stress oxydatif	↘	2,41	↘	4,29	↘	1,78
231	Carbonic anhydrase 3	gi   5921194	29413,8	6,97	réponse au stress oxydatif	↘	1,90	↘	2,81	↘	1,48
244	Catalase	gi   115707	59739,6	7,15	réponse au stress oxydatif	↘	1,97	↘	2,46	↘	1,25
253	Catalase	gi   115707	59739,6	7,15	réponse au stress oxydatif	↘	2,36	↘	2,66	↘	1,13
257	Catalase	gi   115707	59739,6	7,15	réponse au stress oxydatif	↘	2,09	↘	4,74	↘	2,27
260	Catalase	gi   115707	59739,6	7,15	réponse au stress oxydatif	↘	1,96	↘	2,55	↘	1,30
<b>Repliement des protéines</b>											
58	Endoplasmic reticulum resident protein 29	gi   2507015	28558,3	6,08	repliement des protéines	↘	2,01	↘	3,13	↘	1,56
194	Heat shock cognate 71 kDa protein	gi   51702273	70911,8	5,37	repliement des protéines	↘	1,64	↘	2,41	↘	1,47
265	protein disulfide isomerase associated 3	gi   1352384	56607,5	5,88	repliement des protéines	↘	1,92	↘	3,03	↘	1,58

**Tableau 2 : Liste des protéines identifiées dans les spots protéiques différentiels et variations de leur abondance relative selon les groupes. Les valeurs en jaune sont les valeurs significatives (p value < 0,05).**

Les protéines identifiées ont été regroupées suivant les processus biologiques dans lesquels elles sont impliquées. Ceci a été réalisé par l'extraction des annotations fonctionnelles contenues dans la base de données Gene Ontology ([geneontology.org](http://geneontology.org)). Cette extraction a été automatisée au laboratoire par des développements bioinformatiques implémentés et disponibles sur une interface web ([msda.unistra.fr](http://msda.unistra.fr)).

Nous avons ainsi pu déterminer que les protéines différentielles sont connues pour être impliquées dans le métabolisme des sucres, celui des lipides et celui des protéines, mais aussi dans la réponse au stress oxydatif et enfin dans le repliement des protéines (Tableau 2).

Les niveaux des enzymes impliquées dans le métabolisme des sucres étaient essentiellement diminués par le jeûne par rapport à l'état nourri. En particulier, la répression des enzymes de la glycolyse et de la néoglucogenèse tendait à être accentuée en phase 3 par rapport à la phase 2, en termes de nombre de protéines réprimées et/ou d'amplitude des variations. Cependant, les différences observées entre phase 3 et phase 2 n'étaient pas significatives. Ces diminutions sont en accord avec le modèle du jeûne prolongé, puisque le substrat énergétique majeur n'est plus le glucose dès la phase 2 [134]. En effet, à ce stade ce sont les substrats lipidiques qui prennent le relais, ce qui requiert une forte lipolyse, qui sera ensuite diminuée en phase 3 au profit de la mobilisation des protéines corporelles. Ainsi, la répression des enzymes de la lipolyse en phase 3 et pas en phase 2 est tout à fait cohérente avec les adaptations métaboliques au jeûne prolongé. Le fait qu'on n'observe pas d'augmentation en phase 2, alors qu'on aurait pu s'y attendre, reflète le fait que les animaux de notre groupe P2 étaient plutôt en fin de phase 2. Ces résultats sont en accord avec des travaux précédents [143]. Les niveaux de l'apolipoprotéine AI étaient augmentés en phase 2, mais normalisés en phase 3 en comparaison de l'état nourri. Le même profil de régulation a déjà été retrouvé pour ce qui est des niveaux d'ARNm hépatiques (site principal de production) de l'apolipoprotéine AI. Cette protéine est un des composants essentiels des HDL (High Density Lipoprotein), qui sont impliquées dans l'efflux de cholestérol vers le foie. Outre des altérations du métabolisme du cholestérol au cours du jeûne, cela pourrait suggérer que le cholestérol contenu dans les membranes du tissu adipeux est mobilisé en phase 2. Enfin, la diminution des niveaux des protéines impliquées dans les processus de traduction des protéines reflètent la diminution globale des synthèses protéiques en P3.

Les altérations métaboliques qui interviennent au cours du jeûne peuvent favoriser la production d'espèces réactives de l'oxygène (ROS) [144]. Les principaux effets délétères des ROS peuvent intervenir sur les lipides, les protéines et l'ADN. Les défenses antioxydantes sont notamment le fait de nombreuses enzymes antioxydantes. Les niveaux de certaines des enzymes antioxydantes que nous avons identifiées dans le tissu adipeux des rats étaient diminués dès la phase 2 du jeûne. On remarque cependant pour plusieurs de ces enzymes qu'elles ont été identifiées dans des spots protéiques distincts sur les gels. Ces spots distincts contiennent probablement des isoformes de ces enzymes, qui ne sont pas toutes altérées dès la phase 2. En revanche, la répression de l'expression de toutes ces isoformes semblent se généraliser en phase 3, de même que les niveaux des autres enzymes antioxydantes étaient aussi réprimés à ce stade. La seule exception concerne la peroxiredoxine 2 dont les niveaux sont augmentés significativement en P3. On remarquera également que la Lys-63-specific deubiquitinease BRCC36 qui est impliquée dans la réparation de l'ADN voit ses niveaux diminués en phase 2 comme en phase 3. Il restera à vérifier si le tissu adipeux des animaux privés de nourriture est effectivement en état de stress oxydatif, en mesurant par

exemple les produits de la peroxydation lipidique, de l'oxydation des protéines ou de la dégradation de l'ADN.

Enfin l'expression de certaines protéines impliquées dans le repliement correct des protéines était diminuée de manière significative en phase 3, mais pas en phase 2. A ce stade, l'accumulation possible de protéines non fonctionnelles (mauvais repliement) pourrait se révéler toxique pour les cellules.

Tous ces résultats suggèrent une accentuation des variations en phase 3 par rapport à la phase 2. Les résultats les plus pertinents seront prochainement confirmés par une autre méthode (Western-blot).

## 9. Conclusion

Notre étude du tissu adipeux nous a permis d'observer des différences dans l'expression de certaines protéines au cours des trois phases du jeûne dans le tissu adipeux, jetant les bases des régulations qui interviennent dans ces conditions au niveau moléculaire. Globalement, certaines des étapes des voies métaboliques qui sont régulées par le jeûne prolongée ont été identifiées. Par ailleurs, cette étude révèle qu'en phase 3 les cellules adipeuses pourraient être incapables de répondre à un stress oxydatif marqué.

Du point de vue analytique, l'approche gel 2D DiGE a permis d'analyser et de comparer plusieurs protéomes avec une très bonne sensibilité (150 µg de protéines déposées par gel) et ceci avec une bonne reproductibilité due à la normalisation par le standard interne. Si la technique 2D-DIGE est très sensible, l'analyse par nanoLC-MS/MS des spots de Gel DiGE ne permet pas d'obtenir des identifications renforcées par un nombre important de peptides. Nous avons donc réalisé un gel 2D classique coloré au bleu de Coomassie et sur lequel nous avons déposé une quantité plus importante de protéines (1 mg) afin d'obtenir des identifications avec un nombre plus important de peptides. La correspondance entre les spots sur ce gel coloré au bleu et les gels DiGE est effectuée manuellement. Des analyses par nanoLC-MS/MS ont été effectuées sur certains spots issus des gels DiGE et du gel bleu afin de vérifier cette correspondance.

L'approche d'identification utilisant deux algorithmes de recherche différents a, encore une fois, montré la complémentarité des deux algorithmes et permis de renforcer la confiance dans les résultats d'identification.

En perspective, des analyses par Western blot seront effectuées afin de valider par une méthode orthogonale nos résultats d'analyse différentielle. De plus, l'analyse des échantillons de tissus sous-cutanés est en cours de réalisation. Elle permettra de mieux comprendre les différences physiologiques qui existent entre les diverses localisations anatomiques du tissu adipeux.

## Chapitre 4 Application de la stratégie qN-TOP à l'étude du protéome du foie du rat soumis au jeûne prolongé

### 1. Résumé

Après avoir validé notre stratégie qN-TOP sur des protéines modèles, nous avons appliqué cette stratégie à un échantillon complexe afin de l'évaluer dans le cadre d'une étude différentielle des processus protéolytiques. Nous avons choisi comme modèle le protéome hépatique du rat soumis au jeûne prolongé. Notre étude a permis d'identifier les problèmes de notre stratégie et les améliorations à lui apporter. Nous avons ainsi pu proposer des solutions de quantification basées sur des analyses en deux temps pour dissocier l'étape de quantification de l'étape d'identification dans le but d'augmenter sa robustesse.

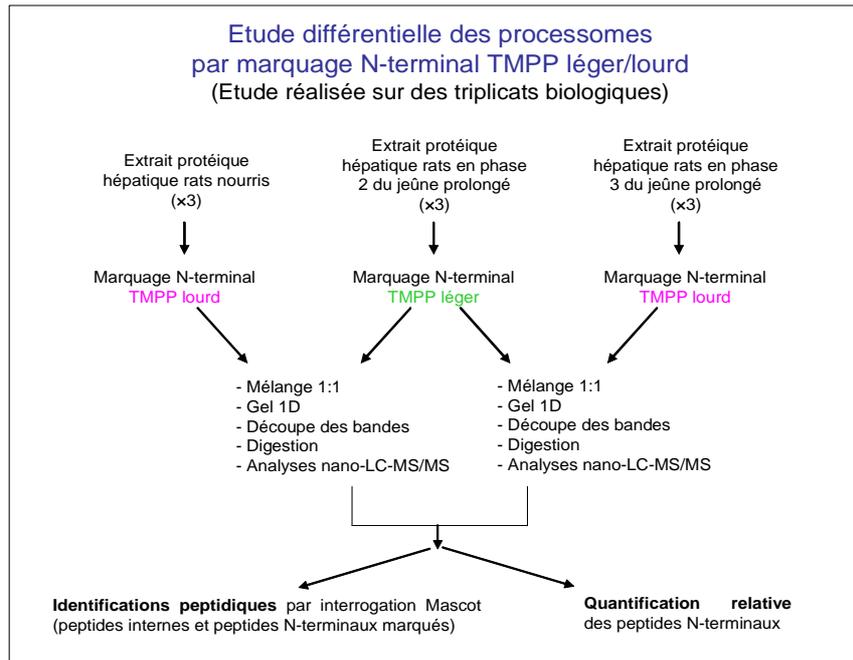
### 2. Contexte de l'étude

Nous avons vu, dans le chapitre 2 de la partie II, le développement de la stratégie qN-TOP. Cette stratégie, qui est basée sur le marquage chimique spécifique des extrémités N-terminales des protéines, permet une étude différentielle des processus protéolytiques grâce à l'utilisation de réactifs chimiques marqués aux isotopes stables. Afin d'évaluer l'efficacité de cette stratégie nous l'avons appliquée à une étude du protéome hépatique chez le rat soumis au jeûne prolongé. Nous avons vu dans le chapitre précédent l'intérêt de ce modèle et les trois phases P1, P2 et P3 du jeûne correspondant à la mobilisation séquentielle des substrats énergétiques. Nous nous sommes intéressés au protéome hépatique puisque le foie, un organe très riche en protéines, est un carrefour métabolique essentiel. Le protéome hépatique est donc le plus susceptible de subir des processus protéolytiques en réponse au jeûne expérimental. Ces processus protéolytiques génèrent de nouvelles extrémités N-terminales pour les protéines subissant ces processus. Le protéome hépatique est donc un bon exemple de protéome particulièrement complexe pour une étude de quantification relative en dégradomique. La stratégie qN-TOP facilite d'une part l'identification et la caractérisation des extrémités N-terminales et permet d'autre part la quantification des peptides N-terminaux ouvrant la voie à la réalisation d'études comparatives des processus protéolytiques entre différentes conditions biologiques telles que les différentes phases métaboliques du jeûne (voir chapitre 2 de la partie II).

Dans ce chapitre, nous discuterons les difficultés rencontrées dans la réalisation de cette étude et surtout dans le traitement des données générées, ainsi que les solutions et les résultats obtenus.

### 3. Stratégie analytique

Après le traitement des animaux, le prélèvement des foies et l'extraction des protéines, cette étude reposait sur le marquage différentiel des extrémités N-terminales des protéines selon la stratégie qN-TOP (Figure 1).



**Figure 1 : Stratégie d'analyse différentielles des clivages protéolytiques subit par les protéines du foie du rat dénutri par marquage au TMPP léger/lourd**

#### 3.1. Préparation des échantillons

Neuf rats mâles âgés de 8 semaines sont maintenus dans des cages individuelles dans une salle climatisée (25 °C) avec des photopériodes constantes de 12 heures de lumière et 12 heures d'obscurité. Les rats sont nourris à l'aide d'une diète de laboratoire composé en pourcentage de masse de 50 % de carbohydrate, 5 % de matière grasse et 24 % de protéine. Les rats ont eu un accès *ad libitum* à l'eau de boisson tout au cours de l'expérience

Les rats (255 – 275 g) sont soit sacrifiés directement (n = 3) pour servir de control en tant qu'animaux nourris, soit sacrifiés après un jeûne expérimental jusqu'à l'atteinte de la phase 2 (n = 3) et la phase 3 (n = 3). La détermination des phases métaboliques est effectuée à l'aide de la pesée journalière des rats. Après dislocation cervicale et décapitation, le foie est disséqué, pesé et congelé à l'aide d'azote liquide avant d'être conservés à -80 °C. 300 mg de tissu hépatique pour chaque prélèvement (3 nourris, 3 en phase 2 et 3 en phase 3) sont ensuite broyés dans l'azote liquide et les protéines sont extraites dans 3,5 mL de solution tampon de marquage au TMPP-Ac-OSu après ajout d'un cocktail d'inhibiteurs de protéases. Les concentrations en protéines dans chaque extrait sont estimées à l'aide de la méthode de Bradford.

### 3.2. Marquage différentielle et analyse par nanoLC-MS/MS

Pour chaque préparation (3 rats nourris, 3 rats en P2 et 3 en P3), 80 µg de protéines sont marqués selon le protocole qN-TOP (Partie II, Chapitre 2) au TMPP-Ac-OSu lourd ou léger après réduction et alkylation des cystéines. Les prélèvements issus des rats nourris et des rats en phase 3 sont marqués au TMPP-Ac-OSu léger. Les prélèvements issus des rats en phase 2 sont marqués au TMPP-Ac-OSu lourd. Les échantillons sont ensuite mélangés en quantités égales de la façon suivante : Nourri marqué léger avec P2 marqué lourd et P2 marqué lourd avec P3 marqué léger. Les échantillons sont ensuite déposés sur un gel 1D et après migration, chaque piste de gel est découpée de façon systématique en 35 bandes de 2 mm chacune à l'aide d'un emporte-pièce.

Après réduction alkylation et digestion à la trypsine des protéines présentes dans les bandes de gel, les peptides générés par la digestion tryptique sont analysés par nanoLC-MS/MS sur un système nanoLC Chip-Cube (Agilent) couplé à un spectromètre de masse de type trappe d'ions (Amazon, Bruker).

### 3.3. Identification des protéines et caractérisation des peptides N-terminaux

Pour l'identification des peptides internes, la recherche avec le moteur de recherche Mascot est effectuée à l'aide d'une banque composite « Target-Decoy » contenant les séquences des protéines de la taxonomie *Rattus norvegicus* téléchargées du site du NCBI. La recherche est effectuée avec une tolérance de 0,25 Da sur la masse des ions précurseurs et sur celle des ions fragments. Un site de clivage manqué par la trypsine est autorisé par peptide et la carbamidométhylation des résidus cystéines est recherchée en tant que modification obligatoire, tandis que l'oxydation de la méthionine est recherchée en tant que modification variable. Les seuils de score pour les identifications des peptides sont établis de façon à avoir un FDR inférieur à 2 %.

Pour l'identification des peptides N-terminaux, nous avons tout d'abord généré une sous-banque contenant les séquences des protéines identifiées dans la première recherche qui a servi à l'identification des peptides internes. Les recherches sont réalisées avec une enzyme semi-spécifique, c'est-à-dire les peptides de digestion *in silico* peuvent avoir une seule extrémité qui correspond à un site de clivage de la trypsine. L'oxydation des méthionines, l'acétylation N-terminale des protéines et le marquage au TMPP léger ou lourd sur l'amine N-terminale des peptides et les résidus lysine et tyrosine sont recherchés en tant que modifications variables. Les spectres générant une identification d'un peptide marqué au TMPP sont réinspectés individuellement.

Le traitement des données d'identification et l'établissement des seuils de score sont réalisés à l'aide du logiciel Scaffold (Protéome Software). Ce logiciel permet d'intégrer les résultats de plusieurs moteurs de recherche et d'établir les seuils de score en fonction du moteur de recherche utilisé. Pour Mascot, il permet d'établir un seuil de score de corrélation (« ion score ») en fonction de l'état de charge de l'ion précurseur, un seuil de différence entre l'« ion score » et le score probabiliste (« identity score »), ainsi qu'un seuil de probabilité local calculé d'après les approches empiriques de Bayes. Ces seuils sont établis de façon à établir un FDR évalué à moins de 1 % pour les protéines.

## 4. Résultats et discussion

Dans cette étude différentielle des processus protéolytiques du protéome hépatique du rat soumis au jeûne prolongé, nous avons pu appliquer la stratégie qN-TOP à des échantillons biologiques complexes. Cette étude nous permet d'évaluer les capacités de cette stratégie, et de l'affiner en détectant les problèmes qui peuvent être rencontrés dans ce type d'étude et en leur trouvant des solutions. Nous avons vu dans le chapitre 2 de la partie II le problème de gamme dynamique que nous avons rencontré pour la quantification qN-TOP des protéines modèles à l'aide de l'instrument Q-TOF de Waters (Synapt HDMS). Nos expériences ont montré que le Synapt HDMS présente une gamme dynamique linéaire de 6, c'est-à-dire que lorsqu'un ion est plus que 6 fois plus intense que son homologue, la réponse n'est plus linéaire. Au moment de la réalisation de cette étude, nous disposions en plus du Q-TOF Synapt HDMS de Waters, d'un instrument de type trappe d'ions (Amazon, Bruker). La quantité d'ions dans la trappe peut être contrôlée et par suite la trappe ne sature pas et présente une gamme dynamique très étendue [145]. C'est pour cette raison que nous avons effectué nos analyses sur la trappe d'ions.

### 4.1. Identification des protéines et des processus protéolytiques

La recherche à l'aide du moteur de recherche Mascot dans la banque de données NCBI restreinte à la taxonomie *Rattus Norvegicus* et constituée en moitié de séquences « Decoy » a permis d'identifier 1750 protéines uniques avec un FDR évalué à 1,5 % à l'aide de la stratégie « Target-Decoy ». Nous avons appliqué deux seuils différents, un seuil pour les protéines identifiées avec un seul peptide unique et un autre seuil, moins élevé, pour les protéines identifiées avec deux peptides uniques et plus.

La recherche dans la sous banque contenant les séquences des 1750 protéines identifiées dans la première recherche a permis d'identifier sans ambiguïté 381 extrémités N-terminales uniques marquées au TMPP sous sa forme lourde ou légère. Parmi ces 381 extrémités N-terminales, 226 correspondent à des protéines n'ayant pas subi de processus protéolytiques. Les 155 extrémités N-terminales restantes, correspondent à des extrémités issues d'un clivage protéolytique.

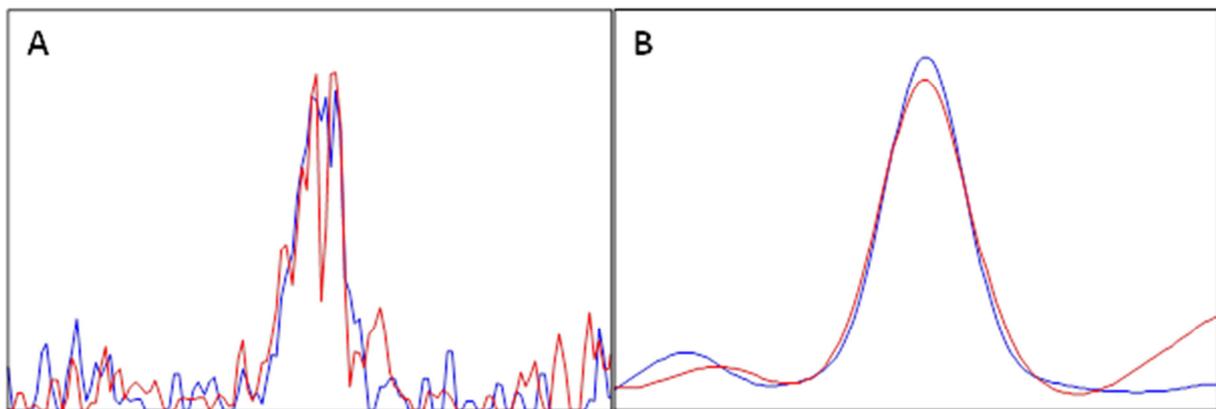
Sur les 381 extrémités N-terminales identifiées, 19 % ont été identifiées à la fois sous la forme marquée avec le TMPP léger et sous la forme marquée avec le TMPP lourd. 46 % sont identifiées uniquement sous la forme marquée avec le TMPP léger et 35 % sont identifiées uniquement sous la forme marquée avec le TMPP lourd. Cette distribution indique qu'il ne semble pas y avoir de biais sur les identifications en fonction du type de marquage, puisque les nombres d'identification sous forme marquées au TMPP léger et lourd sont comparables (243 et 213 respectivement).

### 4.2. Quantification des peptides N-terminaux

Après cette étape d'identification, il faut quantifier les peptides N-terminaux marqués au TMPP léger/lourd. Ceci est effectué par l'extraction des courants d'ions des peptides marqués et le calcul des ratios des aires de leurs pics chromatographiques. Cette étape nécessite l'utilisation d'un logiciel capable d'identifier les couples de peptides léger/lourd et de calculer leurs ratios d'abondance. Nous avons utilisé pour ceci le logiciel Proteinscape (Bruker). Ce logiciel permet, entre autres, de quantifier les protéines analysées par une approche SILE (Stable Isotope Labelling

Experiment). Cependant nous avons pu observer plusieurs problèmes de traitement de signal, ainsi que des problèmes liés à notre choix du spectromètre de masse et à notre stratégie d'analyse nanoLC-MS/MS. Nous n'avons pas pu donc quantifier de façon satisfaisante les changements protéolytiques entre les différents échantillons (foie de rat nourri et en phase 2 et 3 du jeûne).

Le premier problème que nous avons constaté est que nous avons pour un peptide donné, plusieurs ratios calculés. Ceci est dû au fait que le spectromètre de masse effectue des cycles constitués d'un balayage MS suivi de six balayages MS/MS. Or l'extraction des courants d'ions s'effectue uniquement avec les données MS ; ceci conduit à l'obtention de pics chromatographiques qui ne sont pas lisses (Figure 2-A). La solution à ce problème est le lissage des chromatogrammes à l'aide d'un algorithme de lissage Gaussien proposé par le logiciel Proteinscape. Nous n'avons pas pu cependant évaluer si ce lissage induisait des biais dans la quantification. Le résultat d'un lissage est visible sur la figure 2-B.



**Figure 2 : Chromatogramme d'extrait du courant d'ions d'un couple léger/lourd d'un peptide sans lissage (A) et avec lissage à l'aide d'un algorithme de lissage Gaussien (B).**

Il a donc été nécessaire de réeffectuer la quantification en utilisant l'algorithme de lissage des chromatogrammes malgré le risque que ce lissage induise un biais dans le calcul des aires des pics.

Nous avons aussi constaté que les tolérances autorisées sur la masse et le temps de rétention des ions, lors de l'extraction des courants d'ions, ont un effet important sur les résultats de la quantification. La tolérance sur la masse dépend de la précision de la mesure de masse du spectromètre de masse utilisé. Or dans notre étude nous avons utilisé un spectromètre de type trappe d'ions. Si ce spectromètre de masse présente une bonne vitesse de balayage et une gamme dynamique étendue, il possède néanmoins une faible résolution (6000 au maximum) et par suite une faible précision de mesure ( $\pm 0,25$  Da). Il est donc nécessaire d'autoriser une tolérance de 0,25 Da pour l'extraction des courants d'ions ce qui conduit à avoir des interférences sur les pics chromatographiques issus de ces extractions et faussant ainsi la quantification (figure 3). Ceci montre l'importance d'effectuer les analyses sur un spectromètre de masse de haute résolution tel qu'un LTQ-Orbitrap ou un Q-TOF récent. Ces instruments présentent des précisions de mesure de masse inférieures à 5 ppm permettant ainsi de diminuer considérablement la tolérance sur la masse des ions et donc de limiter ce problème d'interférence d'une façon significative.

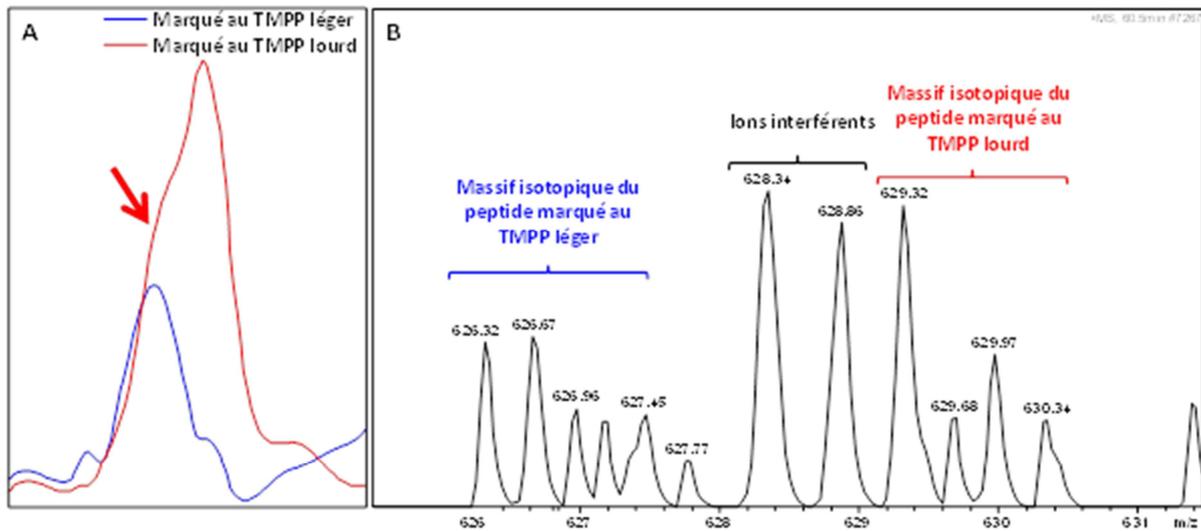


Figure 3 : A) Chromatogrammes d'extraction de courant d'ions d'un peptide marqué au TMPP léger (bleu, 626,3 m/z) et lourd (rouge, 636,3 m/z) montrant une importante interférence dans le courant de l'ion lourd (épaulement). B) Spectre MS correspondant au point indiqué par la flèche en (A). On observe que le massif isotopique de l'ion 628,3 interfère avec celui de l'ion 629,32.

La tolérance sur le temps de rétention dépend de la résolution chromatographique. Il est évident que plus la résolution chromatographique est élevée, moins il y aura de risque de coélution d'espèces isobares ou de masse proche qui interfèrent avec les ions d'intérêt. Cependant, il est nécessaire d'avoir assez de points, donc de spectres MS, pour décrire le pic chromatographique (au moins 9 [145]). Or généralement, dans une expérience de type SILE, l'identification et la quantification se font dans la même analyse LC-MS/MS (voir paragraphe 4.1 du chapitre 3 de la partie I). Le spectromètre de masse enchaîne des cycles composés d'un balayage MS et de plusieurs balayages MS/MS (Figure 4).

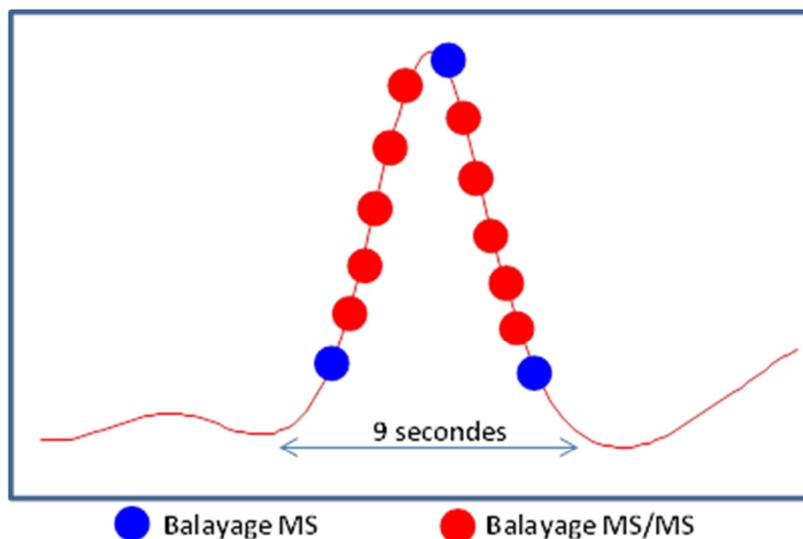
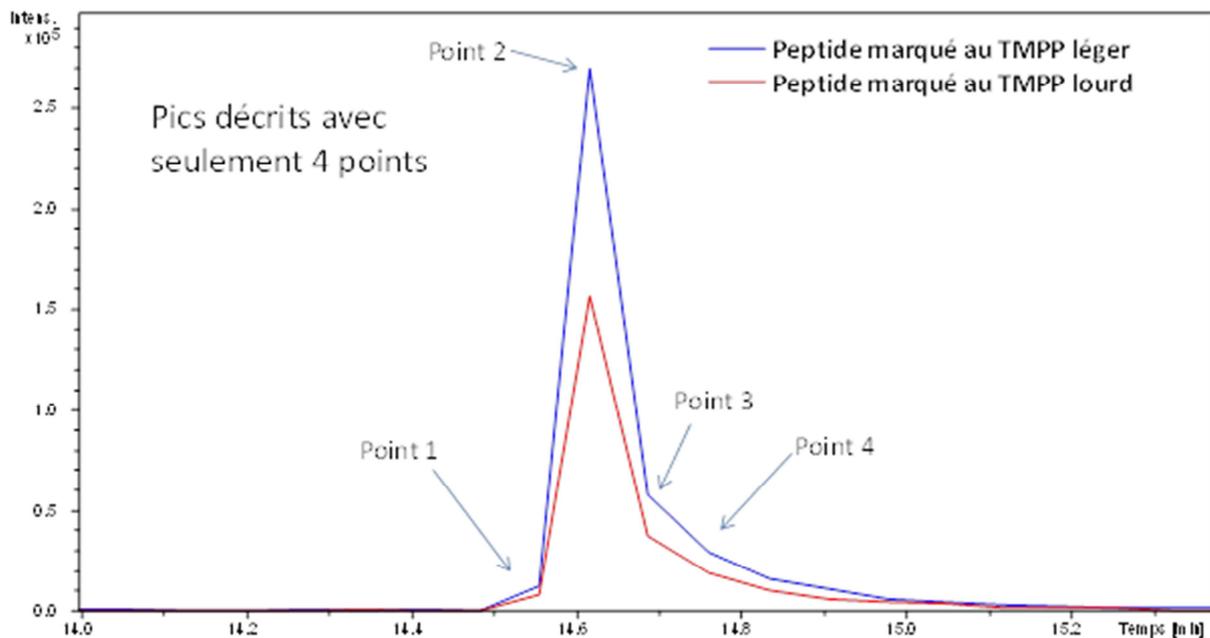


Figure 4 : enchaînement des cycles de balayage MS et MS/MS au cours d'un pic chromatographique d'une largeur de 9 secondes à la base. Seulement trois spectres MS sont réalisés durant le temps de ce pic.

Les spectres MS/MS permettent l'identification des peptides tandis que la quantification est effectuée en extrayant le courant d'ions de l'ion précurseur, et donc les spectres MS. Dépendant de

la vitesse de balayage du spectromètre de masse, ces cycles ont une durée comprise entre 2 et 4 secondes. Le raccourcissement de la durée de ces cycles pénalise directement l'identification puisque la réduction du nombre de MS/MS réduit le nombre des identifications des peptides et la réduction de la durée de balayage des spectres MS/MS réduit la qualité des spectres. Après l'acquisition par notre laboratoire d'un spectromètre de masse de haute résolution de type Q-TOF (MaXis, Bruker) et qui présente une gamme dynamique étendue (3 logs), nous avons réappliqué la stratégie qN-TOP sur des protéines modèles que nous avons analysé sur ce spectromètre de masse (chapitre 2, partie 2). Ce spectromètre de masse est couplé à un système UHPLC (nanoAcquity, Waters) qui permet de travailler avec des colonnes de séparation contenant des phases stationnaires avec des particules de 1,7  $\mu$ . Ce système chromatographique permet une très bonne résolution chromatographique (largeur à mi-hauteur inférieur à 6 secondes). Or, dans une analyse LC-MS/MS avec des cycles MS et MS/MS d'environ 2 secondes (1 MS et 5 MS/MS), on ne peut obtenir plus que 4 spectres MS pour décrire un pic chromatographique avec une largeur à mi-hauteur de 6 secondes et une largeur à la base de neuf secondes. Le résultat est que le pic chromatographique est mal défini (figure 5).



**Figure 5 :** Chromatogrammes d'extraction des courants d'ions du peptide N-terminal de la chaîne  $\alpha$  de l'hémoglobine marqué au TMPP léger (bleu) et lourd (rouge) après analyse LC-MS/MS sur un système UHPLC-Q-TOF (Waters nanoAcquity couplé au Bruker MaXis).

Afin de résoudre ce problème et d'avoir un pic bien défini, la meilleure solution semble être de procéder en deux temps comme dans les stratégies de quantification basé sur la MS dirigée (« directed MS ») (voir paragraphe 4.2 du chapitre 3 de la partie 2). Dans un premier temps, on effectue une analyse LC-MS/MS dans le but d'identifier un nombre maximal de peptides. A partir de cette analyse, on établit une base de données contenant les masses et les temps de rétentions des couples léger-lourd identifiés. Dans un second temps, on effectue une analyse LC-MS en injectant le même échantillon. Dans une analyse LC-MS, nous aurons assez de points pour bien décrire les pics chromatographiques (la durée d'un balayage MS est de 0,1 à 0,5 secondes). La quantification est ensuite effectuée à partir de l'analyse LC-MS en se servant des données d'identification contenues dans la base de données générée à partir de l'analyse LC-MS/MS. Cette dissociation des deux étapes, identification et quantification, permet d'augmenter la robustesse de la quantification [146].

## 5. Conclusion

Cette étude nous a permis d'identifier plusieurs améliorations à apporter notre stratégie de quantification qN-TOP. Il est clair que l'utilisation d'un spectromètre de masse de haute résolution et présentant une bonne précision de mesure de masse est indispensable pour éviter les problèmes d'interférences que nous avons observés avec la trappe d'ions. Ces interférences nous ont empêchés d'avoir une quantification correcte des processus protéolytiques. Nous avons pu cependant détecter une trentaine de processus protéolytiques présents d'une façon différentielle sans pouvoir les quantifier avec précision.

Le deuxième point très important est la bonne définition des pics chromatographiques. Pour ceci nous proposons une stratégie d'acquisition des données en deux temps. Une première analyse d'identification par LC-MS/MS suivie par une analyse LC-MS du même digest. Les données de quantification seront alors générées à partir de l'analyse LC-MS.

## Chapitre 5 Etude fonctionnelle de ReP1-NCXSQ, une protéine qui régule l'échangeur Na<sup>+</sup>/Ca<sup>2+</sup> du nerf du calamar

*Collaboration avec l'équipe de Dr. Alberto Podjarny de l'Institut de Génétique et de Biologie Moléculaire et Cellulaire de Strasbourg (IGBMC)*

### 1. Résumé

Nous avons développé et appliqué une stratégie originale combinant spectrométrie de masse en condition non-dénaturante et LC-MS de haute résolution afin d'étudier les interactions entre la protéine ReP1-NCXSQ et les acides gras. Nous avons pu ainsi déterminer l'acide gras permettant de réguler l'activité de cette protéine.

### 2. Contexte de l'étude

La protéine ReP1-NCXSQ (désignée par ReP1-WT dans la suite du chapitre) régule l'activité de l'échangeur Na<sup>+</sup>/Ca<sup>2+</sup> du nerf de Calamar NCXSQ1. Cette régulation nécessite la présence d'ATP. Dans le but de déterminer le mécanisme d'action de cette protéine et d'identifier son ligand bioactif, nous avons effectué en collaboration avec l'équipe du Dr. A. Podjarny (IGBMC) des études d'analyses de séquences, de structure cristallographique et de spectrométrie de masse de cette protéine et de son mutant inactif Y128F. Les analyses préliminaires de cette protéine par spectrométrie de masse supramoléculaire ont montré qu'elle interagit avec un ligand inconnu d'une masse moléculaire qui correspond à celle d'un acide gras. La fixation des acides gras de façon fortuite par les protéines a été reportée dans la littérature [147-149]. Dans le but de confirmer cette hypothèse et d'identifier cet acide gras, nous avons développé et appliqué avec succès une stratégie originale combinant spectrométrie de masse supramoléculaire et LC-MS de haute résolution (LC-HRMS). Cette étude a fait l'objet d'une publication dans le journal Acta Crystallographica, section D. Dans ce chapitre, après une brève introduction à la spectrométrie de masse supramoléculaire, je décris brièvement la stratégie d'analyse que nous avons développée et les principaux résultats qu'elle a permis d'obtenir.

Pour de plus amples informations et pour les résultats de l'étude cristallographique effectuée par nos collaborateurs de l'IGBMC, se référer à la publication en annexe de ce chapitre.

### 3. Brève introduction à l'étude de complexe non-covalent par spectrométrie de masse supramoléculaire

La spectrométrie de masse supramoléculaire permet d'étudier, dans des conditions non-dénaturantes, les interactions non-covalentes entre différentes biomolécules. Ces interactions peuvent être de différentes natures : liaisons hydrogènes, interactions de type Van der Waals, interactions ioniques et interactions par effet hydrophobe. La nature labile de ce type d'interactions nécessite l'utilisation d'une source d'ionisation douce capable d'ioniser et de transférer en phase gazeuse ces complexes tout en préservant leur intégrité [150]. La source MALDI, qui est une source d'ionisation source, n'est cependant pas adaptée à l'étude de ces complexes. En effet, la nature acide de la matrice utilisée pour former les dépôts cristallins d'échantillon induit la dissociation des complexes non-covalents, contrairement à la source ESI dont la capacité à maintenir leur intégrité a été démontrée dès 1991 [151]. L'analyseur TOF est l'analyseur de choix en MS supramoléculaire à cause de sa gamme de  $m/z$  étendue compatible avec des édifices de hauts poids moléculaires. La spectrométrie de masse supramoléculaire fait donc appel quasi exclusivement à la source ESI couplée à un analyseur de type TOF ou Q-TOF [150].

#### 3.1. La préparation de l'échantillon

La préparation d'échantillon est une étape clé pour l'analyse de complexes non-covalents par MS. La préservation de l'intégrité structurale des complexes en solution est une condition *sine qua non* à leur analyse et détection en phase gazeuse. Les tampons couramment utilisés pour la purification des protéines et des complexes protéiques contiennent des sels non-volatils incompatibles avec l'ionisation ESI. Une étape d'échange de tampon est donc indispensable. Le tampon d'échange doit être compatible avec l'ionisation ESI et capable de préserver l'état natif des protéines et des complexes non-covalents. De ce fait, le choix du tampon d'échange est limité à l'acétate d'ammonium et l'hydrogénocarbonate d'ammonium qui sont des sels volatils capables de stabiliser les protéines en solutions aqueuses. Les choix du pH et de la force ionique du tampon ainsi que de la technique d'échange (dialyse, gel filtration, ultrafiltration) sont échantillons dépendants et sont optimisés au cas par cas.

#### 3.2. Les paramètres instrumentaux

L'analyse de complexes non-covalents intacts par spectrométries de masse nécessite une optimisation fine des paramètres instrumentaux et donc une connaissance profonde du fonctionnement des spectromètres de masse. Il est en effet nécessaire de contrôler finement l'énergie interne communiquée aux ions lors de l'analyse. Si celle-ci est trop importante, elle entraîne la dissociation des complexes, tandis que si elle est trop faible, les ions seront mal désolvatés impliquant une baisse de la résolution et des spectres peu informatifs. Il s'agit donc de trouver un compromis qui permet à la fois de conserver l'intégrité des complexes tout en assurant une désolvatation suffisante. Les paramètres clés permettant ceci sont les tensions d'accélération et la pression dans l'interface. Les tensions d'accélération confèrent une énergie cinétique aux ions afin de les transférer, à travers l'interface (à l'aide des guides d'ions), vers l'analyseur de masse. Cette

énergie cinétique est convertie en énergie interne lors des collisions des ions avec les molécules de gaz. L'augmentation des tensions d'accélération conduit donc les ions à subir des collisions plus énergétiques favorisant ainsi leur désolvatation, mais aussi la dissociation des liaisons non-covalentes (Figure 1-a). La pression dans l'interface permet quant à elle de moduler le libre parcours moyen des ions. L'augmentation de cette pression entraîne une diminution de la distance d'accélération et donc de l'énergie cinétique (vitesse) acquise entre les collisions. Par conséquent, les ions subissent des collisions plus fréquentes mais moins énergétiques permettant de maintenir l'intégrité des complexes (Figure 1-b) [150].

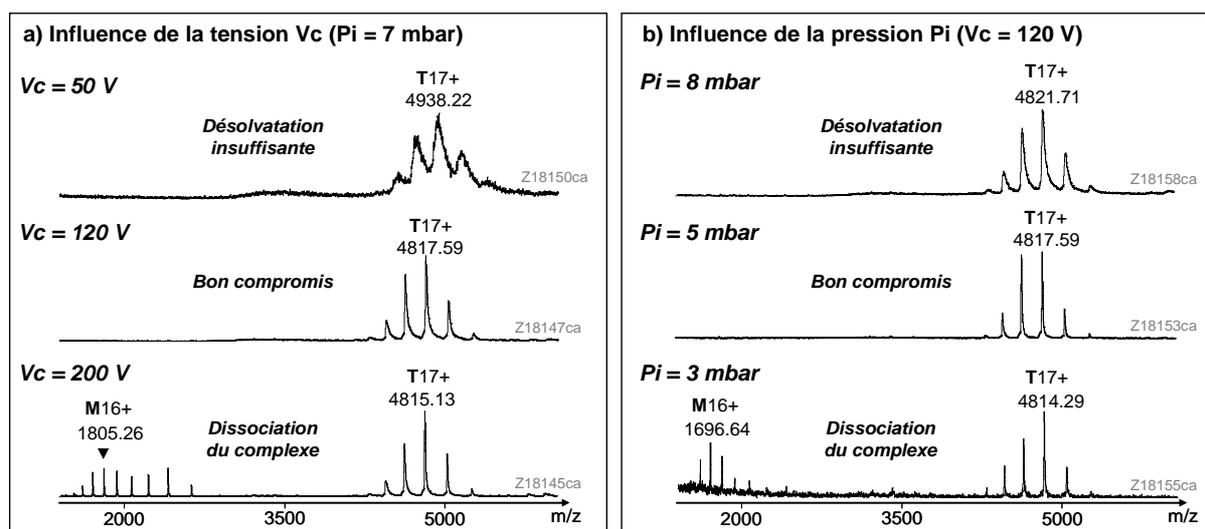


Figure 1 : Influence de la tension Vc et de la pression dans l'interface Pi sur la désolvatation et la préservation de l'intégrité des complexes non-covalents d'après C. Atmanène. Analyse d'un tétramère d'une protéine de 28868 Da [150].

Les stratégies d'analyse en spectrométrie de masse supramoléculaire sont amplement détaillées dans la thèse de C. Atmanène [150].

## 4. Stratégie analytique mise en œuvre

### 4.1. Caractérisation des complexes par spectrométrie de masse supramoléculaire

Après optimisation de l'échange de tampon, les analyses de MS supramoléculaire ont été effectuées sur un spectromètre de masse de type TOF (LCT, Waters) équipé d'un automate passeur d'échantillon avec source nanoESI sur puce microfluidique (Triversa Nanomate, Advion Biosystems).

Les mesures de masse des protéines en conditions dénaturantes sont tout d'abord réalisées afin de vérifier les masses et l'intégrité des protéines. Les masses mesurées de ReP1-WT (15597,6 ± 0,1 Da) et de son mutant ReP1-Y128F (15582,0 ± 0,1 Da) étaient en accord avec les masses calculées à partir de leurs séquences en acides aminés. Des mesures de masse ont ensuite été effectuées en conditions non-dénaturantes. Les protéines ont été analysées dans un tampon acétate d'ammonium. La pression dans la source du spectromètre de masse et la tension d'accélération ont été optimisées à 750 Pa et 10 V respectivement. La masse du ligand est déterminée par le calcul de la différence entre la masse du complexe mesurée en conditions non-dénaturantes et la masse de la protéine

mesurée en conditions dénaturantes. Ces analyses ont montré que ReP1 WT et Rep1 Y128F interagissent avec plusieurs ligands qui correspondent à des masses d'acides gras (acide palmitique, palmitoléique et oléique).

Des expériences d'ajouts dosés de ces trois acides gras ont été effectuées afin d'étudier la capacité de chacun à interagir avec les protéines.

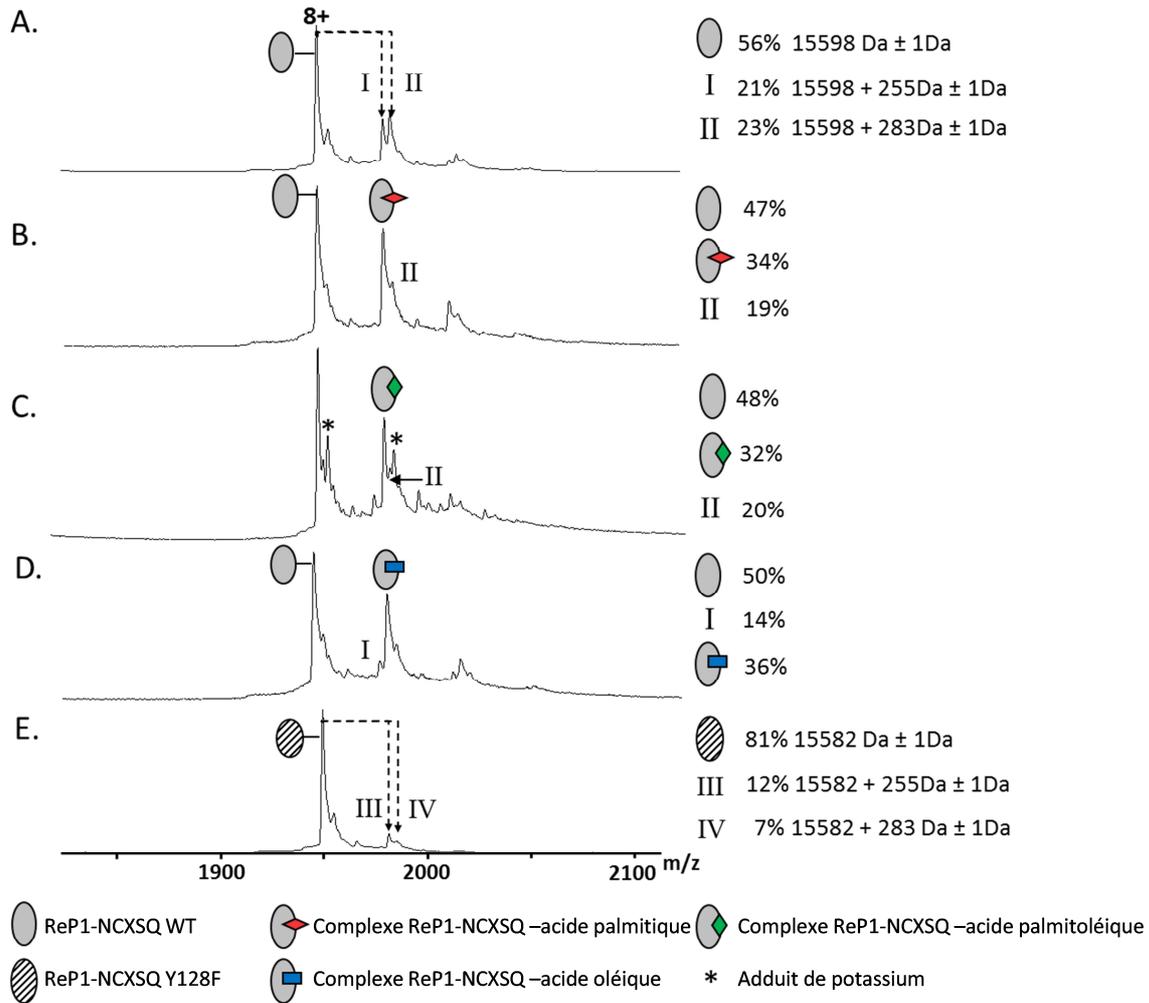
#### 4.2. Caractérisation des acides gras par LC-HRMS

Nous avons développé cette stratégie pour la caractérisation des acides gras qui interagissent avec ReP1 WT et Y128F. Elle consiste à séparer les protéines et les acides gras par RP-HPLC et à analyser les acides gras en spectrométrie de masse de haute résolution (15000) en mode d'ionisation négatif. Les complexes sont donc injectés sur une colonne de phase inverse C18 avec des particules de 3,5  $\mu\text{m}$  et une porosité de 80 Å. La faible porosité des particules exclue les protéines de la phase stationnaire, tandis que les acides gras sont retenus et puis élués avec le gradient de la phase mobile. Ceux-ci sont analysés à l'aide du spectromètre de masse de type Q-TOF (MicroTOFQ, Bruker) qui permet de déterminer leur masse exacte à  $\pm 10$  ppm. La mesure de masse exacte permet de déterminer la formule empirique des composés analysés. Une analyse d'un blanc qui consiste en la matrice dans laquelle les protéines sont dissoutes est réalisée avant chaque injection. Le chromatogramme du blanc est soustrait de celui de l'analyse des complexes protéine-acide gras. La détermination du ligand naturel de ReP1 WT est réalisée en comparant l'analyse de ReP1 WT incubée avec les vésicules lipidiques de calamar en présence et en absence d'ATP.

#### 4.3. Résultats et discussion

L'analyse par spectrométrie de masse supramoléculaire en conditions non-dénaturantes de la protéine ReP1 WT a permis de montrer que cette protéine interagit avec des ligands inconnus. En effet, 21 % et 23 % de l'intensité des ions détectés pour ReP1 WT correspondaient à une différence de masse de  $+255 \pm 1$  et  $+283 \pm 1$  Da respectivement par rapport à la masse de l'*holoprotéine* (protéine sans ligand) (Figure 2-A). Pour la protéine inactive ReP1 Y128F, on observe aussi des pics correspondant à des ions avec les mêmes différences de masse, cependant, les intensités de ces ions (12 et 7 % pour le 255 et le 283 respectivement) étaient nettement plus faibles que dans le cas de ReP1 WT (Figure 2-E). Cette faible intensité de l'*apoprotéine* dans le cas de ReP1 Y128F indique que les ligands ont moins d'affinité pour cette protéine et donc que le résidu tyrosine 128 intervient dans la fixation de ces ligands.

Les différences de masse mesurées correspondent probablement à la fixation d'acide gras par les protéines. Nous avons donc effectué des ajouts dosés acides gras différents aux solutions contenant ReP1 WT et ReP1 Y128F: acide palmitoléique (253,2 Da), acide palmitique (255,2 Da) et acide oléique (283,2 Da). L'analyse en conditions non-dénaturantes de ces mélanges a permis d'observer que l'intensité des pics correspondant à l'incrément de la masse de l'acide gras ajouté augmentait de façon proportionnelle avec la quantité ajoutée sans jamais pouvoir atteindre la saturation de la protéine (Figure 2-B, C et D). Le mutant ReP1 Y128F montrait aussi, dans ces expériences, une moindre affinité pour les acides gras que la protéine sauvage.



**Figure 2 : Analyse par MS non-dénaturante de ReP1-NCXSQ et de son mutant ReP1-NCXSQ Y128F. Zoom sur les pics correspondant à l'état de charge 8+. Les protéines sont injectées à 5 μM. A) ReP1-NCXSQ WT. B, C et D) ReP1-NCXSQ WT incubé avec trois équivalent molaire d'acide palmitique, palmitoléique et oléique respectivement. E) ReP1-NCXSQ Y128F.**

Pour caractériser ces acides gras et déterminer le ligand naturel permettant d'activer cette protéine, nous avons effectué des expériences de LC-HRMS. La protéine ReP1 WT NCQX ainsi que la protéine mutante ont été analysés dans trois conditions différentes :

- Protéine non incubée avec les vésicules lipidiques de calamar ;
- Protéine incubée avec les vésicules lipidiques de calamar en absence de MgATP ;
- Protéine incubée avec les vésicules lipidiques de calamar en présence de MgATP.

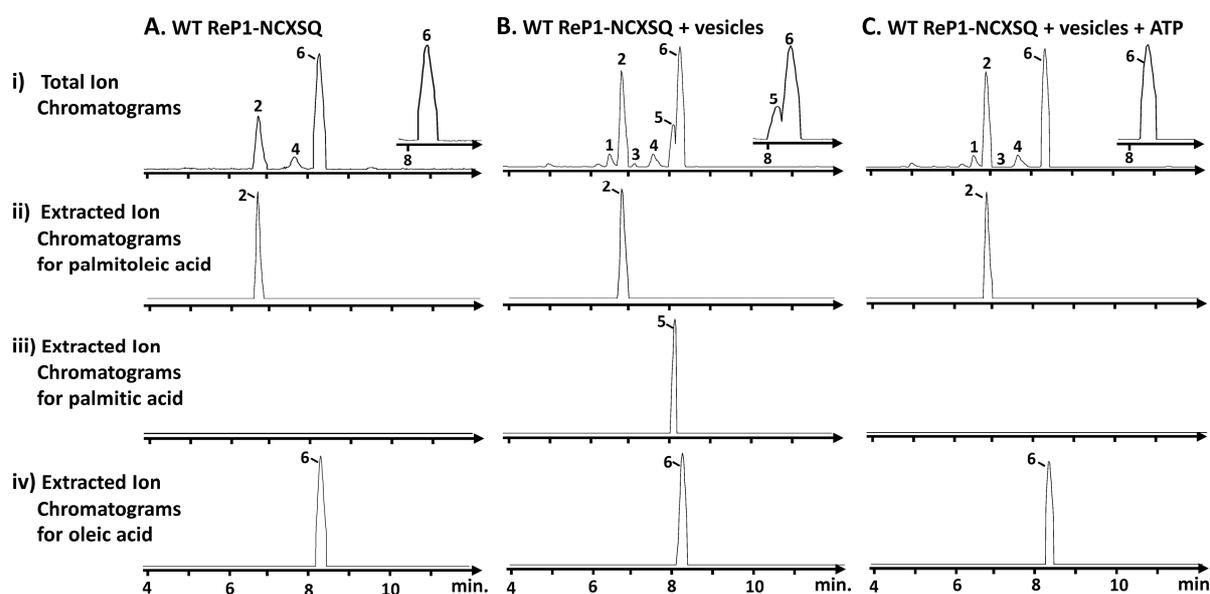
La mesure de masse à haute résolution (15000 à m/z 622) permet de déterminer la masse exacte des composés analysés à ± 10 ppm. Ceci permet de déterminer leur formule empirique et de comparer leur massif isotopique avec le massif isotopique de la molécule empirique proposée. Nous avons donc réussi à l'aide de cette stratégie à identifier les acides gras présents en solution (Tableau 1 et Figure 3 et 4).

L'analyse LC-HRMS de ReP1 WT non incubée avec les vésicules lipidiques montre que l'acide palmitique est absent de la solution (Figure 3-A). Après incubation de cette protéine avec les

vésicules lipidiques, on observe des chromatogrammes entièrement superposables avec celui obtenu pour la protéine non incubée (Figure 3-B et C). Cependant, pour la protéine incubée en absence de MgATP, on observe l'apparition d'un pic supplémentaire (Figure 3-B). Le seul composé présent dans ce pic est identifié par HRMS et correspond à l'acide palmitique ( $m/z$  255,2324). La reproductibilité de ces résultats a été vérifiée en réalisant trois répétitions indépendantes de ces expériences. Ces résultats ont été interprétés de la manière suivante : ReP1 WT purifiée n'interagit pas avec un acide palmitique, cependant après incubation de ReP1 WT avec les vésicules lipidiques du calamar en absence de MgATP, de l'acide palmitique présent dans les vésicules est piégé dans une poche spécifique (démonstré par cristallographie, voir la publication à la fin du chapitre). En présence de MgATP, l'acide palmitique n'est plus détecté par LC-HRMS, il est donc probablement transféré à l'échangeur NCXSQ1.

No. du pic	Acide gras	Formule moléculaire	Masse monoisotopique théorique (Da)*	Masse monoisotopique mesurée (Da)*	Précision (ppm)
1	Acide Docosahexaénoïque (DHA)	C <sub>22</sub> H <sub>32</sub> O <sub>2</sub>	327,2324	327,2319	2,5
2	Acide palmitoléique	C <sub>16</sub> H <sub>30</sub> O <sub>2</sub>	253,2168	253,2165	3,1
3	Acide Linoléique	C <sub>18</sub> H <sub>32</sub> O <sub>2</sub>	279,2324	279,2320	2,9
4	Acide dihomog- $\gamma$ -linoléique (DGLA) ou acide éicosatétraénoïque (ETA) + espèce inconnue	C <sub>20</sub> H <sub>34</sub> O <sub>2</sub> ?	305,2480 -	305,2473 267,2320	4,6 -
5	Acide palmitique	C <sub>16</sub> H <sub>32</sub> O <sub>2</sub>	255,2324	255,2322	3,0
6	Acide oléique	C <sub>18</sub> H <sub>34</sub> O <sub>2</sub>	281,2481	281,2478	2,3

**Tableau 1 : Liste des acides gras identifiés à l'aide des mesures HRMS. La colonne numéro de pic fait référence aux pics observés sur les chromatogrammes des figures 3 et 4. \* les masses sont celles des acides gras déprotonés (ions [M-H]).**



**Figure 3 : Analyse LC-HRMS et identification des acides gras. Chromatogrammes des acides gras interagissant avec ReP1 WT (A) et ReP1 WT incubée avec les vésicules lipidiques de calamar en absence (B) et en présence (C) de MgATP (voir tableau 1 pour la liste des acides gras).**

Les mêmes analyses ont été réalisées avec le mutant ReP1 Y128F dans les mêmes conditions. Les résultats nous permettent de constater que ce mutant n'est plus capable de fixer l'acide gras. Les chromatogrammes obtenus dans les trois conditions d'incubation sont entièrement superposables (Figure 4). Ceci montre que le résidu tyrosine 128 est directement impliqué dans l'interaction avec l'acide palmitique et dans son transport.

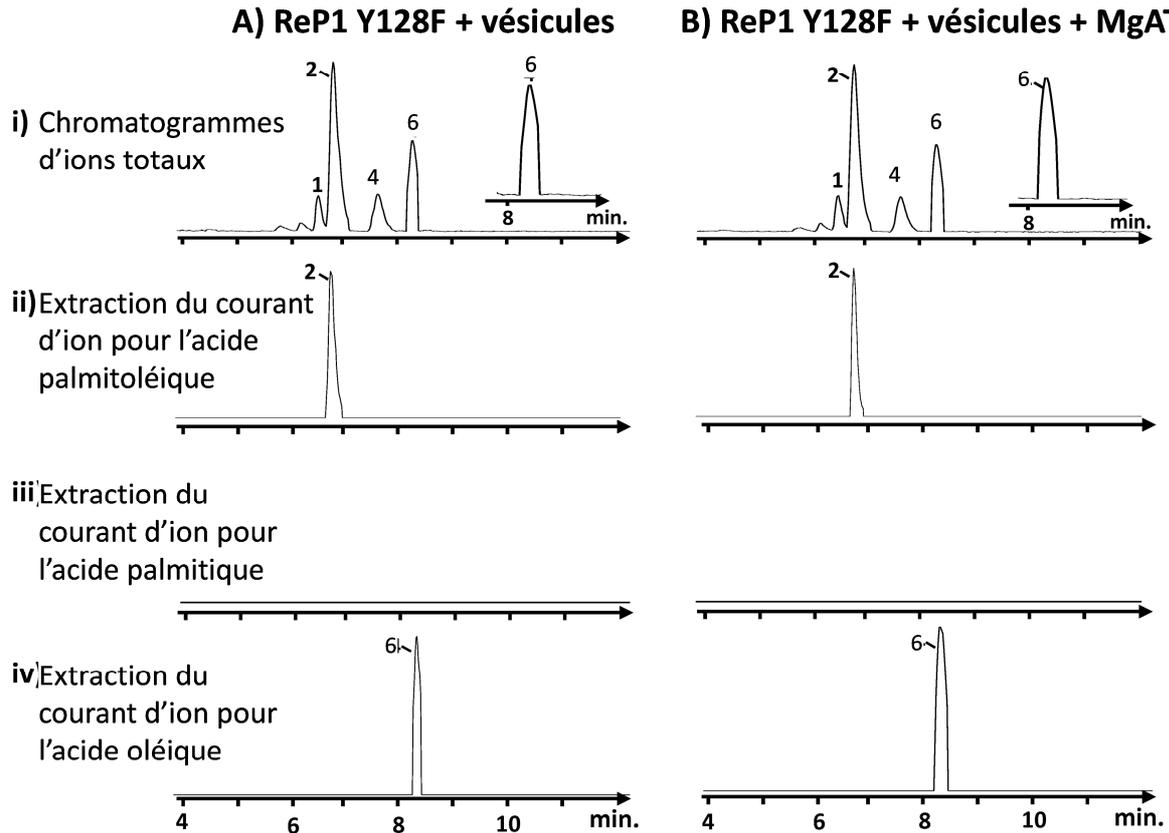


Figure 4 : Analyse LC-HRMS et identification des acides gras. Chromatogrammes des acides gras interagissant avec ReP1 Y128F incubée avec les vésicules lipidiques de calamar en absence de MgATP (A) et en présence de MgATP (B) (voir tableau 1 pour la liste des acides gras).

## 5. Conclusion

Notre stratégie de spectrométrie de masse supramoléculaire combinée à la caractérisation des acides gras par MS à haute résolution (HRMS) a permis d'identifier l'acide palmitique comme étant le ligand biologique de la protéine ReP1-NCXSQ et de mettre en évidence le rôle indispensable du résidu tyrosine 128 dans l'interaction et le transport de ce ligand. Ces résultats combinés à ceux des études cristallographiques ont permis de mieux comprendre la structure de cette protéine et de proposer un nouveau mécanisme de régulation de l'échangeur Na<sup>+</sup>/Ca<sup>2+</sup> basé sur le transport d'acide palmitique. L'expérience clé pour l'identification de ce ligand biologique était l'analyse LC-HRMS des protéines incubées avec les vésicules de calamar en présence et en absence de MgATP. Elle a permis de montrer que l'acide palmitique présent dans les vésicules est piégé dans la poche active de la protéine (vérifié par cristallographie) et libéré en présence de MgATP. Ses résultats ont fait l'objet d'une publication dans le journal Acta Crystallographica Section D.



## **Publication**

### Structural and functional studies of ReP1-NCXSQ, a protein regulating the squid nerve Na<sup>+</sup>/Ca<sup>2+</sup> exchanger

Cousido-Siah, A., Ayoub, D., Berberían, G., Bollo, M., Van Dorsselaer, A., Debaene, F., DiPolo, R., Petrova, T., Schulze-Briese, C., Olieric, V., Esteves, A., Mitschler, A., Sanglier-Cianférani, S., Beaugé, L., Podjarny, A.



## Structural and functional studies of ReP1-NCXSQ, a protein regulating the squid nerve $\text{Na}^+/\text{Ca}^{2+}$ . Erratum

Alexandra Cousido-Siah,<sup>a</sup> Daniel Ayoub,<sup>b,c</sup> Graciela Berberían,<sup>d</sup> Mariana Bollo,<sup>d</sup> Alain Van Dorsselaer,<sup>b,c</sup> François Debaene,<sup>b,c</sup> Reinaldo DiPolo,<sup>e</sup> Tatiana Petrova,<sup>f</sup> Clemens Schulze-Briese,<sup>g</sup> Vincent Olieric,<sup>g</sup> Adriana Esteves,<sup>h</sup> André Mitschler,<sup>a</sup> Sarah Sanglier-Cianférani,<sup>b,c</sup> Luis Beaugé<sup>d</sup> and Alberto Podjarny<sup>a\*</sup>

<sup>a</sup>Department of Structural Biology and Genomics, IGBMC, CNRS, INSERM, Université de Strasbourg, Illkirch, France, <sup>b</sup>Laboratoire de Spectrométrie de Masse BioOrganique (LSMBO), Université de Strasbourg, IPHC, 25 Rue Becquerel, 67087 Strasbourg, France, <sup>c</sup>CNRS, UMR7178, 67037 Strasbourg, France, <sup>d</sup>Laboratorio de Biofísica, Instituto de Investigación Médica 'Mercedes y Martín Ferreyra' (INIMEC-CONICET), Córdoba, Argentina, <sup>e</sup>Centro de Biofísica y Bioquímica, Instituto Venezolano de Investigaciones Científicas, Caracas, Venezuela, <sup>f</sup>Institute of Mathematical Problems of Biology, Russian Academy of Sciences, Pushchino

142290, Russian Federation, <sup>g</sup>SLS, PSI, Villigen, Switzerland, and <sup>h</sup>Sección Bioquímica, Facultad de Ciencias, UdelaR, Montevideo, Uruguay

Correspondence e-mail: podjarny@igbmc.fr

A correction is published to Table 2 in the article by Cousido-Siah *et al.* [(2012), *Acta Cryst.* **D68**, 1098–1107].

On page 1103 of the article by Cousido-Siah *et al.* (2012) the decimal points are missing in the molecular masses given in Table 2. The correct masses are included in the corrected Table 2 given below. The conclusions of the article remain unchanged.

### References

Cousido-Siah, A., Ayoub, D., Berberían, G., Bollo, M., Van Dorsselaer, A., Debaene, F., DiPolo, R., Petrova, T., Schulze-Briese, C., Olieric, V., Esteves, A., Mitschler, A., Sanglier-Cianférani, S., Beaugé, L. & Podjarny, A. (2012). *Acta Cryst.* **D68**, 1098–1107.

**Table 2**

Summary of identified fatty acids from HRMS measurements (Fig. 3).

Theoretical molecular masses are indicated as  $[M - H]^-$  ions. n.i., not identified.

Peak No.	Fatty acid	Molecular formula	Theoretical monoisotopic mass (Da)	Measured monoisotopic mass (Da)
1	Docosahexaenoic acid (DHA)	$\text{C}_{22}\text{H}_{32}\text{O}_2$	327.232	327.232 (4)
2	Palmitoleic acid	$\text{C}_{16}\text{H}_{30}\text{O}_2$	253.216	253.215 (4)
3	Linoleic acid	$\text{C}_{18}\text{H}_{32}\text{O}_2$	279.232	279.232 (2)
4	Dihomo- $\gamma$ -linoleic acid (DGLA) or eicosatetraenoic acid (ETA) + n.i.	$\text{C}_{20}\text{H}_{34}\text{O}_2$	305.248	305.247 (1)
5	Palmitic acid	$\text{C}_{16}\text{H}_{32}\text{O}_2$	255.232	255.232
6	Oleic acid	$\text{C}_{18}\text{H}_{34}\text{O}_2$	281.248	281.247 (4)

Alexandra Cousido-Siah,<sup>a</sup> Daniel Ayoub,<sup>b,c</sup> Graciela Berberían,<sup>d</sup> Mariana Bollo,<sup>d</sup> Alain Van Dorsselaer,<sup>b,c</sup> François Debaene,<sup>b,c</sup> Reinaldo DiPolo,<sup>e</sup> Tatiana Petrova,<sup>f</sup> Clemens Schulze-Briese,<sup>g</sup> Vincent Olieric,<sup>g</sup> Adriana Esteves,<sup>h</sup> André Mitschler,<sup>a</sup> Sarah Sanglier-Cianféran,<sup>b,c</sup> Luis Beauge<sup>d</sup> and Alberto Podjarny<sup>a\*</sup>

<sup>a</sup>Department of Structural Biology and Genomics, IGBMC, CNRS, INSERM, Université de Strasbourg, Illkirch, France, <sup>b</sup>Laboratoire de Spectrométrie de Masse BioOrganique (LSMBO), Université de Strasbourg, IPHC, 25 Rue Becquerel, 67087 Strasbourg, France, <sup>c</sup>CNRS, UMR7178, 67037 Strasbourg, France, <sup>d</sup>Laboratorio de Biofísica, Instituto de Investigación Médica 'Mercedes y Martín Ferreyra' (INIMEC-CONICET), Córdoba, Argentina, <sup>e</sup>Centro de Biofísica y Bioquímica, Instituto Venezolano de Investigaciones Científicas, Caracas, Venezuela, <sup>f</sup>Institute of Mathematical Problems of Biology, Russian Academy of Sciences, Pushchino 142290, Russian Federation, <sup>g</sup>SLS, PSI, Villigen, Switzerland, and <sup>h</sup>Sección Bioquímica, Facultad de Ciencias, UdelaR, Montevideo, Uruguay

Correspondence e-mail: podjarny@igbmc.fr

## Structural and functional studies of ReP1-NCXSQ, a protein regulating the squid nerve Na<sup>+</sup>/Ca<sup>2+</sup> exchanger

The protein ReP1-NCXSQ was isolated from the cytosol of squid nerves and has been shown to be required for MgATP stimulation of the squid nerve Na<sup>+</sup>/Ca<sup>2+</sup> exchanger NCXSQ1. In order to determine its mode of action and the corresponding biologically active ligand, sequence analysis, crystal structures and mass-spectrometric studies of this protein and its Tyr128Phe mutant are reported. Sequence analysis suggests that it belongs to the CRABP family in the FABP superfamily. The X-ray structure at 1.28 Å resolution shows the FABP β-barrel fold, with a fatty acid inside the barrel that makes a relatively short hydrogen bond to Tyr128 and shows a double bond between C9 and C10 but that is disordered beyond C12. Mass-spectrometric studies identified this fatty acid as palmitoleic acid, confirming the double bond between C9 and C10 and establishing a length of 16 C atoms in the aliphatic chain. This acid was caught inside during the culture in *Escherichia coli* and therefore is not necessarily linked to the biological activity. The Tyr128Phe mutant was unable to activate the Na<sup>+</sup>/Ca<sup>2+</sup> exchanger and the corresponding crystal structure showed that without the hydrogen bond to Tyr128 the palmitoleic acid inside the barrel becomes disordered. Native mass-spectrometric analysis confirmed a lower occupancy of the fatty acid in the Tyr128Phe mutant. The correlation between (i) the lack of activity of the Tyr128Phe mutant, (ii) the lower occupancy/disorder of the bound palmitoleic acid and (iii) the mass-spectrometric studies of ReP1-NCXSQ suggests that the transport of a fatty acid is involved in regulation of the NCXSQ1 exchanger, providing a novel insight into the mechanism of its regulation. In order to identify the biologically active ligand, additional high-resolution mass-spectrometric studies of the ligands bound to ReP1-NCXSQ were performed after incubation with squid nerve vesicles both with and without MgATP. These studies clearly identified palmitic acid as the fatty acid involved in regulation of the Na<sup>+</sup>/Ca<sup>2+</sup> exchanger from squid nerve.

Received 8 February 2012

Accepted 8 May 2012

### PDB References:

ReP1-NCXSQ, 3ppt;

Tyr128Phe mutant, 3pp6.

### 1. Introduction

Na<sup>+</sup>/Ca<sup>2+</sup> exchangers are highly regulated by processes taking place at the large intracellular loop. The two exchangers that have been most studied functionally and structurally are those

from mammalian heart (NCX1<sup>1</sup>) and squid nerve (NCXSQ1) (Blaustein & Lederer, 1999; DiPolo & Beaugé, 2006). Both were studied in their native state *in situ* (giant patch clamp, internally dialyzed squid giant axons or inside-out plasma membrane vesicles) or in alien cells expressing the cloned exchangers (Berberían *et al.*, 2012). The basis of these regulations are related to the existence of nontransporting Ca<sup>2+</sup>-regulatory sites located on the intracellular 'regulatory' loop; these sites must be occupied in order for any transport mode to take place. Ionic (H<sub>i</sub><sup>+</sup> and H<sub>i</sub><sup>+</sup> + Na<sub>i</sub><sup>+</sup>) inhibition acts by impairing the binding of Ca<sup>2+</sup> to those sites located in two domains named CBD1 and CBD2 (Hilge *et al.*, 2006; Nicoll *et al.*, 2006); as expected, their effects are counteracted by increasing the concentration of intracellular calcium [Ca<sup>2+</sup>]<sub>i</sub>. In addition, there is an MgATP up-regulation which, by an as yet unknown intimate mechanism, protects the Ca<sup>2+</sup>-regulatory site from intracellular H<sub>i</sub><sup>+</sup> and H<sub>i</sub><sup>+</sup> + Na<sub>i</sub><sup>+</sup> inhibition (Doering & Lederer, 1994; Hilgemann *et al.*, 1992; DiPolo & Beaugé, 2002). Although the ionic regulations of the exchangers are similar in the two aforementioned preparations, they differ drastically in the case of MgATP. In mammalian heart NCX1 up-regulation takes place *via* the production and binding to the exchanger of phosphatidylinositol 4,5-bisphosphate [PtdIns(4,5)P<sub>2</sub>]. On the other hand, in NCXSQ1 PtdIns(4,5)P<sub>2</sub> is ineffective (DiPolo & Beaugé, 2002), while MgATP up-regulation requires the presence of a soluble cytosolic protein (Beaugé *et al.*, 1996) present in both nerve axoplasm and optic ganglia. This protein of 132 amino acids, named ReP1-NCXSQ (regulatory protein of the squid nerve sodium calcium exchanger), was cloned, overexpressed and purified. Functional experiments showed that to be active ReP1-NCXSQ must be phosphorylated by a membrane-bound staurosporin-insensitive kinase and that MgATP might be necessary for this phosphorylation (Berberían *et al.*, 2007). It should be noted that, in contrast to these previous functional studies, the protein that we studied here was found not to be phosphorylated. This protein was observed in several squid species, including *Loligo pealei*, *Doryteuthis plei* and *Sepiotheutis sepioidea*. Originally identified as a lipocalin (Berberían *et al.*, 2009), this protein is actually placed at the frontier between the CRABP and FABP families (see §3). FABPs and CRABPs are members of the same superfamily. A few CRABPs have been identified from invertebrates, one of which binds both palmitic acid and retinoic acid with high affinity (Gu *et al.*, 2002; Folli *et al.*, 2005). Members of both groups share a superimposable three-dimensional structure and relevant sequence similarity. Almost all members of the

FABP family bind fatty acids, while CRABPs usually bind retinoic acid. Depending on the structural elements and the database considered for classification, they can be included in the following families: cytosolic fatty-acid signature-containing proteins (PS00214, PROSITE; <http://www.expasy.org/prosite>), fatty acid-binding protein-like (SCOP; <http://scop.mrc-lmb.cam.ac.uk/scop/>) and homologous superfamily code 2.40.128.20 (CATH; <http://www.biochem.ucl.ac.uk/bsm/cath>). Recently, Ono and Odani have proposed joining the FABPs, CRABPs and retinal CRBPs in a new superfamily named FABPs (Ono & Odani, 2010). Therefore, this protein belongs to the FABP superfamily and is placed at the frontier between CRABPs and FABPs.

In this work, X-ray structural and mass-spectrometric studies of ReP1-NCXSQ suggest that it acts as a carrier of fatty acids and that palmitic acid is involved in the regulation of the Na<sup>+</sup>/Ca<sup>2+</sup> exchanger from squid nerve. Despite the important advances made regarding the structure and phospholipid interactions of ReP1-NCXSQ1, the molecular mechanism underlying regulation remains unknown.

## 2. Experimental procedures

### 2.1. Mutagenesis (Tyr128Phe)

The point mutation of ReP1-NCXSQ was obtained using Phusion polymerase (Finnzymes, Finland) and pET28a-ReP1-NCXSQ as a template. Following PCR, *DpnI* (Fermentas Inc., USA) incubation for 2 h at 310 K was used to digest parental DNA and select for mutation-containing amplified DNA. The introduction of the desired mutation into the recombinant plasmid was determined using automatic sequencing by the IGBMC core facility. The oligonucleotides used were from Invitrogen (Argentina). The primers used were as follows: forward primer, 5'-gtg tgc acc cgt att ttt aaa cgc gtg gat ctc-3'; reverse primer, 5'-gag atc cac gcg ttt aaa aat acg ggt gca cac-3'.

### 2.2. ReP1-NCXSQ purification, preparation of nerve membrane vesicles, incubation and transport studies

The expression and purification of His-tagged native and mutant ReP1-NCXSQ were performed as described previously (Berberían *et al.*, 2009). Membrane vesicles were obtained from the optic nerves of two squid species (*S. sepioidea* and *L. pealei*), subjected to homogenization and differential centrifugation as described elsewhere (Beaugé *et al.*, 1996; DiPolo *et al.*, 1997) and loaded with 300 mM NaCl, 0.1 mM Tris-EDTA, 30 mM Tris-MOPS (pH 7.3 at 293 K). About 40% of them were in the inside-out orientation, making it possible to study intracellular exchanger-ligand interactions by their presence in the extravesicular solutions.

For incubation with the vesicles, 25 µl of vesicles was incubated with 175 µl ReP1-NCXSQ at 10 mg ml<sup>-1</sup> for 20 s at room temperature with 25 µl 250 mM NMG-Cl, 50 mM HEPES-Tris pH 7.3, 1.14 mM CaCl<sub>2</sub>, 1.5 mM Tris-EGTA, 10 mM MgCl<sub>2</sub>, 2 mM vanadate with or without 25 µl 5 mM MgATP. The reaction was stopped with 4 mM Tris-EDTA and 1 mM vanadate (both final concentrations) at 273 K.

<sup>1</sup> Abbreviations used: ReP1-NCXSQ, cytosolic regulatory protein of the squid nerve Na<sup>+</sup>/Ca<sup>2+</sup> exchanger; CRABP, cellular retinoic acid-binding protein; FABP, fatty acid-binding protein, NCX1, mammalian heart Na<sup>+</sup>/Ca<sup>2+</sup> exchanger; NCXSQ1, squid nerve Na<sup>+</sup>/Ca<sup>2+</sup> exchanger; MgATP, magnesium adenosine 5'-triphosphate; LBP, lipid-binding protein; CRBP, cellular retinol-binding protein; H<sub>i</sub><sup>+</sup>, intracellular H<sup>+</sup>; Na<sub>i</sub><sup>+</sup>, intracellular Na<sup>+</sup>; Ca<sub>i</sub><sup>2+</sup>, intracellular Ca<sup>2+</sup>; PtdIns(4,5)P<sub>2</sub>, phosphatidylinositol 4,5-bisphosphate; NMG, *N*-methyl-D-glucamine; CHES, *N*-cyclohexyl-2-aminoethanesulfonic acid; MOPS, 3-(*N*-morpholino)propanesulfonic acid; HRMS, high-resolution mass spectrometry; ESI-MS, electrospray ionization mass spectrometry; LC-MS, liquid chromatography coupled to mass spectrometry.

**Table 1**

X-ray data-collection and refinement statistics for WT and Tyr128Phe mutant ReP1-NCXSQ crystals.

Values in parentheses are for the highest resolution shell.

Protein	WT	Tyr128Phe mutant
<b>Data collection</b>		
X-ray beamline and source	X06DA [PXIII], SLS	X06DA [PXIII], SLS
Wavelength (Å)	0.9000	0.91960
Space group	$P2_12_12_1$	$P3_2$
Asymmetric unit content	Monomer	Trimer
Unit-cell parameters (Å)	$a = 40.32, b = 49.58,$ $c = 60.47$	$a = 80.26, c = 59.34$
Resolution range (Å)	50.0–1.28 (1.33–1.28)	50.0–1.90 (1.97–1.90)
No. of observations	175673	103423
No. of unique reflections	31531	33356
Multiplicity	5.6 (5.2)	3.1 (3.1)
Completeness (%)	98.7 (97.6)	98.8 (98.3)
$R_{\text{merge}}^1$ (%)	4.3 (37.6)	3.5 (37.5)
Mean $I/\sigma(I)$	32.8 (4.1)	32.3 (3.2)
<b>Refinement</b>		
Refinement resolution range (Å)	30.25–1.28	34.76–1.90
$R$ factor (%)	13.4	20.8
$R_{\text{free}}$ (%)	17.1	26.6
No. of reflections used in refinement	31468	31990
Model used for molecular replacement	1yiv	1yiv
R.m.s.d. bonds (Å)	0.008	0.007
R.m.s.d. angles (°)	1.291	1.099
No. of protein atoms	1267	3801
No. of solvent molecules	252	363
<b>Ramachandran plot</b>		
Favoured (%)	98.53	97.02
Accepted (%)	1.47	2.68
Outliers (%)	0	0.3

$$^1 R_{\text{merge}} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$$

The membrane vesicles were separated by centrifugation at 12 000 rev min<sup>-1</sup> for 30 min.

The assays for ion transport consisted of measuring the Na<sup>+</sup>-gradient-dependent <sup>45</sup>Ca<sup>2+</sup> uptake (forward mode in inside-out vesicles; Berberian *et al.*, 2009). The external solutions consisted of low (20 mM) or high (300 mM) NaCl, 1 mM free Mg<sup>2+</sup>, 0.2 mM vanadate, 0.15 mM EGTA, 20 mM Tris-MOPS (pH 7.3 at 293–295 K) and 1 μM Ca<sup>2+</sup>; when present, MgATP was at 1 mM concentration. Iso-osmolar amounts of NMG-Cl compensated for the reduction in Na<sup>+</sup> concentration. The Na<sup>+</sup>-gradient-dependent Ca<sup>2+</sup> uptake was taken as the difference between that observed at low and high Na<sup>+</sup> concentration; the incubation time was 10 s and the temperature was 293–295 K. The transport studies were performed for the wild type (WT) and the mutant.

### 2.3. Crystallization and structure determination of WT and Tyr128Phe mutant ReP1-NCXSQ

Crystals of WT ReP1-NCXSQ were grown by hanging-drop vapour diffusion at 277 K. The protein was concentrated to 10 mg ml<sup>-1</sup> in 10 mM Tris-HCl pH 7.4, 150 mM NaCl. For crystal growth, a 2 μl drop of protein solution was mixed with an equal volume of precipitating solution (200 mM KCl, 35%

PEG 2000 MME) and equilibrated against 0.5 ml of the same solution. Typically, rod-like crystals appeared in 2–3 d and continued to grow to maximum dimensions of 300 × 100 × 50 μm. They were cryoprotected by transfer into precipitating solution plus 5% ethylene glycol and were flash-cooled in liquid ethane prior to data collection.

An initial data set was collected to 1.8 Å resolution with the laboratory source. The data were processed with *HKL-2000* (Otwinowski & Minor, 1997) and the structure was solved by molecular replacement (Navaza, 1994) using PDB entry 1yiv (Hunter *et al.*, 2005). The correct sequence was fitted with *ARP/wARP* (Perrakis *et al.*, 1997) and the structure was refined with *REFMAC5* (Murshudov *et al.*, 2011). A second data set was collected to a resolution of 1.28 Å on the X06DA beamline at the Swiss Light Source (SLS). The structure was further refined at 1.28 Å resolution with *REFMAC5* and *PHENIX* (Adams *et al.*, 2002) to an  $R$  factor of 13.4% and an  $R_{\text{free}}$  of 17.1%. The model and the diffraction data have been deposited in the Protein Data Bank with code 3ppt. Note that the differences in the occupancies of protein and ligand atoms in the PDB deposition are the result of independent refinement. For data-collection and refinement statistics, see Table 1.

For the Tyr128Phe mutant of ReP1-NCXSQ, the protein was also concentrated to 10 mg ml<sup>-1</sup> in 10 mM Tris-HCl pH 7.4, 150 mM NaCl, but the crystals were obtained using a different crystallization condition. Rod-like crystals appeared in 100 mM CHES pH 9.5, 30% PEG 3000. A data set was collected to 1.9 Å resolution on the X06DA beamline at SLS. The structure was solved by molecular replacement to an  $R$  factor of 20.8% and an  $R_{\text{free}}$  of 26.6%. The model and the diffraction data have been deposited in the Protein Data Bank with code 3pp6. For data-collection and refinement statistics, see Table 1.

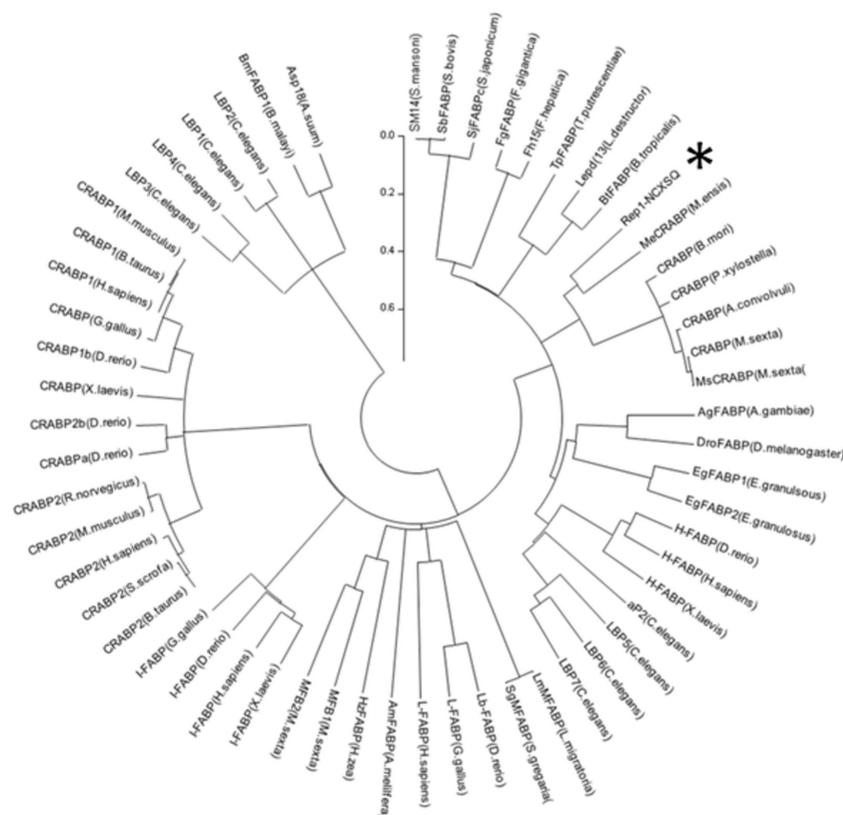
### 2.4. Native mass spectrometry

Prior to mass-spectrometric (MS) experiments, the proteins were buffer-exchanged against 10 mM ammonium acetate pH 6.8 using microcentrifuge gel-filtration columns (Zeba 0.5 ml; Thermo Scientific, Rockford, Illinois, USA). Protein concentrations were determined spectrophotometrically. Native MS analyses were performed using an electrospray time-of-flight (ESI-TOF) mass spectrometer (LCT; Waters, Manchester, England) equipped with an automated chip-based nanoESI source (Triversa Nanomate; Advion Biosciences, Ithaca, New York, USA) operating in the positive-ion mode. External calibration was performed using the multiply charged ions produced by 2 μM horse heart myoglobin diluted in 1:1(v:v) water:acetonitrile acidified with 1%(v/v) formic acid. The molecular mass, integrity and homogeneity of the proteins were first checked under denaturing conditions by diluting the proteins to 2 μM in 1:1(v:v) water:acetonitrile acidified with 1%(v/v) formic acid. The measured molecular masses of WT ReP1-NCXSQ (15 597.6 ± 0.1 Da) and ReP1-NCXSQ Tyr128Phe (15 582.0 ± 0.1 Da) were found to be in agreement with the masses calculated on the basis of the amino-acid sequences with deletion of the N-terminal methionine. Native

MS experiments require careful instrumental tuning to find the best compromise between ion desolvation/transmission through the mass spectrometer and the preservation of weak noncovalent interactions in the gas phase. In this study, instrumental optimization especially concerned the accelerating voltage applied on the sample cone ( $V_c$ ) and the pressure in the first pumping stage of the mass spectrometer ( $P_1$ ). While  $P_1$  was fixed at 750 Pa,  $V_c$  was fixed at 10 V for protein–fatty acid complex detection. Native MS analysis were performed by diluting buffer-exchanged proteins to 5  $\mu$ M in H<sub>2</sub>O. Data analysis was performed with *MassLynx* 4.0 (Waters, Manchester, England). The peak intensities at the three charge states +7, +8 and +9 of the different species (apo-protein and holoprotein) were used to calculate the ratios of each detected species.

## 2.5. High-resolution LC-MS analysis

LC-MS analyses were performed on an Agilent 1100 HPLC system (Agilent Technologies, Palo Alto, California, USA) coupled to an electrospray quadrupole/time of flight-type mass spectrometer (MicroTOF Q; Bruker Daltonics, Bremen, Germany). 1  $\mu$ g protein was injected onto a Zorbax SB-C18 Rapid Resolution Cartridge 2.1  $\times$  30 mm, 3.5  $\mu$ m particle-size, 80  $\text{\AA}$  bore-diameter column (Agilent Technologies). The solvent system consisted of 5 mM aqueous ammonium acetate (solvent *A*) and a mixture of 75% acetonitrile, 20% 2-propanol and 5% 50 mM aqueous ammonium acetate (solvent *B*). While the proteins were excluded from the column owing to the narrow bore diameter (80  $\text{\AA}$ ), the fatty acids bound to the proteins were trapped on the column at 15% solvent *B* and elution was performed at a flow rate of 0.25 ml min<sup>-1</sup> with a 15–90% solvent *B* gradient over 4 min followed by 90% solvent *B* for 4 min and 100% solvent *B* for a further 4 min before reconditioning of the column at 15% solvent *B*. The MicroTOF Q was operated in the negative-ion mode. The performance and resolution were verified using Tunemix (Agilent Technologies, USA). The resolution at  $m/z$  602 was 15 000. Mass calibration of the TOF was achieved prior to analysis between  $m/z$  50 and 1200 using lithium formate (10 mM) in order to obtain a mass precision of  $\pm 10$  p.p.m. The capillary voltage was set to 4500 V. The complete system was fully controlled by *Compass/HyStar* software (Bruker Daltonics, Germany). Prior to protein injection, a blank consisting of 1  $\mu$ l of the matrix in which the protein was dissolved (1 mM ammonium acetate) was injected. The blank was then subtracted from the protein analysis using the software *Metabolite Detect* (Bruker Daltonics, Germany).



**Figure 1**

Phylogenetic tree derived from neighbour-joining analysis of invertebrate sequences and selected representatives of vertebrate CRABPs and FABPs. FABPs from invertebrates: SbFABP from *Schistosoma bovis* (Sb); SjFABPc from *S. japonicum* (Sj); Sm14 from *S. mansoni* (Sm); FgFABP from *Fasciola gigantica* (Fg); Fh15 from *F. hepatica* (Fh); LBP5–7 from *Caenorhabditis elegans* (Ce); ap2-like from *C. elegans*; TpFABP from *Tyrophagus putrescentiae* (Tp); LBP1–4 from *C. elegans* (Ce); BmFABP1 from *Brugia malayi* (Bm); Asp18 from *Ascaris suum* (As). FABPs from vertebrates: heart-type FABPs (H-FABPs) from *Danio rerio* (Dr), *Homo sapiens* (Hs) and *Xenopus laevis* (Xl); intestinal-type FABPs (I-FABPs) from *H. sapiens* (Hs), *Gallus gallus* (Gg), *D. rerio* (Dr); liver-type FABPs (L-FABPs) from *H. sapiens* (Hs) and *G. gallus* (Gg); Lb-FABP, liver basic type FABP from *D. rerio* (Dr). Vertebrate and invertebrate CRABPs are listed in Supplementary Figs. S1, S2 and S3. ReP1-NCXSQ is marked with an asterisk. The scale indicates the size of the branch lengths, which is proportional to the divergence time (in arbitrary units in this case, as the molecular clock is not set).

Data analysis was performed using *DataAnalysis* 4 (Bruker Daltonics, Germany). In order to identify the fatty acids eluted from the column, we used *SmartFormula* (a function of *DataAnalysis* 4), which can calculate an empirical formula from an exact mass measurement. We calculated the formulae of all peaks eluted from the column allowing  $\pm 10$  p.p.m. error on the mass of the generated formulae.

## 2.6. Sequence alignment

An initial *BLASTx* search (default conditions) at the National Center for Biotechnology Information (NCBI) showed that ReP1-NCXSQ has the greatest number of hits with cytosolic retinoic acid-binding proteins (CRABPs) and fatty acid-binding proteins (FABPs). In order to discern which proteins belong to which subgroups, we aligned ReP1-NCXSQ

with vertebrate and invertebrate CRABPs and FABPs using the *ClustalX* v.2.0.10 software (Larkin *et al.*, 2007). *ClustalX* was used because of the lack of structural data for invertebrate homologues. Selected sequences from GenBank were used to develop a neighbour-joining tree using *MEGA* (Kumar *et al.*, 1993). The significance of branching points was assessed by bootstrapping with 500 pseudoreplicates.

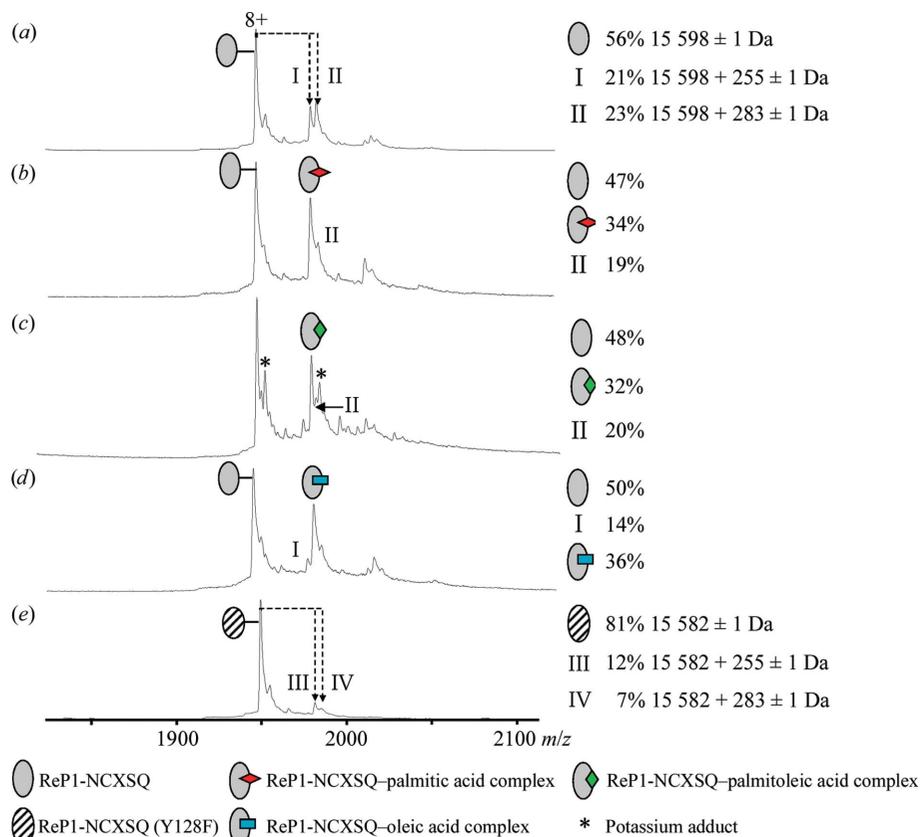
### 3. Results

#### 3.1. ReP1-NCXSQ belongs to the CRABP family and the FABP superfamily

Alignment of ReP1-NCXSQ with CRABPs from invertebrates (Supplementary Fig. S1<sup>2</sup>) revealed a similarity of 70% and an identity of 37%, while that with vertebrate CRABPs gave a similarity of 66% and an identity of 30% (Supplementary Fig. S2<sup>2</sup>), suggesting that it belongs to the CRABP family of the FABP superfamily (Ono & Odani, 2010). In particular, it presents the fingerprint of the FABP superfamily and most of its conserved amino acids (Furuhashi & Hotamisligil, 2008; Esteves & Ehrlich, 2006). Supplementary Fig. S3<sup>2</sup> shows the conserved amino acids of H-FABPs and CRABPs that are shared with the deduced ReP1-NCXSQ sequence. The sequence alignment from *BLAST* shows that the closest sequence is that of the retinoic acid/retinol-binding protein from shrimp (gb|AAL68638.1|AF458289\_1). It also shows a strong similarity, both in sequence and structure, to myelin P2 proteins from vertebrates.

When ReP1-NCXSQ was aligned with invertebrate FABPs similarity and identity were considerably lower (23% similarity and 7.5% identity; Supplementary Fig. S4<sup>2</sup>). To further explore the position of the squid protein in the protein superfamily, a phylogenetic reconstruction was performed. The ReP1-NCXSQ sequence clusters with *Metapenaeus ensis* CRABP and forms a sister group to a cluster of insect CRABPs. Both clades are also clustered with invertebrate FABPs (Fig. 1). Moreover, if we restrict the comparison to proteins from arthropods and even insects, this trend is supported (Supplementary Figs. S5 and S6<sup>2</sup>). A detailed discussion of the assignment as FABP or CRABP is given in the Supplementary Material.<sup>2</sup>

<sup>2</sup> Supplementary material has been deposited in the IUCr electronic archive (Reference: EN5495). Services for accessing this material are described at the back of the journal.



**Figure 2**

Native mass-spectrometric analysis of ReP1-NCXSQ and its Tyr128Phe mutant. ESI-MS mass spectra obtained under nondenaturing conditions (1 mM ammonium acetate buffer pH 6.8) for (a) WT ReP1-NCXSQ, (b) WT ReP1-NCXSQ incubated with three molar equivalents of palmitic acid, (c) WT ReP1-NCXSQ incubated with three molar equivalents of palmitoleic acid, (d) WT ReP1-NCXSQ incubated with three molar equivalents of oleic acid and (e) ReP1-NCXSQ Tyr128Phe mutant. The accelerating cone voltage was set to 10 V and the interface pressure to 750 Pa in order to preserve weak protein–ligand interactions. Proteins were injected at 5 µM.

#### 3.2. Mass-spectrometric characterization of ReP1-NCXSQ and its Tyr128Phe mutant

Prior to crystallization studies, ReP1-NCXSQ and its Tyr128Phe mutant were characterized in depth by a combination of MS-based techniques. Native MS analyses were first conducted in order to detect any unknown ligand arising from the *Escherichia coli* production medium and bound to recombinant ReP1-NCXSQ. Liquid chromatography coupled to high-resolution mass spectrometry (LC-HRMS) was then used to identify bound compounds.

Fig. 2 presents mass spectra obtained under native conditions for both recombinant ReP1-NCXSQ and its Tyr128Phe mutant. In the wild-type recombinant protein (WT ReP1-NCXSQ), 44% of the ions detected correspond to species with an unknown ligand bound to the protein. Among these, 21% had a mass increase of 255 ± 1 Da, while the other 23% had a mass increase of 283 ± 1 Da (Fig. 2a).

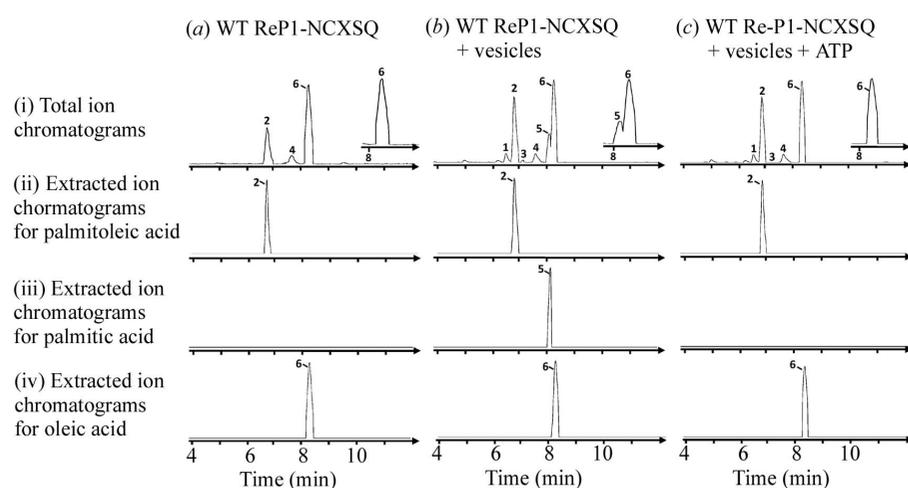
As fatty acids are molecules that are known to fortuitously interact with proteins such as nuclear receptors (Bourguet *et al.*, 2000; de Urquiza *et al.*, 2000; Elviri *et al.*, 2001; Potier *et al.*, 2003) and because ReP1-NCXSQ belongs to the FABP

**Table 2**

Summary of identified fatty acids from HRMS measurements (Fig. 3).

Theoretical molecular masses are indicated as  $[M - H]^-$  ions. n.i., not identified.

Peak No.	Fatty acid	Molecular formula	Theoretical monoisotopic mass (Da)	Measured monoisotopic mass (Da)
1	Docosahexaenoic acid (DHA)	$C_{22}H_{32}O_2$	327232	327232 (4)
2	Palmitoleic acid	$C_{16}H_{30}O_2$	253216	253215 (4)
3	Linoleic acid	$C_{18}H_{32}O_2$	279232	279232 (2)
4	Dihomo- $\gamma$ -linoleic acid (DGLA) or eicosatetraenoic acid (ETA) + n.i.	$C_{20}H_{34}O_2$ n.i.	305248 n.i.	305247 (1) 267232 (3)
5	Palmitic acid	$C_{16}H_{32}O_2$	255232	255232
6	Oleic acid	$C_{18}H_{34}O_2$	281248	281247 (4)



**Figure 3**

Identification of fatty acids bound to ReP1-NCXSQ by high-resolution LC-MS analysis (LC-HRMS). LC chromatograms of fatty-acid binding to (a) wild-type ReP1-NCXSQ, (b) ReP1-NCXSQ pre-incubated with nerve vesicles without MgATP and (c) ReP1-NCXSQ pre-incubated with nerve vesicles in the presence of MgATP. (i) Represent total ion chromatograms and (ii–iv) correspond to extracted ion chromatograms of palmitoleic (ii), palmitic (iii) and oleic (iv) acids, respectively.

superfamily, palmitic/palmitoleic acids could be proposed as possible interacting compounds corresponding to the  $15\ 853 \pm 1$  Da complex, while linoleic/oleic acid may correspond to the  $15\ 881 \pm 1$  Da compound (Fig. 2a). The binding capabilities of palmitoleic, palmitic and oleic acids to both ReP1-NCXSQ and the Tyr128Phe mutant were investigated by the direct addition of three molar equivalents of each individual fatty acid to wild-type ReP1-NCXSQ protein (Figs. 2b–2e). As expected, depending on the fatty acid spiked, subsequent increases of the  $m/z$  ratio corresponding to either palmitoleic, palmitic or oleic acid bound to WT ReP1-NCXSQ were observed (Figs. 2b–2d). Interestingly, for the Tyr128Phe mutant the native MS results show a lower occupancy for fatty acids. Only 12% of the detected species could correspond to an ReP1-NCXSQ–palmitic/palmitoleic acid complex and 7% to an ReP1-NCXSQ–linoleic/oleic acid complex (Fig. 2e).

High-resolution LC-MS experiments (LC-HRMS) were then performed to deduce molecular formulae from exact mass measurements. LC chromatograms obtained with the

WT ReP1-NCXSQ protein (Fig. 3a and Table 2) show that there is no peak corresponding to either palmitic or linoleic acid. Conversely, a peak corresponding to palmitoleic acid was clearly identified. Together, these MS results prompted us to conclude that oleic and palmitoleic acids correspond to the  $15\ 881 \pm 1$  and  $15\ 853 \pm 1$  Da complexes detected, respectively.

### 3.3. Mass-spectrometric characterization of the fatty acid involved in the regulation of NCXSQ

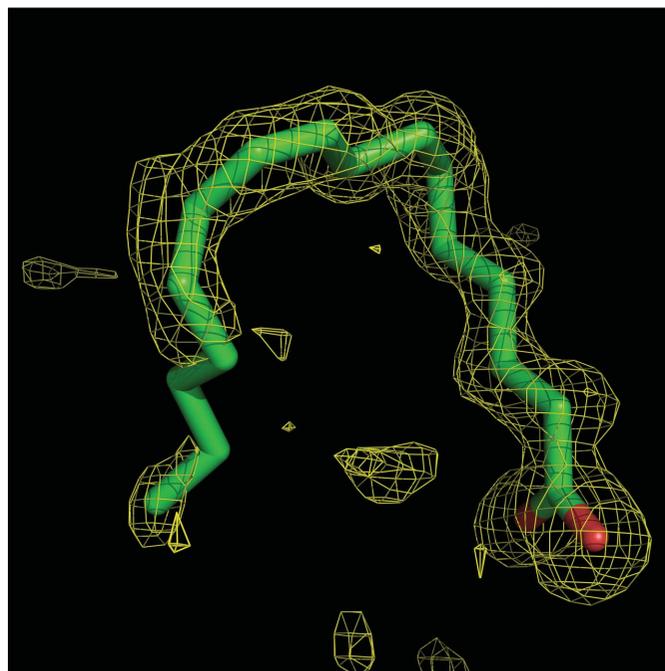
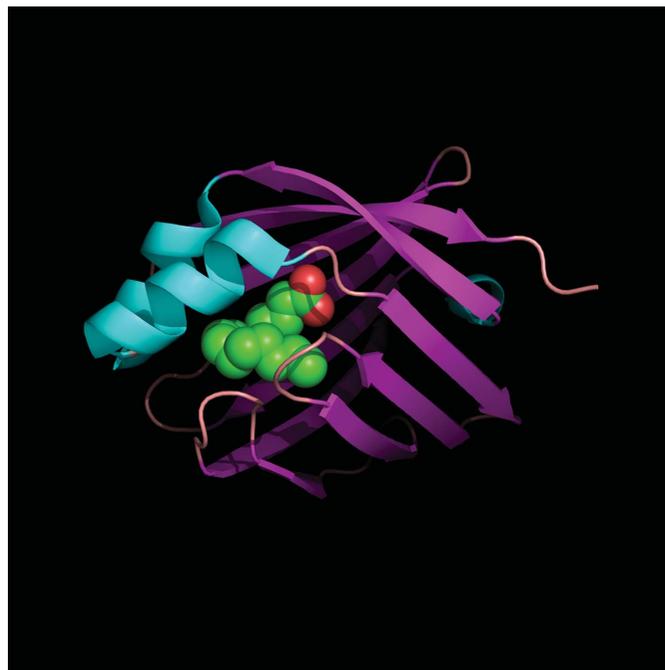
In order to assess the physiological relevance of the identified fatty acids, LC-HRMS experiments were also conducted using the protein incubated with squid nerve vesicles. As previously described (Fig. 3a), no palmitic acid was detected within the WT ReP1-NCXSQ sample.

After incubation of ReP1-NCXSQ with squid nerve vesicles, very similar LC chromatograms were obtained both in the absence and the presence of MgATP (Figs. 3b and 3c). However, in the absence of MgATP an additional peak (peak 5 in Fig. 3b) was detected that was absent in the presence of MgATP. HRMS analyses clearly identified palmitic acid as the only component present in peak 5 (Fig. 3 and Table 2). This result can be interpreted in the following way: (i) in WT ReP1-NCXSQ there is no palmitic acid (no peak 5 in Fig. 3a); (ii) after vesicle incubation without MgATP palmitic acid binds to the protein (peak 5 in Fig. 3b); (iii) after vesicle incubation with MgATP palmitic acid is no longer detected and may thus have been transferred to the exchanger (no peak 5 and narrow peak 6 in Fig. 3c).

In order to investigate the role of Tyr128 in fatty-acid binding after vesicle incubation, similar LC-HRMS experiments were performed on the Tyr128Phe mutant (Supplementary Fig. S8). Under the same conditions, mutated ReP1-NCXSQ is no longer able to bind palmitic acid. This result

**Table 3**  
Effect of WT and mutant ReP1-NCXSQ on the MgATP stimulation of Na<sup>+</sup>-gradient-dependent Ca<sup>2+</sup> uptake.

Protein type	MgATP stimulation of Na <sup>+</sup> -gradient-dependent Ca <sup>2+</sup> uptake (nmol mg <sup>-1</sup> in 10 s)
WT, no MgATP	0.51 ± 0.01
WT, with MgATP	1.25 ± 0.03
Tyr128Phe, with MgATP	0.53 ± 0.025



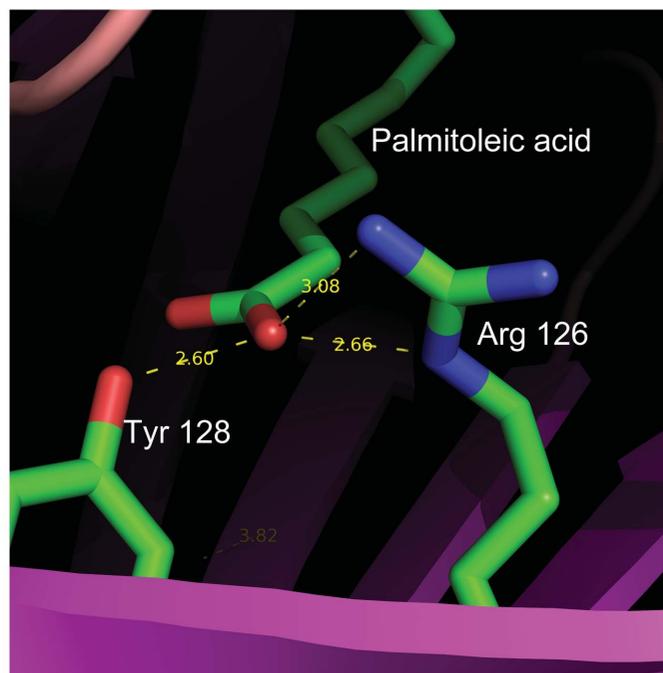
strongly suggests that Tyr128 is directly involved in palmitic acid transport.

In summary, even when the actual details of the interaction of ReP1-NCXSQ with the exchanger remain elusive (Berberian *et al.*, 2009), the LC-HRMS strongly suggests that ReP1-NCXSQ transports palmitic acid and that Tyr128 is essential for this function.

### 3.4. Crystal structures of wild-type ReP1-NCXSQ and its Tyr128Phe mutant

The crystal structures of wild-type ReP1-NCXSQ and the Tyr128Phe mutant and their abilities to bind the fatty acid inside the barrel were analyzed. The structures fold in a β-sandwich with ten β-strands, as is typical for lipid-binding proteins (Fig. 4a). For the wild type, the difference maps clearly show the presence of a lipid arising from the *E. coli* production medium which is ordered up to C12 (Fig. 4b). In this chain, the high-resolution structure (1.28 Å) of WT ReP1-NCXSQ allows the identification of a double bond between C9 and C10 (1.33 Å) of the aliphatic chain. The major conformation of the carboxylate head makes two strong contacts with Arg126 (2.66 Å) and Tyr128 (2.60 Å) (Fig. 4c). The O—O distance of 2.60 Å is close to the limit of a low-barrier hydrogen bond (2.55 Å; Schiøtt *et al.*, 1998). Such an interaction has been observed previously with a distance of 2.50 Å (Marr *et al.*, 2006).

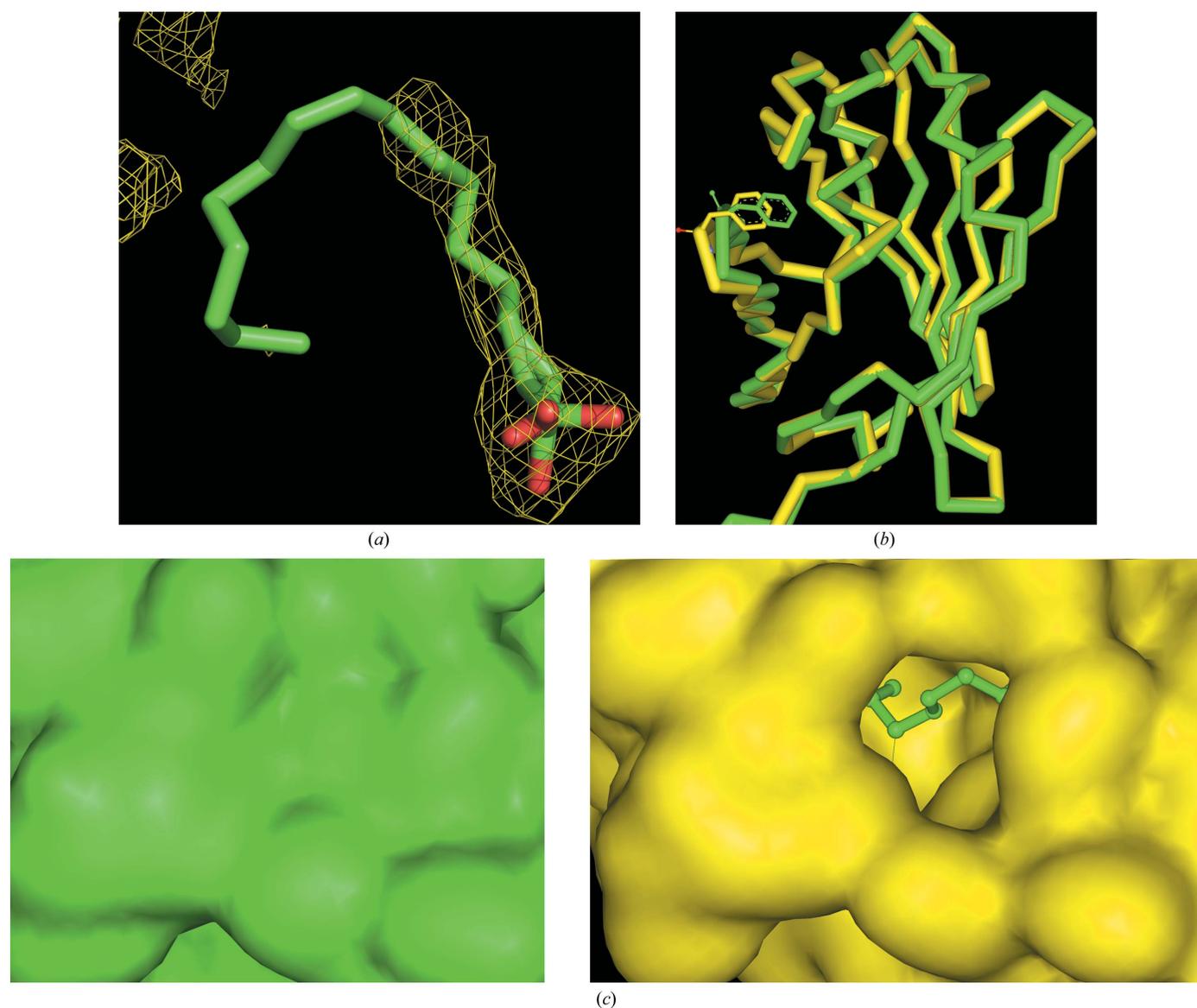
Following the indications of native mass-spectrometric and LC-HRMS experiments, electron-density signals observed for



**Figure 4**  
Crystal structure of ReP1-NCXSQ. (a) Ribbon representation of the structure of WT ReP1-NCXSQ with the palmitoleic acid shown as a CPK model. Note that the structure folds in a β-sandwich with ten β-strands that is typical of lipid-binding proteins. (b) Difference map of WT ReP1-NCXSQ showing the presence of palmitoleic acid, contoured at 3.0σ. Note the clear signal for the carboxylate head and the aliphatic chain up to C12 and the disorder at the end of the aliphatic chain. (c) Hydrogen-bond contacts of the head of palmitoleic acid with Arg126 and Tyr128.

ReP1-NCXSQ were interpreted as the disordered tail (C13–C16) of palmitoleic acid. This disorder had previously been observed with oleic acid (Sacchetti *et al.*, 1992). This fatty acid tightly fits the interior of the protein. Similar observations have been made with palmitic acid (Sacchetti *et al.*, 1989; Scapin *et al.*, 1993; Jakobsson *et al.*, 2003). The fact that the fatty acid in the interior of the ReP1-NCXSQ barrel seems to be stabilized by a relatively short hydrogen bond connecting to Tyr128 offers the possibility of investigating whether this bond is necessary for promotion of the stimulation of Na<sup>+</sup>/Ca<sup>2+</sup> exchanger by MgATP. The functional analysis results described in Table 3 show that the MgATP up-regulation of the exchanger is strongly diminished by Tyr128Phe-mutated ReP1-NCXSQ protein compared with the WT.

The Tyr128Phe mutant crystal structure provides difference maps showing a signal for the palmitoleic acid, with disorder of the carboxylate head and of the aliphatic chain C7–C16 (Fig. 5*a*). C atoms C7–C12, which are clearly seen in the WT map, are not seen in the Tyr128Phe mutant. The carboxylate head is significantly more disordered than in the WT structure. Fig. 5(*b*) shows a comparison between the WT and Tyr128Phe mutant structures, showing that they are very similar except in the loop containing Phe58; this amino-acid residue, which is a main component of the so-called portal region, is displaced by almost 5 Å. This can be better observed in Figs. 5(*c*) and 5(*d*), where the movement of Phe58 opens a channel leading into the cavity. This flexibility of the portal region has previously been observed in other structures (Reese-Wagoner *et al.*, 1999;



**Figure 5** Comparison between WT and Tyr128Phe mutant ReP1-NCXSQ. (*a*) Difference map showing the presence of palmitoleic acid in the Tyr128Phe mutant, contoured at  $3.0\sigma$ . Note the disordered signal for the carboxylate head and that the clear signal for palmitoleic acid stops after C8, showing disorder from C8 to the end of the aliphatic chain. (*b*) Comparison of the structures of WT (green) and Tyr128Phe mutant ReP1-NCXSQ (yellow) showing the C $\alpha$  trace and Phe58. (*c*) Comparison of the surface around Phe58 for WT (green, left) and Tyr128Phe (yellow, right) ReP1-NCXSQ. The palmitoleic acid (WT position) is shown in green ball-and-stick representation.

Xu *et al.*, 1993). This channel is closed in the WT–palmitoleic acid complex. In other words, in the WT structure the position of Phe58 is stabilized by a hydrophobic contact between Phe58 CE1 and C10 of palmitoleic acid (3.57 Å distance), whereas in the Tyr128Phe mutant the palmitoleic acid tail is disordered, leading to a double conformation of the Phe58 loop with equal occupancies.

It is also worth mentioning the closed position of Phe58 when bound to WT ReP1-NCXSQ. A similar position has only previously been observed when activating ligands were bound to FABP4 (Gillilan *et al.*, 2007).

## 4. Discussion

This work presents crystallographic and mass-spectrometric results on the ReP1-NCXSQ protein, which is essential for MgATP up-regulation of the squid nerve Na<sup>+</sup>/Ca<sup>2+</sup> exchanger. A sequence alignment shows that it shares primary structure with the CRABP family of the FABP superfamily. The fingerprint of the FABPs (see Furuhashi & Hotamisligil, 2008, and references therein) and most of the conserved amino acids described for FABPs from invertebrates (Esteves & Ehrlich, 2006) are present in ReP1-NCXSQ.

The presence in this protein of the characteristic binding residues involved in the interaction with ligand carboxylate head groups, a feature pertaining to several FABPs and to CRABPs, and the close phylogenetic relationships with both the vertebrate heart-type FABPs and CRBPs/CRABPs makes it difficult to assign it to either the FABPs or the CRABPs. On the other hand, the three-dimensional structures of FABPs and CRABPs are superimposable (Supplementary Fig. S7). Taking together sequence and phylogenetic data and binding results, we assume that ReP1-NCXSQ is at the frontier between the subfamilies, with several elements that lead us to place it among the CRABPs.

Recombinant ReP1-NCXSQ expressed in *E. coli* contains palmitoleic acid in the barrel. The fatty acid is held by hydrogen bonds to Arg126 and Tyr128 (shortest bond). The point mutation Tyr128Phe suggested that Tyr128 is a critical residue that is essential for fatty acid binding in the barrel and activity studies of the WT and the Tyr128Phe mutant showed that Tyr128 is necessary for MgATP activation of the Na<sup>+</sup>/Ca<sup>2+</sup> exchanger. The crystal and mass-spectrometric analyses indicate the presence of palmitoleic acid attached to Tyr128 and Arg126 by hydrogen bonds and indicate that the Tyr128Phe mutant loses, to a large extent if not completely, the capability to retain the ligand inside the barrel when purified from the expression system. This is accompanied by disorder of the palmitoleic acid and a structural change at the level of the so-called portal region (the Phe58 loop), which is displaced by almost 5 Å, opening a channel leading into the cavity.

Identification of the biological ligand is not obvious. The most likely candidates are (i) palmitic acid, which is abundant in squid axons (Zlatanov *et al.*, 2006), (ii) PtdIns(4,5)P<sub>2</sub>, which regulates the cardiac exchanger, and (iii) retinoic acid (RA), as suggested by the fact that ReP1-NCXSQ has a sequence

profile that places it next to the CRABP family. In this regard, it is interesting that ReP1-NCXSQ binds polyphosphoinositides and phosphatidic acid but does not seem to bind other membrane components such as cholesterol, sphingomyelin, phosphoserine, phosphoethanolamide or PtdIns (Berberian *et al.*, 2009).

MeCRABP was the first crustacean member of the retinoid/fatty acid-binding protein family to be cloned. A study indicated that the protein binds to both RA and retinal with similar affinity (Gu *et al.*, 2002). Note, however, that evidence for roles of retinoic signalling outside the chordates and the evolutionary origin of retinoic signalling itself still remain elusive.

Palmitic acid is indeed a good candidate for the regulatory function; there are several examples in the literature in which it acts as a regulatory biological ligand, sometimes on membrane transporters. (i) In *Saccharomyces cerevisiae* the YPTI gene encoding a RAS-like guanine nucleotide-binding protein is essential for cell viability. Mutations of two consecutive cysteines at the carboxyl-terminus, either by deletion or substitution by serines, resulted in a nonfunctional protein and also a loss of palmitoylation as estimated by [<sup>3</sup>H]-palmitic acid binding (Molenaar *et al.*, 1988). (ii) In rat L6 cells, saturated fatty acids, including palmitate, induce peripheral insulin resistance by impairing GLUT4 translocation in response to insulin (Tsuchiya *et al.*, 2010). (iii) In the skeletal muscle L6 cell line, the acute effects of palmitic acid as followed by immunofluorescence changes included rapidly induced GLUT4 translocation and stimulated glucose uptake (Pu *et al.*, 2011). (iv) In mammalian heart, inhibition of the sodium pump by phospholemman (PLM) required palmitoylation of the PLM; the absence of palmitoylation rendered PLM ineffective as, unexpectedly, did reduced PLM phosphorylation. Therefore, palmitoylation of all FXD proteins, which have conserved intracellular cysteines, may be a universal means of regulating the sodium pump (Tulloch *et al.*, 2011). (v) In neurons, dopamine transporters (DATs) are palmitoylated. Inhibition of palmitic acid incorporation in DAT synaptosomes with 2-bromopalmitate caused a marked inhibition of transport  $V_{max}$  with no losses of DAT protein or changes in DAT surface levels; this indicated a direct effect of the loss of palmitoylation by reducing transport kinetics.

The key experiment for identifying the biological ligand is LC-HRMS analysis of the protein incubated with nerve vesicles with and without MgATP. The resulting LC chromatograms are almost exactly superposable (Figs. 3b and 3c), with a striking difference in the peak corresponding to palmitic acid, which disappears for the protein incubated with nerve vesicles and MgATP. These results show that WT ReP1-NCXSQ does not bind palmitic acid, but palmitic acid is trapped during the incubation with nerve vesicles. In the presence of MgATP and vesicles the palmitic acid is released, probably to the exchanger. This suggests that palmitic acid may play a crucial role in the activity of ReP1-NCXSQ, but the mechanism of the regulation remains unknown, in particular how MgATP stimulates both lipid release and activation of the Na<sup>+</sup>/Ca<sup>2+</sup> exchanger.

In conclusion, the present findings suggest a novel mechanism based on the transport of palmitic acid for fine-tuning of the regulation of NCXSQ1 by ReP1-NCXSQ.

This work was supported by grants from the US National Science Foundation (MCB 0444598), Fondo Nacional para Investigaciones Científicas y Tecnológicas (PICT-05-38073) and Consejo Nacional de Investigaciones Científicas Y Técnicas (PIP 2010–2012: GI11220090100063), Argentina, Fondo Nacional para Ciencia y Técnica (S1-9900009046 and G-2001000637) and Fundación Polar, Venezuela. We thank the IGBMC Structural Genomics Platform staff (in particular Pierre Poussin) and l'Établissement Français du Sang-Alsace (in particular Philippe Ohlmann and Christian Gacher). This work was supported by CNRS, INSERM, Université de Strasbourg, Région Alsace and Hôpital Civil de Strasbourg. We also thank Myriam Siravegna and Dr Daniel Raimunda (INIMEC–CONICET) for their helpful expertise. DA was supported by a studentship from the French Ministère de la Recherche. TP was supported by Russian Foundation for Basic Research (grant number RFBR 10-04-00254-a).

## References

- Adams, P. D., Grosse-Kunstleve, R. W., Hung, L.-W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K. & Terwilliger, T. C. (2002). *Acta Cryst.* **D58**, 1948–1954.
- Beaugé, L., Delgado, D., Rojas, H., Berberían, G. & DiPolo, R. (1996). *Ann. N. Y. Acad. Sci.* **779**, 208–216.
- Berberían, G., Bollo, M., Montich, G., Roberts, G., Degiorgis, J. A., DiPolo, R. & Beaugé, L. (2009). *Biochim. Biophys. Acta*, **1788**, 1255–1262.
- Berberían, G., DiPolo, R. & Beaugé, L. (2007). *Ann. N. Y. Acad. Sci.* **1099**, 152–165.
- Berberían, G., Podjarny, A., DiPolo, R. & Beaugé, L. (2012). *Prog. Biophys. Mol. Biol.* **108**, 47–63.
- Blaustein, M. P. & Lederer, W. J. (1999). *Physiol. Rev.* **79**, 763–854.
- Bourguet, W., Andry, V., Iltis, C., Klaholz, B., Potier, N., Van Dorsselaer, A., Chambon, P., Gronemeyer, H. & Moras, D. (2000). *Protein Expr. Purif.* **19**, 284–288.
- DiPolo, R. & Beaugé, L. (2002). *J. Physiol.* **539**, 791–803.
- DiPolo, R. & Beaugé, L. (2006). *Physiol. Rev.* **86**, 155–203.
- DiPolo, R., Berberían, G., Delgado, D., Rojas, H. & Beaugé, L. (1997). *FEBS Lett.* **401**, 6–10.
- Doering, A. E. & Lederer, W. J. (1994). *J. Physiol.* **480**, 9–20.
- Elvir, L., Zagnoni, I., Careri, M., Cavazzini, D. & Rossi, G. L. (2001). *Rapid Commun. Mass Spectrom.* **15**, 2186–2192.
- Esteves, A. & Ehrlich, R. (2006). *Comp. Biochem. Physiol. C Toxicol. Pharmacol.* **142**, 262–274.
- Folli, C., Ramazzina, I., Percudani, R. & Berni, R. (2005). *Biochim. Biophys. Acta*, **1747**, 229–237.
- Furuhashi, M. & Hotamisligil, G. S. (2008). *Nature Rev. Drug Discov.* **7**, 489–503.
- Gillilan, R. E., Ayers, S. D. & Noy, N. (2007). *J. Mol. Biol.* **372**, 1246–1260.
- Gu, P.-L., Gunawardene, Y. I. N. S., Chow, B. C. K., He, J. G. & Chan, S.-M. (2002). *Gene*, **288**, 77–84.
- Hilge, M., Aelen, J. & Vuister, G. W. (2006). *Mol. Cell*, **22**, 15–25.
- Hilgemann, D. W., Matsuoka, S., Nagel, G. A. & Collins, A. (1992). *J. Gen. Physiol.* **100**, 905–932.
- Hunter, D. J. B., Macmaster, R., Roszak, A. W., Riboldi-Tunnicliffe, A., Griffiths, I. R. & Freer, A. A. (2005). *Acta Cryst.* **D61**, 1067–1071.
- Jakobsson, E., Alvite, G., Bergfors, T., Esteves, A. & Kleywegt, G. J. (2003). *Biochim. Biophys. Acta*, **1649**, 40–50.
- Kumar, S., Tamura, K. & Nei, M. (1993). *Comput. Appl. Biosci.* **10**, 189–191.
- Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. & Higgins, D. G. (2007). *Bioinformatics*, **23**, 2947–2948.
- Marr, E., Tardie, M., Carty, M., Brown Phillips, T., Wang, I.-K., Soeller, W., Qiu, X. & Karam, G. (2006). *Acta Cryst.* **F62**, 1058–1060.
- Molenaar, C. M., Prange, R. & Gallwitz, D. (1988). *EMBO J.* **7**, 971–976.
- Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* **D67**, 355–367.
- Navaza, J. (1994). *Acta Cryst.* **A50**, 157–163.
- Nicoll, D. A., Sawaya, M. R., Kwon, S., Cascio, D., Philipson, K. D. & Abramson, J. (2006). *J. Biol. Chem.* **281**, 21577–21581.
- Ono, T. & Odani, S. (2010). *Proc. Jpn Acad. Ser. B Phys. Biol. Sci.* **86**, 220–228.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Perrakis, A., Sixma, T. K., Wilson, K. S. & Lamzin, V. S. (1997). *Acta Cryst.* **D53**, 448–455.
- Potier, N., Billas, I. M. L., Steinmetz, A., Schaeffer, C., van Dorsselaer, A., Moras, D. & Renaud, J.-P. (2003). *Protein Sci.* **12**, 725–733.
- Pu, J., Peng, G., Li, L., Na, H., Liu, Y. & Liu, P. (2011). *J. Lipid Res.* **52**, 1319–1327.
- Reese-Wagoner, A., Thompson, J. & Banaszak, L. (1999). *Biochim. Biophys. Acta*, **1441**, 106–116.
- Sacchettini, J. C., Gordon, J. I. & Banaszak, L. J. (1989). *J. Mol. Biol.* **208**, 327–339.
- Sacchettini, J. C., Scapin, G., Gopaul, D. & Gordon, J. I. (1992). *J. Biol. Chem.* **267**, 23534–23545.
- Scapin, G., Young, A. C., Kromminga, A., Veerkamp, J. H., Gordon, J. I. & Sacchettini, J. C. (1993). *Mol. Cell Biochem.* **123**, 3–13.
- Schiøtt, B., Iversen, B. B., Madsen, G. K. H. & Bruice, T. C. (1998). *J. Am. Chem. Soc.* **120**, 12117–12124.
- Tsuchiya, Y., Hatakeyama, H., Emoto, N., Wagatsuma, F., Matsushita, S. & Kanzaki, M. (2010). *J. Biol. Chem.* **285**, 34371–34381.
- Tulloch, L. B., Howie, J., Wypijewski, K. J., Wilson, C. R., Bernard, W. G., Shattock, M. J. & Fuller, W. (2011). *J. Biol. Chem.* **286**, 36020–36031.
- Urquiza, A. M. de, Liu, S., Sjöberg, M., Zetterström, R. H., Griffiths, W., Sjövall, J. & Perlmann, T. (2000). *Science*, **290**, 2140–2144.
- Xu, Z., Bernlohns, D. & Banaszak, L. (1993). *J. Biol. Chem.* **268**, 7874–7884.
- Zlatanos, S., Laskaridis, K., Feist, C. & Sagredos, A. (2006). *Mol. Nutr. Food Res.* **50**, 967–970.



## Références Bibliographiques

1. HUPO, <http://www.hupo.org/research/hpp/>.
2. Service, R. F., Proteomics. Proteomics ponders prime time. *Science* **2008**, 321 (5897), 1758-1761.
3. Anderson, N. L., Anderson, N. G., Pearson, T. W., Borchers, C. H., Paulovich, A. G., Patterson, S. D., Gillette, M., Aebersold, R., and Carr, S. A., A human proteome detection and quantitation project. *Mol Cell Proteomics* **2009**, 8 (5), 883-886.
4. Initial sequencing and analysis of the human genome. *Nature* **2001**, 409 (6822), 860-921.
5. Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., Skupski, M., Subramanian, G., Thomas, P. D., Zhang, J., Gabor Miklos, G. L., Nelson, C., Broder, S., Clark, A. G., Nadeau, J., McKusick, V. A., Zinder, N., Levine, A. J., Roberts, R. J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Francesco, V. D., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A. E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T. J., Higgins, M. E., Ji, R.-R., Ke, Z., Ketchum, K. A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G. V., Milshina, N., Moore, H. M., Naik, A. K., Narayan, V. A., Neelam, B., Nuskern, D., Rusch, D. B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z. Y., Wang, A., Wang, X., Wang, J., Wei, M.-H., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S. C., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M. L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y.-H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N. N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J. F., Guigó, R., Campbell, M. J., Sjolander, K. V., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y.-H., Coyne, M., Dahlke, C., Mays, A. D., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M.,

- Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A., and Zhu, X., The Sequence of the Human Genome. *Science* **2001**, *291* (5507), 1304-1351.
6. The Universal Protein Resource (UniProt) 2009. *Nucleic Acids Res* **2009**, *37* (Database issue), D169-174.
  7. Clamp, M., Fry, B., Kamal, M., Xie, X., Cuff, J., Lin, M. F., Kellis, M., Lindblad-Toh, K., and Lander, E. S., Distinguishing protein-coding and noncoding genes in the human genome. *Proc Natl Acad Sci U S A* **2007**, *104* (49), 19428-19433.
  8. The call of the human proteome. *Nat Methods* **2010**, *7* (9), 661.
  9. A gene-centric human proteome project: HUPO--the Human Proteome organization. *Mol Cell Proteomics* **2010**, *9* (2), 427-429.
  10. Venter, J. C., Multiple personal genomes await. *Nature* **2010**, *464* (7289), 676-677.
  11. Legrain, P., Aebersold, R., Archakov, A., Bairoch, A., Bala, K., Beretta, L., Bergeron, J., Borchers, C., Corthals, G. L., Costello, C. E., Deutsch, E. W., Domon, B., Hancock, W., He, F., Hochstrasser, D., Marko-Varga, G., Salekdeh, G. H., Sechi, S., Snyder, M., Srivastava, S., Uhlen, M., Hu, C. H., Yamamoto, T., Paik, Y. K., and Omenn, G. S., The human proteome project: Current state and future direction. *Mol Cell Proteomics* **2011**.
  12. Hochstrasser, D., Should the Human Proteome Project be gene- or protein-centric? *J Proteome Res* **2008**, *7* (12), 5071.
  13. Rabilloud, T., Hochstrasser, D., and Simpson, R. J., Is a gene-centric human proteome project the best way for proteomics to serve biology? *Proteomics* **2010**, *10* (17), 3067-3072.
  14. Picotti, P., Bodenmiller, B., Mueller, L. N., Domon, B., and Aebersold, R., Full dynamic range proteome analysis of *S. cerevisiae* by targeted proteomics. *Cell* **2009**, *138* (4), 795-806.
  15. Taylor, R. W., and Turnbull, D. M., Mitochondrial DNA mutations in human disease. *Nat Rev Genet* **2005**, *6* (5), 389-402.
  16. Rodolfo, C., Ciccocanti, F., Giacomo, G. D., Piacentini, M., and Fimia, G. M., Proteomic analysis of mitochondrial dysfunction in neurodegenerative diseases. *Expert Review of Proteomics* **2010**, *7* (4), 519-542.
  17. Singh, K. K., Kulawiec, M., Still, I., Desouki, M. M., Geradts, J., and Matsui, S., Inter-genomic cross talk between mitochondria and the nucleus plays an important role in tumorigenesis. *Gene* **2005**, *354*, 140-146.
  18. Rabilloud, T., Kieffer, S., Procaccio, V., Louwagie, M., Courchesne, P. L., Patterson, S. D., Martinez, P., Garin, J., and Lunardi, J., Two-dimensional electrophoresis of human placental mitochondria and protein identification by mass spectrometry: toward a human mitochondrial proteome. *Electrophoresis* **1998**, *19* (6), 1006-1014.
  19. Mootha, V. K., Bunkenborg, J., Olsen, J. V., Hjerrild, M., Wisniewski, J. R., Stahl, E., Bolouri, M. S., Ray, H. N., Sihag, S., Kamal, M., Patterson, N., Lander, E. S., and Mann, M., Integrated analysis of protein composition, tissue diversity, and gene regulation in mouse mitochondria. *Cell* **2003**, *115* (5), 629-640.
  20. Foster, L. J., de Hoog, C. L., Zhang, Y., Xie, X., Mootha, V. K., and Mann, M., A mammalian organelle map by protein correlation profiling. *Cell* **2006**, *125* (1), 187-199.
  21. Pagliarini, D. J., Calvo, S. E., Chang, B., Sheth, S. A., Vafai, S. B., Ong, S. E., Walford, G. A., Sugiana, C., Boneh, A., Chen, W. K., Hill, D. E., Vidal, M., Evans, J. G., Thorburn, D. R., Carr, S. A., and Mootha, V. K., A mitochondrial protein compendium elucidates complex I disease biology. *Cell* **2008**, *134* (1), 112-123.
  22. Vogle, F. N., Wortelkamp, S., Zahedi, R. P., Becker, D., Leidhold, C., Gevaert, K., Kellermann, J., Voos, W., Sickmann, A., Pfanner, N., and Meisinger, C., Global analysis of the mitochondrial N-proteome identifies a processing peptidase critical for protein stability. *Cell* **2009**, *139* (2), 428-439.
  23. Palmfeldt, J., Vang, S., Stenbroen, V., Pedersen, C. B., Christensen, J. H., Bross, P., and Gregersen, N., Mitochondrial proteomics on human fibroblasts for identification of metabolic imbalance and cellular stress. *Proteome Sci* **2009**, *7*, 20.

24. Taylor, S. W., Fahy, E., Zhang, B., Glenn, G. M., Warnock, D. E., Wiley, S., Murphy, A. N., Gaucher, S. P., Capaldi, R. A., Gibson, B. W., and Ghosh, S. S., Characterization of the human heart mitochondrial proteome. *Nat Biotechnol* **2003**, *21* (3), 281-286.
25. Ferreira, R., Vitorino, R., Alves, R. M., Appell, H. J., Powers, S. K., Duarte, J. A., and Amado, F., Subsarcolemmal and intermyofibrillar mitochondria proteome differences disclose functional specializations in skeletal muscle. *Proteomics* **2010**, *10* (17), 3142-3154.
26. Dolezal, P., Likic, V., Tachezy, J., and Lithgow, T., Evolution of the molecular machines for protein import into mitochondria. *Science* **2006**, *313* (5785), 314-318.
27. Neupert, W., and Brunner, M., The protein import motor of mitochondria. *Nat Rev Mol Cell Biol* **2002**, *3* (8), 555-565.
28. Rehling, P., Brandner, K., and Pfanner, N., Mitochondrial import and the twin-pore translocase. *Nat Rev Mol Cell Biol* **2004**, *5* (7), 519-530.
29. Bolender, N., Sickmann, A., Wagner, R., Meisinger, C., and Pfanner, N., Multiple pathways for sorting mitochondrial precursor proteins. *EMBO Rep* **2008**, *9* (1), 42-49.
30. Claros, M. G., and Vincens, P., Computational method to predict mitochondrially imported proteins and their targeting sequences. *Eur J Biochem* **1996**, *241* (3), 779-786.
31. Gaston, D., Tsaousis, A. D., and Roger, A. J., Predicting proteomes of mitochondria and related organelles from genomic and expressed sequence tag data. *Methods Enzymol* **2009**, *457*, 21-47.
32. Anderson, L., and Seilhamer, J., A comparison of selected mRNA and protein abundances in human liver. *Electrophoresis* **1997**, *18* (3-4), 533-537.
33. Zamzami, N., and Kroemer, G., The mitochondrion in apoptosis: how Pandora's box opens. *Nat Rev Mol Cell Biol* **2001**, *2* (1), 67-71.
34. Emanuelsson, O., Brunak, S., von Heijne, G., and Nielsen, H., Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc* **2007**, *2* (4), 953-971.
35. Halligan, B. D., Geiger, J. F., Vallejos, A. K., Greene, A. S., and Twigger, S. N., Low cost, scalable proteomics data analysis using Amazon's cloud computing services and open source search algorithms. *J Proteome Res* **2009**, *8* (6), 3148-3153.
36. Geer, L. Y., Markey, S. P., Kowalak, J. A., Wagner, L., Xu, M., Maynard, D. M., Yang, X., Shi, W., and Bryant, S. H., Open Mass Spectrometry Search Algorithm. *Journal of Proteome Research* **2004**, *3* (5), 958-964.
37. Carr, S., Aebersold, R., Baldwin, M., Burlingame, A., Clauser, K., and Nesvizhskii, A., The need for guidelines in publication of peptide and protein identification data: Working Group on Publication Guidelines for Peptide and Protein Identification Data. *Mol Cell Proteomics* **2004**, *3* (6), 531-533.
38. Latterich, M., Publishing proteomic data. *Proteome Sci* **2006**, *4*, 8.
39. Van Kuilenburg, A. B. P., Van Beeumen, J. J., Van Der Meer, N. M., and Muijsers, A. O., Subunits VIIa,b,c of human cytochrome c oxidase. *European Journal of Biochemistry* **1992**, *203* (1-2), 193-199.
40. Wickham, M. E., Rug, M., Ralph, S. A., Klonis, N., McFadden, G. I., Tilley, L., and Cowman, A. F., Trafficking and assembly of the cytoadherence complex in Plasmodium falciparum-infected human erythrocytes. *EMBO J* **2001**, *20* (20), 5636-5649.
41. Lopez-Estraño, C., Bhattacharjee, S., Harrison, T., and Haldar, K., Cooperative domains define a unique host cell-targeting signal in Plasmodium falciparum-infected erythrocytes. *Proceedings of the National Academy of Sciences of the United States of America* **2003**, *100* (21), 12402-12407.
42. Sargeant, T. J., Marti, M., Caler, E., Carlton, J. M., Simpson, K., Speed, T. P., and Cowman, A. F., Lineage-specific expansion of proteins exported to erythrocytes in malaria parasites. *Genome Biol* **2006**, *7* (2), R12.
43. Marti, M., Good, R. T., Rug, M., Knuepfer, E., and Cowman, A. F., Targeting malaria virulence and remodeling proteins to the host erythrocyte. *Science* **2004**, *306* (5703), 1930-1933.

44. Chang, H. H., Falick, A. M., Carlton, P. M., Sedat, J. W., DeRisi, J. L., and Marletta, M. A., N-terminal processing of proteins exported by malaria parasites. *Mol Biochem Parasitol* **2008**, *160* (2), 107-115.
45. Boddey, J. A., Moritz, R. L., Simpson, R. J., and Cowman, A. F., Role of the Plasmodium Export Element in Trafficking Parasite Proteins to the Infected Erythrocyte. *Traffic* **2009**, *10* (3), 285-299.
46. Joannin, N., Kallberg, Y., Wahlgren, M., and Persson, B., RSpred, a set of Hidden Markov Models to detect and classify the RIFIN and STEVOR proteins of Plasmodium falciparum. *BMC Genomics* **2011**, *12*, 119.
47. Schmelzle, K., and White, F. M., Phosphoproteomic approaches to elucidate cellular signaling networks. *Curr Opin Biotechnol* **2006**, *17* (4), 406-414.
48. (2011) World Malaria Report 2011. World Health Organization
49. Murray, C. J., Rosenfeld, L. C., Lim, S. S., Andrews, K. G., Foreman, K. J., Haring, D., Fullman, N., Naghavi, M., Lozano, R., and Lopez, A. D., Global malaria mortality between 1980 and 2010: a systematic analysis. *The Lancet* **2012**, *379* (9814), 413-431.
50. Greenwood, B., Bojang, K., Whitty, C., and Targett, G., Malaria. *Lancet* **2005**, *365* (9469), 1487-1498.
51. Gething, P., Patil, A., Smith, D., Guerra, C., Elyazar, I., Johnston, G., Tatem, A., and Hay, S., A new world malaria map: Plasmodium falciparum endemicity in 2010. *Malaria Journal* **2011**, *10* (1), 378.
52. Hay, S. I., and Snow, R. W., The Malaria Atlas Project: Developing Global Maps of Malaria Risk. *PLoS Med* **2006**, *3* (12), e473.
53. Sallares, R., Role of environmental changes in the spread of malaria in Europe during the Holocene. *Quaternary International* **2006**, *150* (1), 21-27.
54. Poncon, N., Toty, C., L'Ambert, G., Le Goff, G., Brengues, C., Schaffner, F., and Fontenille, D., Biology and dynamics of potential malaria vectors in Southern France. *Malaria Journal* **2007**, *6* (1), 18.
55. Lindsay, S., and Thomas, C., Global Warming and Risk of Vivax Malaria in Great Britain. *Global Change and Human Health* **2001**, *2* (1), 80-84.
56. Sainz-Elipse, S., Latorre, J., Escosa, R., Masia, M., Fuentes, M., Mas-Coma, S., and Bargues, M., Malaria resurgence risk in southern Europe: climate assessment in an historically endemic area of rice fields at the Mediterranean shore of Spain. *Malaria Journal* **2010**, *9* (1), 221.
57. Patz, J., P. Martens, Ph.D.: Health and Climate Change: Modelling the Impacts of Global Warming and Ozone Depletion. *Climatic Change* **1998**, *39* (1), 39-46.
58. Reiter, P., Thomas, C. J., Atkinson, P. M., Hay, S. I., Randolph, S. E., Rogers, D. J., Shanks, G. D., Snow, R. W., and Spielman, A., Global warming and malaria: a call for accuracy. *The Lancet Infectious Diseases* **2004**, *4* (6), 323-324.
59. Rogers, D. J., and Randolph, S. E., The Global Spread of Malaria in a Future, Warmer World. *Science* **2000**, *289* (5485), 1763-1766.
60. Takken, W., and Knols, B. G. J. (eds). (2007) *Emerging Pests and Vector-borne Diseases in Europe*, Vol. 1, hardback
61. Joy, D. A., Feng, X., Mu, J., Furuya, T., Chotivanich, K., Krettli, A. U., Ho, M., Wang, A., White, N. J., Suh, E., Beerli, P., and Su, X.-z., Early Origin and Recent Expansion of Plasmodium falciparum. *Science* **2003**, *300* (5617), 318-321.
62. <http://www.dpd.cdc.gov/dpdx/Default.htm>. Centre de Contrôle et de Prévention des Maladies, Etats-Unis d'Amérique
63. Miller, L. H., Baruch, D. I., Marsh, K., and Doumbo, O. K., The pathogenic basis of malaria. *Nature* **2002**, *415* (6872), 673-679.
64. Maier, A. G., Cooke, B. M., Cowman, A. F., and Tilley, L., Malaria parasite proteins that remodel the host erythrocyte. *Nat Rev Microbiol* **2009**, *7* (5), 341-354.
65. Maier, A. G., Rug, M., O'Neill, M. T., Brown, M., Chakravorty, S., Szeszak, T., Chesson, J., Wu, Y., Hughes, K., Coppel, R. L., Newbold, C., Beeson, J. G., Craig, A., Crabb, B. S., and Cowman,

- A. F., Exported Proteins Required for Virulence and Rigidity of Plasmodium falciparum-Infected Human Erythrocytes. *Cell* **2008**, *134* (1), 48-61.
66. Craig, A., and Scherf, A., Molecules on the surface of the Plasmodium falciparum infected erythrocyte and their role in malaria pathogenesis and immune evasion. *Mol Biochem Parasitol* **2001**, *115* (2), 129-143.
67. Lanzer, M., Wickert, H., Krohne, G., Vincensini, L., and Braun Breton, C., Maurer's clefts: a novel multi-functional organelle in the cytoplasm of Plasmodium falciparum-infected erythrocytes. *Int J Parasitol* **2006**, *36* (1), 23-36.
68. Haldar, K., Kamoun, S., Hiller, N. L., Bhattacharje, S., and van Ooij, C., Common infection strategies of pathogenic eukaryotes. *Nat Rev Micro* **2006**, *4* (12), 922-931.
69. Dluzewski, A. R., Fryer, P. R., Griffiths, S., Wilson, R. J., and Gratzer, W. B., Red cell membrane protein distribution during malarial invasion. *J Cell Sci* **1989**, *92* ( Pt 4), 691-699.
70. Dluzewski, A. R., Mitchell, G. H., Fryer, P. R., Griffiths, S., Wilson, R. J., and Gratzer, W. B., Origins of the parasitophorous vacuole membrane of the malaria parasite, Plasmodium falciparum, in human red blood cells. *J Cell Sci* **1992**, *102* ( Pt 3), 527-532.
71. McPherson, R. A., Donald, D. R., Sawyer, W. H., and Tilley, L., Proteolytic digestion of band 3 at an external site alters the erythrocyte membrane organisation and may facilitate malarial invasion. *Mol Biochem Parasitol* **1993**, *62* (2), 233-242.
72. Roggwiler, E., Betoulle, M. E., Blisnick, T., and Braun Breton, C., A role for erythrocyte band 3 degradation by the parasite gp76 serine protease in the formation of the parasitophorous vacuole during invasion of erythrocytes by Plasmodium falciparum. *Mol Biochem Parasitol* **1996**, *82* (1), 13-24.
73. Wickramarachchi, T., Devi, Y. S., Mohammed, A., and Chauhan, V. S., Identification and Characterization of a Novel Plasmodium falciparum Merozoite Apical Protein Involved in Erythrocyte Binding and Invasion. *PLoS One* **2008**, *3* (3), e1732.
74. Li, X., Chen, H., Oh, S. S., and Chishti, A. H., A Presenilin-like protease associated with Plasmodium falciparum micronemes is involved in erythrocyte invasion. *Mol Biochem Parasitol* **2008**, *158* (1), 22-31.
75. Singh, S., Plassmeyer, M., Gaur, D., and Miller, L. H., Mononeme: a new secretory organelle in Plasmodium falciparum merozoites identified by localization of rhomboid-1 protease. *Proc Natl Acad Sci U S A* **2007**, *104* (50), 20043-20048.
76. Haldar, K., and Mohandas, N., Erythrocyte remodeling by malaria parasites. *Curr Opin Hematol* **2007**, *14* (3), 203-209.
77. Duffy, P. E., and Fried, M., Plasmodium falciparum adhesion in the placenta. *Curr Opin Microbiol* **2003**, *6* (4), 371-376.
78. Atkinson, C. T., and Aikawa, M., Ultrastructure of malaria-infected erythrocytes. *Blood Cells* **1990**, *16* (2-3), 351-368.
79. Kriek, N., Tilley, L., Horrocks, P., Pinches, R., Elford, B. C., Ferguson, D. J., Lingelbach, K., and Newbold, C. I., Characterization of the pathway for transport of the cytoadherence-mediating protein, PfEMP1, to the host cell surface in malaria parasite-infected erythrocytes. *Mol Microbiol* **2003**, *50* (4), 1215-1227.
80. Bhattacharjee, S., van Ooij, C., Balu, B., Adams, J. H., and Haldar, K., Maurer's clefts of Plasmodium falciparum are secretory organelles that concentrate virulence protein reporters for delivery to the host erythrocyte. *Blood* **2008**, *111* (4), 2418-2426.
81. Kirchhausen, T., Three ways to make a vesicle. *Nat Rev Mol Cell Biol* **2000**, *1* (3), 187-198.
82. Wickner, W., and Schekman, R., Protein translocation across biological membranes. *Science* **2005**, *310* (5753), 1452-1456.
83. Burghaus, P. A., and Lingelbach, K., Luciferase, when fused to an N-terminal signal peptide, is secreted from transfected Plasmodium falciparum and transported to the cytosol of infected erythrocytes. *J Biol Chem* **2001**, *276* (29), 26838-26845.

84. Waller, R. F., Reed, M. B., Cowman, A. F., and McFadden, G. I., Protein trafficking to the plastid of *Plasmodium falciparum* is via the secretory pathway. *EMBO J* **2000**, *19* (8), 1794-1802.
85. Przyborski, J. M., and Lanzer, M., Protein transport and trafficking in *Plasmodium falciparum*-infected erythrocytes. *Parasitology* **2005**, *130* (Pt 4), 373-388.
86. Hiller, N. L., Bhattacharjee, S., van Ooij, C., Liolios, K., Harrison, T., Lopez-Estrano, C., and Haldar, K., A host-targeting signal in virulence proteins reveals a secretome in malarial infection. *Science* **2004**, *306* (5703), 1934-1937.
87. Przyborski, J. M., Miller, S. K., Pfahler, J. M., Henrich, P. P., Rohrbach, P., Crabb, B. S., and Lanzer, M., Trafficking of STEVOR to the Maurer's clefts in *Plasmodium falciparum*-infected erythrocytes. *EMBO J* **2005**, *24* (13), 2306-2317.
88. Spycher, C., Rug, M., Klonis, N., Ferguson, D. J. P., Cowman, A. F., Beck, H.-P., and Tilley, L., Genesis of and Trafficking to the Maurer's Clefts of *Plasmodium falciparum*-Infected Erythrocytes. *Molecular and Cellular Biology* **2006**, *26* (11), 4074-4085.
89. Sargeant, T., Marti, M., Caler, E., Carlton, J., Simpson, K., Speed, T., and Cowman, A., Lineage-specific expansion of proteins exported to erythrocytes in malaria parasites. *Genome Biology* **2006**, *7* (2), R12.
90. Lingelbach, K., and Joiner, K. A., The parasitophorous vacuole membrane surrounding *Plasmodium* and *Toxoplasma*: an unusual compartment in infected cells. *J Cell Sci* **1998**, *111* (Pt 11), 1467-1475.
91. Przyborski, J., and Lanzer, M., Parasitology. The malarial secretome. *Science* **2004**, *306* (5703), 1897-1898.
92. Crabb, B. S., de Koning-Ward, T. F., and Gilson, P. R., Protein export in *Plasmodium* parasites: from the endoplasmic reticulum to the vacuolar export machine. *Int J Parasitol* **2010**, *40* (5), 509-513.
93. Lingelbach, K., and Przyborski, J. M., The long and winding road: protein trafficking mechanisms in the *Plasmodium falciparum* infected erythrocyte. *Mol Biochem Parasitol* **2006**, *147* (1), 1-8.
94. Bullen, H. E., Charnaud, S. C., Kalanon, M., Riglar, D. T., Dekiwadia, C., Kangwanrangsan, N., Torii, M., Tsuboi, T., Baum, J., Ralph, S. A., Cowman, A. F., de Koning-Ward, T. F., Crabb, B. S., and Gilson, P. R., Biosynthesis, localization, and macromolecular arrangement of the *Plasmodium falciparum* translocon of exported proteins (PTEX). *J Biol Chem* **2012**, *287* (11), 7871-7884.
95. Cohen, G. M., Caspases: the executioners of apoptosis. *Biochem J* **1997**, *326* (Pt 1), 1-16.
96. Li, S. J., and Hochstrasser, M., A new protease required for cell-cycle progression in yeast. *Nature* **1999**, *398* (6724), 246-251.
97. Huttenlocher, A., Palecek, S. P., Lu, Q., Zhang, W., Mellgren, R. L., Lauffenburger, D. A., Ginsberg, M. H., and Horwitz, A. F., Regulation of cell migration by the calcium-dependent protease calpain. *J Biol Chem* **1997**, *272* (52), 32719-32722.
98. <http://ebischoff.free.fr/Palu/palu2.html>.
99. Trager, W., and Jensen, J. B., Human malaria parasites in continuous culture. *Science* **1976**, *193* (4254), 673-675.
100. Lambros, C., and Vanderberg, J. P., Synchronization of *Plasmodium falciparum* erythrocytic stages in culture. *J Parasitol* **1979**, *65* (3), 418-420.
101. Staalsoe, T., Giha, H. A., Dodoo, D., Theander, T. G., and Hviid, L., Detection of antibodies to variant antigens on *Plasmodium falciparum*-infected erythrocytes by flow cytometry. *Cytometry* **1999**, *35* (4), 329-336.
102. Ansoorge, I., Benting, J., Bhakdi, S., and Lingelbach, K., Protein sorting in *Plasmodium falciparum*-infected red blood cells permeabilized with the pore-forming protein streptolysin O. *Biochem J* **1996**, *315* (Pt 1), 307-314.
103. Fidock, D. A., Nomura, T., Talley, A. K., Cooper, R. A., Dzekunov, S. M., Ferdig, M. T., Ursos, L. M., Sidhu, A. B., Naude, B., Deitsch, K. W., Su, X. Z., Wootton, J. C., Roepe, P. D., and

- Wellems, T. E., Mutations in the *P. falciparum* digestive vacuole transmembrane protein PfCRT and evidence for their role in chloroquine resistance. *Mol Cell* **2000**, *6* (4), 861-871.
104. Reed, M. B., Saliba, K. J., Caruana, S. R., Kirk, K., and Cowman, A. F., Pgh1 modulates sensitivity and resistance to multiple antimalarials in *Plasmodium falciparum*. *Nature* **2000**, *403* (6772), 906-909.
105. Sa, J. M., Twu, O., Hayton, K., Reyes, S., Fay, M. P., Ringwald, P., and Wellems, T. E., Geographic patterns of *Plasmodium falciparum* drug resistance distinguished by differential responses to amodiaquine and chloroquine. *Proc Natl Acad Sci U S A* **2009**, *106* (45), 18883-18889.
106. Raj, D. K., Mu, J., Jiang, H., Kabat, J., Singh, S., Sullivan, M., Fay, M. P., McCutchan, T. F., and Su, X. Z., Disruption of a *Plasmodium falciparum* multidrug resistance-associated protein (PfMRP) alters its fitness and transport of antimalarial drugs and glutathione. *J Biol Chem* **2009**, *284* (12), 7687-7696.
107. Kuhn, Y., Sanchez, C. P., Ayoub, D., Saridaki, T., van Dorselaer, A., and Lanzer, M., Trafficking of the phosphoprotein PfCRT to the digestive vacuolar membrane in *Plasmodium falciparum*. *Traffic* **2010**, *11* (2), 236-249.
108. Summers, R. L., Nash, M. N., and Martin, R. E., Know your enemy: understanding the role of PfCRT in drug resistance could lead to new antimalarial tactics. *Cell Mol Life Sci* **2012**.
109. Sanchez, C. P., Dave, A., Stein, W. D., and Lanzer, M., Transporters as mediators of drug resistance in *Plasmodium falciparum*. *Int J Parasitol* **2010**, *40* (10), 1109-1118.
110. Goldberg, D. E., Slater, A. F., Cerami, A., and Henderson, G. B., Hemoglobin degradation in the malaria parasite *Plasmodium falciparum*: an ordered process in a unique organelle. *Proc Natl Acad Sci U S A* **1990**, *87* (8), 2931-2935.
111. Goldberg, D. E., Hemoglobin degradation. *Curr Top Microbiol Immunol* **2005**, *295*, 275-291.
112. Liu, J., Istvan, E. S., Gluzman, I. Y., Gross, J., and Goldberg, D. E., *Plasmodium falciparum* ensures its amino acid supply with multiple acquisition pathways and redundant proteolytic enzyme systems. *Proc Natl Acad Sci U S A* **2006**, *103* (23), 8840-8845.
113. Lew, V. L., Tiffert, T., and Ginsburg, H., Excess hemoglobin digestion and the osmotic stability of *Plasmodium falciparum*-infected red blood cells. *Blood* **2003**, *101* (10), 4189-4194.
114. Jani, D., Nagarkatti, R., Beatty, W., Angel, R., Slebodnick, C., Andersen, J., Kumar, S., and Rathore, D., HDP-a novel heme detoxification protein from the malaria parasite. *PLoS Pathog* **2008**, *4* (4), e1000053.
115. Fitch, C. D., Ferriprotoporphyrin IX, phospholipids, and the antimalarial actions of quinoline drugs. *Life Sci* **2004**, *74* (16), 1957-1972.
116. Cabrera, M., Paguio, M. F., Xie, C., and Roepe, P. D., Reduced digestive vacuolar accumulation of chloroquine is not linked to resistance to chloroquine toxicity. *Biochemistry* **2009**, *48* (47), 11152-11154.
117. Bray, P. G., Mungthin, M., Ridley, R. G., and Ward, S. A., Access to hemozoin: the basis of chloroquine resistance. *Mol Pharmacol* **1998**, *54* (1), 170-179.
118. Valderramos, S. G., and Fidock, D. A., Transporters involved in resistance to antimalarial drugs. *Trends Pharmacol Sci* **2006**, *27* (11), 594-601.
119. Sanchez, C. P., Stein, W. D., and Lanzer, M., Is PfCRT a channel or a carrier? Two competing models explaining chloroquine resistance in *Plasmodium falciparum*. *Trends Parasitol* **2007**, *23* (7), 332-339.
120. Roepe, P. D., Molecular and physiologic basis of quinoline drug resistance in *Plasmodium falciparum* malaria. *Future Microbiol* **2009**, *4* (4), 441-455.
121. Klemba, M., Beatty, W., Gluzman, I., and Goldberg, D. E., Trafficking of plasmepsin II to the food vacuole of the malaria parasite *Plasmodium falciparum*. *J Cell Biol* **2004**, *164* (1), 47-56.
122. Banerjee, R., Francis, S. E., and Goldberg, D. E., Food vacuole plasmepsins are processed at a conserved site by an acidic convertase activity in *Plasmodium falciparum*. *Mol Biochem Parasitol* **2003**, *129* (2), 157-165.

- 123.** Kim, Y. M., Lee, M. H., Piao, T. G., Lee, J. W., Kim, J. H., Lee, S., Choi, K. M., Jiang, J. H., Kim, T. U., and Park, H., Prodomain processing of recombinant plasmepsin II and IV, the aspartic proteases of *Plasmodium falciparum*, is auto- and trans-catalytic. *J Biochem* **2006**, *139* (2), 189-195.
- 124.** Kuhn, Y., Rohrbach, P., and Lanzer, M., Quantitative pH measurements in *Plasmodium falciparum*-infected erythrocytes using pHluorin. *Cell Microbiol* **2007**, *9* (4), 1004-1013.
- 125.** Subramanian, S., Sijwali, P. S., and Rosenthal, P. J., Falcipain cysteine proteases require bipartite motifs for trafficking to the *Plasmodium falciparum* food vacuole. *J Biol Chem* **2007**, *282* (34), 24961-24969.
- 126.** Dasaradhi, P. V., Korde, R., Thompson, J. K., Tanwar, C., Nag, T. C., Chauhan, V. S., Cowman, A. F., Mohammed, A., and Malhotra, P., Food vacuole targeting and trafficking of falcipain-2, an important cysteine protease of human malaria parasite *Plasmodium falciparum*. *Mol Biochem Parasitol* **2007**, *156* (1), 12-23.
- 127.** Braulke, T., and Bonifacino, J. S., Sorting of lysosomal proteins. *Biochim Biophys Acta* **2009**, *1793* (4), 605-614.
- 128.** Bonifacino, J. S., and Traub, L. M., Signals for sorting of transmembrane proteins to endosomes and lysosomes. *Annu Rev Biochem* **2003**, *72*, 395-447.
- 129.** FAO. (2010) Rapport: L'état de l'insécurité alimentaire dans le monde; Combattre l'insécurité alimentaire lors des crises prolongées. Rome
- 130.** Harris, R. B., Role of set-point theory in regulation of body weight. *The FASEB Journal* **1990**, *4* (15), 3310-3318.
- 131.** Kennedy, G. C., The role of depot fat in the hypothalamic control of food intake in the rat. *Proc R Soc Lond B Biol Sci.* **1953**, *15* (140), 578-596.
- 132.** Rigaud, D., Hassid, J., Meulemans, A., Poupard, A. T., and Boulier, A., A paradoxical increase in resting energy expenditure in malnourished patients near death: the king penguin syndrome. *Am J Clin Nutr* **2000**, *72*, 355-360.
- 133.** Groscolas, R., and Robin, J. P., Long-term fasting and re-feeding in penguins. *Comp Biochem Physiol A Mol Integr Physiol* **2001**, *128* (3), 645-655.
- 134.** Goodman, M. N., Larsen, P. R., Kaplan, M. M., Aoki, T. T., Young, V. R., and Ruderman, N. B., Starvation in the rat. II. Effect of age and obesity on protein sparing and fuel metabolism. *Am J Physiol* **1980**, *239* (4), E277-E286.
- 135.** Cherel, Y., Robin, J.-P., and Maho, Y. L., Physiology and biochemistry of long-term fasting in birds. *Canadian Journal of Zoology* **1988**, *66* (1), 159-166.
- 136.** Goodman, M., Larsen, P., Kaplan, M., Aoki, T., Young, V., and Ruderman, N., Starvation in the rat. II. Effect of age and obesity on protein sparing and fuel metabolism. *American Journal of Physiology - Endocrinology And Metabolism* **1980**, *239* (4), E277.
- 137.** Le Maho, Y., Robin, J. P., and Cherel, Y., Starvation as a treatment for obesity - the need to conserve body protein. *News Physiol. Sci.* **1988**, *3*, 21-24.
- 138.** Groscolas, R., and Robin, J.-P., Long-term fasting and re-feeding in penguins. *Comparative Biochemistry and Physiology - Part A: Molecular & Integrative Physiology* **2001**, *128* (3), 643-653.
- 139.** Groscolas, R., Decrock, F., Thil, M. A., Fayolle, C., Boissery, C., and Robin, J. P., Refeeding signal in fasting-incubating king penguins: changes in behavior and egg temperature. *Am J Physiol Regul Integr Comp Physiol* **2000**, *279* (6), R2104-2112.
- 140.** Bertile, F., and Raclot, T., ATGL and HSL are not coordinately regulated in response to fuel partitioning in fasted rats. *J Nutr Biochem* PMID: 20621463
- 141.** Bertile, F., Oudart, H., Le Maho, Y., and Raclot, T., Recombinant leptin in the hypothalamic response to late fasting. *Biochem Biophys Res Commun* **2003**, *310* (3), 949-955.
- 142.** Bertile, F., and Raclot, T., Adipose-Derived Factors During Nutritional Transitions. *Current Nutrition & Food Science* **2006**, *2* (2), 127-139.
- 143.** Bertile, F., and Raclot, T., ATGL and HSL are not coordinately regulated in response to fuel partitioning in fasted rats. *J Nutr Biochem* **2011**, *22* (4), 372-379.

144. Skulachev, V. P., Mitochondrial physiology and pathology; concepts of programmed death of organelles, cells and organisms. *Mol Aspects Med* **1999**, *20* (3), 139-184.
145. Nesvizhskii, A. I., A survey of computational methods and error rate estimation procedures for peptide and protein identification in shotgun proteomics. *Journal of Proteomics* **2010**, *73* (11), 2092-2123.
146. Domon, B., and Aebersold, R., Options and considerations when selecting a quantitative proteomics strategy. *Nat Biotechnol* **2010**, *28* (7), 710-721.
147. Potier, N., Billas, I. M., Steinmetz, A., Schaeffer, C., van Dorsselaer, A., Moras, D., and Renaud, J. P., Using nondenaturing mass spectrometry to detect fortuitous ligands in orphan nuclear receptors. *Protein Sci* **2003**, *12* (4), 725-733.
148. Bourguet, W., Andry, V., Iltis, C., Klaholz, B., Potier, N., Van Dorsselaer, A., Chambon, P., Gronemeyer, H., and Moras, D., Heterodimeric complex of RAR and RXR nuclear receptor ligand-binding domains: purification, crystallization, and preliminary X-ray diffraction analysis. *Protein Expr Purif* **2000**, *19* (2), 284-288.
149. de Urquiza, A. M., Liu, S., Sjoberg, M., Zetterstrom, R. H., Griffiths, W., Sjovall, J., and Perlmann, T., Docosahexaenoic acid, a ligand for the retinoid X receptor in mouse brain. *Science* **2000**, *290* (5499), 2140-2144.
150. Atmanène, C. (2009) Nouvelles perspective pour l'analyse et la caractérisation structurale de complexes biologiques par spectrométrie de masse supramoléculaire. in *Thèse de l'Institut Pluridisciplinaire Hubert Curien - Département des Sciences Analytiques*, Université de Strasbourg, Strasbourg
151. Ganem, B., Li, Y. T., and Henion, J. D., Detection of noncovalent receptor-ligand complexes by mass spectrometry. *Journal of the American Chemical Society* **1991**, *113* (16), 6294-6296.



## Conclusion Générale

---



## Conclusion générale

L'objectif de ce travail de thèse était principalement focalisé sur le développement et les applications des approches méthodologiques permettant une caractérisation plus fine des protéomes avec un intérêt particulier autour des processus protéolytiques.

**En ce qui concerne les développements méthodologiques**, nous avons pu améliorer et appliquer avec succès notre stratégie originale de marquage et d'enrichissement en peptides N-terminaux appelée N-TOP (N-Terminal Oriented Proteomics), et ceci au travers de trois développements méthodologiques que nous avons mis au point.

1. Tout d'abord, nous avons rendu possible la dérivation des amines N-terminales dans un gel SDS-PAGE (stratégie Ingel N-TOP) grâce à l'élimination des produits secondaires du réactif. Ceci a permis de s'affranchir ainsi des problèmes liés aux tampons non compatibles avec le marquage en solution et aux faibles quantités d'échantillons.
2. Nous avons ensuite développé une méthode de quantification relative des peptides N-terminaux des protéines afin de déterminer le taux de protéolyse de façon différentielle. Ceci a été réalisé grâce à la synthèse d'un réactif de dérivation marqué aux isotopes stables ( $^{13}\text{C}$ ). L'application de cette stratégie appelée qN-TOP (« quantitative N-TOP ») sur des protéines modèles a permis de montrer que le marquage différentiel s'effectue avec succès et que la gamme linéaire est comparable à celle décrite pour les stratégies de quantification relative par marquage aux isotopes stables (1 à 2 logs).
3. Enfin, dans le but d'avoir une meilleure séparation chromatographique entre les peptides N-terminaux dérivés et les peptides de digestion, nous avons synthétisé un réactif plus hydrophobe. Ce nouveau réactif nous a permis de déplacer, dans nos conditions chromatographiques, les peptides N-terminaux d'environ 8 minutes, par rapport au réactif classique. Ceci doit permettre d'augmenter le nombre d'identification des peptides marqués puisque la sélection par le spectromètre de masse pour la MS/MS sera maximisée.

Nous avons ainsi pu étendre avec succès le champ d'application de la stratégie N-TOP à plusieurs problématiques de biologie. **En ce qui concerne les applications de l'analyse**

**protéomique** nous avons apporté des éléments de réponse dans plusieurs projets menés en collaboration avec des biologistes.

Ainsi, **dans une étude du protéome mitochondriale humain**, nous avons pu, grâce à la stratégie N-TOP, identifier plus de 1700 protéines mitochondriales et caractériser plusieurs centaines de peptides de transit. De plus nous avons pu comparer dans le cadre du même projet deux moteurs de recherche, Mascot et OMSSA, que nous avons utilisé en parallèle pour la recherche dans les banques de données protéiques. Nos résultats ont permis de valider, corriger et annoter les peptides de transit prédit dans les banques de données. Ces résultats ont été implémentés dans la banque de données à haute valeur ajoutée « neXtprot ». En perspective, des études différentielles à travers des mitochondries issues de différents tissus biologiques sont en cours de préparation.

**Sur le parasite responsable du paludisme *Plasmodium falciparum*** trois études ont été réalisées. La première décrit l'application de notre stratégie « Ingel N-TOP » aux protéines exportées par le parasite vers sa cellule hôte. Elle a permis d'identifier les résidus nécessaires à la reconnaissance moléculaire à chaque étape du mécanisme d'export et de mettre en évidence la nécessité d'une maturation à l'aide d'un clivage protéolytique et d'une acétylation de l'amine N-terminale de la protéine maturée. Dans la deuxième étude, nous avons caractérisé le protéome du parasite dans son stade érythrocytaire grâce à la stratégie N-TOP. Nous avons identifié près d'un millier de protéines et plus de 500 extrémités N-terminales. Cette étude sera étendue aux autres compartiments de l'érythrocyte infecté. De plus, l'étude bioinformatique des identifications en cours, permettra de dégager un certain nombre de cibles thérapeutiques potentielles qui pourront ensuite être suivies par SRM. Enfin, dans une troisième étude, nous avons identifié une phosphothréonine indispensable pour la translocation de la protéine transporteur membranaire PfCRT impliquée dans la résistance du parasite aux antibiotiques.

**Une étude protéomique différentielle du tissu adipeux du rat soumis au jeûne prolongé** a été réalisée en combinant une approche basée sur les gels 2D DiGE et l'identification par spectrométrie de masse. Nous avons ainsi pu identifier les protéines présentes dans 28 spots différentielles dans l'étude du tissu épидидymaire. Grâce à l'interface web « msda.unistra.fr », développé au laboratoire, qui permet d'automatiser la recherche dans la base de données « Gene Ontology », nous avons classé les protéines différentielles en familles selon les processus biologiques dans lesquels elles sont impliquées. Ceci nous a permis de mettre en évidence les principaux mécanismes d'adaptation à la privation de nourriture.

**Une évaluation de la stratégie qN-TOP pour l'étude différentielle des clivages protéolytiques a été réalisée sur le foie de rat soumis au jeûne prolongé.** Le protéome hépatique est un bon exemple de protéome particulièrement complexe pour une étude de quantification relative en dégradomique. Nous n'avons pas pu quantifier de façon satisfaisante les changements protéolytiques entre les différents échantillons (foie de rat

nourri et en phase 2 et 3 du jeûne). Cependant cette étude nous a permis d'identifier les principales améliorations à apporter à la stratégie analytique. Il ressort de cette étude que la meilleure stratégie consisterait à réaliser une analyse en deux temps : une première analyse d'identification par LC-MS/MS, suivie par une analyse LC-MS du même digestât. Les données de quantification seront alors générées à partir de l'analyse LC-MS.

Enfin, dans le cadre d'une étude de la **structure d'une protéine du nerf du calamar impliquée dans l'échange sodium/calcium**, nous avons développé une approche de spectrométrie de masse supramoléculaire combinée à une stratégie LC-MS de haute résolution. Ceci nous a permis de caractériser les interactions entre cette protéine et les acides gras contenus dans les vésicules lipidiques du calamar et d'identifier l'acide palmitique comme étant l'acide gras fonctionnel permettant d'activer cette protéine.

**Notre travail a illustré l'intérêt de la méthode de marquage N-terminal pour l'étude des clivages protéolytiques et a montré qu'elle peut être utilisée maintenant en routine, grâce aux améliorations apportées.**

Plusieurs perspectives d'amélioration peuvent encore être proposées. Elles portent sur l'augmentation de la sensibilité, grâce à d'autres réactifs de dérivation avec une hydrophobie modulée par exemple. Elles concernent également un meilleur traitement bioinformatique des données ; notre laboratoire travaille sur de nouveaux algorithmes pour la validation automatique des spectres de peptides marqués au TMPP. Enfin, une stratégie en deux temps (analyse LC-MS/MS puis LC-MS) permettra de réaliser des quantifications sur des échantillons très complexes.

Ces perspectives illustrent que l'analyse protéomique est loin d'avoir atteint son potentiel maximal et que de nombreux autres développements permettront dans les années qui viennent des progrès en sensibilité, en dynamique et en débit similaires à ceux observés dans les analyses génomiques.



## Partie expérimentale

---



## Partie II

Dérivation des protéines en solution par le TMPP-Ac-OSu, le 13C9 – TMPP-Ac-OSU et le TMPP-Pentanoyl-OSu (N-TOP)

Les protéines sont dissoutes dans un tampon tris 50 mM, urée 8 M, thiourée 2 M, SDS 1 % pH=8,20. Après dosage de Bradford (si la concentration est inconnue), réduction à l'aide de 5 mM de tributylphosphine et alkylation à l'aide de 30 mM d'iodoacétamide, 200 équivalents molaires de réactif TMPP sont ajoutés. Après 1 heure de réaction avec agitation mécanique, 0,1 M d'hydroxylamine sont ajoutées, puis l'agitation du mélange est maintenue pendant 1 heure. Les protéines sont ensuite séparées sur un gel 1D. Le pourcentage de polyacrylamide des gels est choisi en fonction des protéines étudiées (16 % pour l'hémoglobine, 10 % pour les échantillons complexes). Pour la stratégie qN-TOP les échantillons dérivés au TMPP léger et lourd sont mélangés en quantité égale avant de les séparer sur le gel 1D. Les bandes sont ensuite découpées, lavées deux fois avec du bicarbonate d'ammonium 25 mM et deux fois avec de l'acétonitrile et puis les ponts disulfures sont réduits à l'aide de DTT (10 mM) et les thiols des cystéines sont alkylés avec 30 mM d'iodoacétamide. Les morceaux de gel sont ensuite déshydratés avec 50  $\mu$ L d'acétonitrile.

### 1. Chapitre 1

#### 1.1. Dérivation des protéines au TMPP-Ac-OSu dans le gel (Ingel N-TOP)

Après séparation des protéines sur un gel 1D, les bandes d'intérêt sont excisées, lavées deux fois avec du bicarbonate d'ammonium 25 mM et deux fois avec de l'acétonitrile et puis les ponts disulfures sont réduits à l'aide de TCEP (5 mM) et les thiols des cystéines sont alkylés avec 30 mM d'iodoacétamide. Les morceaux de gel sont ensuite déshydratés avec 50  $\mu$ L d'acétonitrile. La réaction de dérivation est effectuée en ajoutant à chaque bande de gel déshydratée 100  $\mu$ g de TMPP-Ac-OSu contenu dans un volume maximal de 5  $\mu$ L d'acétonitrile 20 %. Lorsque ce volume est absorbé par les morceaux de gel déshydratés, 50  $\mu$ L de tampon tris 50 mM, urée 8 M, thiourée 2 M, SDS 1 % pH=8,20 sont ajoutées. Après un temps de réaction d'une heure, 0,1 M d'hydroxylamine sont ajoutés pour arrêter la réaction. Les bandes de gel subissent ensuite quatre cycles de lavage/déshydratation automatisés sur

le robot Massprep (Waters). Chaque cycle comprend quatre lavages avec de l'hydrogencarbonate d'ammonium 25 mM et de l'acétonitrile et une étape de déshydratation des morceaux de gel à l'aide de l'acétonitrile.

## 1.2. Digestion enzymatique et analyse LC-MS/MS

Les protéines sont digérées dans le gel à l'aide de la trypsine (0,25 µg / bande de gel). Les peptides de digestion sont ensuite extraits à l'aide de 50 µL d'acétonitrile 60 % dans 40 % d'eau acidifiée avec 1 % HCOOH. Après évaporation de l'acétonitrile au speedvacuum, les peptides sont analysés par LC-MS/MS sur un système nanoLC ChipCube (Agilent Technologies) couplée à une trappe d'ions (Amazon, Bruker). La chip contenait une colonne analytique Zorbax 300SB-C18 (15 cm x 75 µm avec des particules de 5 µm) et une précolonne de 40 nL contenant la même phase. La phase mobile était composée d'un solvant A (2% acétonitrile dans 0,1% HCOOH) et d'un solvant B (98% acétonitrile, 2% eau, 0,1% HCOOH). Les peptides sont injectés sur la colonne à l'aide du solvant A à 3,75 µL minutes et sont ensuite élués sur la colonne analytique à l'aide d'un gradient 8 – 60 % de B dans A en 45 mins (pour les digests de protéines modèles).

La trappe d'ions est utilisée en mode positif avec alternation d'un balayage MS et six balayages MS/MS. Les balayages MS sont effectués dans le mode « standard enhanced » à 8.100 m/zs sur la gamme m/z 250-2000. Le temps d'accumulation maximal était de 200 ms. Pour la MS/MS, les six ions les plus intenses sont sélectionnés avec une préférence pour les deux et trois fois chargés. La fenêtre d'isolation était de 4 m/z. Les ions sélectionnés sont exclus pendant 1 cycle et 0,2 min. Le système était piloté à l'aide du logiciel Hystar.

## 2. Chapitre 2

### 2.1. Evaluation de la linéarité de réponse du MaXis à l'aide des protéines modèles marquées au TMPP lourd et léger

Après dérivation des amines α de l'hémoglobine avec du TMPP léger et du TMPP lourd comme décrit ci-dessus, les protéines marquées léger et lourd sont déposées individuellement et en quantités égales sur un gel 1D (16% acrylamide/bisacrylamide). Les bandes de gel sont excisées et les protéines sont digérées dans le gel à l'aide de la trypsine (0,25 µg / bande de gel). Les peptides de digestion sont ensuite extraits à l'aide de 50 µL d'acétonitrile 60% dans 40% d'eau acidifiée avec 1% HCOOH. Les ratios léger/lourd sont ensuite préparés à partir de ces extraits et analysés sur le système nanoLC (NanoAcquity Waters) Q-TOF (Maxis, Bruker).

Les mélanges de peptides sont injectés à 15 µL/min sur une précolonne 20 x 0.18 mm, 5 µm Symmetry C18 (Waters) et sont ensuite élués à 0,45 µL/min sur une colonne BEH130 C18 column (Waters), 75 µm x 200 mm, 1.7 µm. La phase mobile était composée d'un solvant A (0,1% HCOOH dans l'eau) et d'un solvant B (0,1% HCOOH dans l'acétonitrile).

L'élution est réalisée à l'aide d'un gradient 6-45 % de solvant B dans A en 18 minutes suivi par un plateau d'une minute à 90 % B avant rééquilibrage de la colonne à 99 % B.

Le Maxis équipé d'un nanosprayer était utilisé en mode positif. Le TOF était étalonné de m/z 50 à 2200 à l'aide de Tuning Mix LC (Agilent Technologies). La correction de la dérive de cet étalonnage est effectuée par la diffusion en continu de 2 ions du Tuning Mix (m/z 299,2945 et 922,0098). Pour la MS/MS, le système alternait un balayage MS de m/z 50 à 2200 à 0,2 s et quatre balayages MS/MS sur la même gamme de m/z. La durée des balayages MS/MS est réglée automatiquement en fonction de l'intensité des ions sélectionnés lors des balayages MS. Les quatre ions les plus intenses sont sélectionnés et fragmentés en mode CID. Les ions sélectionnés sont ensuite exclus pendant 0,6 secondes. Le système était piloté à l'aide du logiciel Hystar (Bruker). L'interprétation des données et l'intégration des pics sont effectuées à l'aide du logiciel DataAnalysis (Bruker).

### 3. Chapitre 3

La dérivation des protéines modèles (hémoglobine  $\alpha$  et  $\beta$ ) à l'aide des formes plus hydrophobes du TMPP est effectuée comme décrit dans le protocole ci-dessus. La digestion enzymatique et l'analyse LC-MS/MS sont effectuées comme décrit dans le protocole du chapitre 1 avec une différence : le gradient d'élution pour la chromatographie était de 8 à 60 % en 120 minutes.



## Partie III

### 1. Chapitre 1

Un article est en cours de publication pour ces travaux. Le paragraphe matériel et méthode de cet article est le suivant :

#### ***N-termini labeling and sample preparation***

*The mitochondria suspension is first centrifuged for 30 minutes at 10 000 x g, the supernatant is discarded. The mitochondria are then lysed by dissolution in the labeling buffer (50 mM Tris-HCl, 8 M urea, 2 M thiourea, 1% SDS, pH 8.20). A Bradford assay is performed to estimate the amount of proteins. The disulfide bonds are then reduced using 5 mM tributylphosphine over 1 hour and alkylated using 50 mM iodoacetamide over 1 hour before adding 5 mM of tributylphosphine to eliminate the excess of iodoacetamide. The labeling reaction is then performed by adding 200 molar excess of (N-Succinimidyl)oxycarbonylmethyl)tris(2,4,6-trimethoxyphenyl)phosphonium bromide (TMPP) (Sigma-Aldrich) and incubating for 1 hour. The reaction is then quenched using 100 mM of hydroxylamine over 1 hour. The mitochondrial proteins are then loaded on a 10% SDS polyacrylamide gel electrophoresis (100 µg of protein per gel line). The gel is then stained with colloidal blue and finally each gel line is sliced into 33 slices of 2 mm each using an adapted gel slicing tool and transferred into a 96 well plate. The gel slices were then prepared for enzymatic digestion using a MassPrep automated system (Waters, Manchester, UK). The gel slices were washed twice in 25 mM ammonium hydrogen carbonate and acetonitrile before being dehydrated. Two independent enzymatic digestions were performed in-gel using modified porcine trypsin (Promega, France) and AspN (Roche, Germany) at 37°C overnight. The peptides are then extracted twice using 40:60 water: acetonitrile + 1 % formic acid solution for the first extraction and 100 % acetonitrile for the second one. The volumes of the extracts are then reduced by speed vacuuming. The AspN and tryptic peptides are then analyzed by LC-MS/MS using two different mass analyzers.*

#### ***NanoLC-MS/MS analysis with a Q-TOF type instrument***

*Peptides were analyzed using an UHPLC system (NanoAcquity, Waters, Milford MA, USA) coupled to a high resolution nanoElectrospray-Quadrupole-Time of Flight type mass*

spectrometer (Maxis, Bruker Daltonics, Bremen, Germany). The peptides were trapped on a 20 x 0.18 mm, 5  $\mu\text{m}$  Symmetry C18 precolumn (Waters), and then separated on a ACQUITY UPLC<sup>®</sup> BEH130 C18 column (Waters), 75  $\mu\text{m}$  x 200 mm, 1.7  $\mu\text{m}$  particle size. 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 1 min at 15  $\mu\text{L}/\text{min}$  with 99 % of solvent A and 1% of solvent B. Elution was performed at a flow rate of 450 nL/min, using 1-50 % gradient (solvent B) over 38 min at 45°C followed by 90% (solvent B) over 1 min before the reconditioning of the column at 99 % of solvent A over 6 min.

The Maxis equipped with a nanosprayer was operated in the positive ion mode. Mass calibration of the TOF was achieved prior to analysis from  $m/z$  50 to 2200 using Tunemix (Agilent technologies, Waldbronn, Germany). Online correction of this calibration was performed using two Tunemix ions ( $m/z$  299.2945 and 922.0098) as recalibration masses. The capillary voltage was set to 4500 V and the end plate offset to -500 V. For tandem MS experiment, the system was operated with automatic switching between MS and MS/MS modes. The MS scan is performed over a  $m/z$  range of [50;2200] at 0.2 seconds/scan. The MS/MS time is regulated by the intensity of the selected peak on the MS spectrum; it ranges from 0.1 to 1.4 seconds/scan over the same  $m/z$  range. The five most abundant peptides, preferably doubly and triply charged ions, were selected on each MS spectrum for further isolation and Collision Induced Dissociation (CID) using optimized collision energy depending on the charge state and the  $m/z$  of the ion. The selected peptides are then excluded for 0.8 minutes. The complete system was fully controlled by Compass Hystar (Bruker Daltonics).

#### **NanoLC-MS/MS analysis with an ion trap type instrument**

Peptides were analyzed using an Agilent 1200 series HPLC-Chip/MS system (Agilent Technologies, Palo Alto, CA, USA) coupled to an Amazon ion trap mass spectrometer (Bruker Daltonics, Bremen, Germany). The Chip contained a Zorbax 300SB-C18 (150 mm x 75  $\mu\text{m}$ , with a 5  $\mu\text{m}$  particle size) column and a Zorbax 300SB-C18 (40nL, 5 $\mu\text{m}$ ) enrichment column (Agilent Technologies). The solvent system consisted of 2 % Methanol and 0.1 % formic acid in water (solvent A); 2 % water and 0.1 % formic acid in acetonitrile (solvent B). The peptides were loaded onto the enrichment column at a 3.75  $\mu\text{L}/\text{min}$  flow rate using solvent A. Elution was performed at a flow rate of 0.3  $\mu\text{L}/\text{min}$  with a 10-30 % solvent B linear gradient over 20 min followed by 30-70 % solvent B linear gradient over 80 min followed by an 85 % stage (solvent B) over 6 min before reconditioning the column over 10 min.

The Amazon ion trap was operated in the positive ion mode and calibrated prior to analysis using Tunemix (Agilent Technologies). Two independent experiments were performed; the first using Collision-Induced Dissociation (CID) and the second using Electron Transfer Dissociation (ETD). The MS scanning was performed in the standard enhanced resolution mode at a scan rate of 8100  $m/z$  per second. The mass range was 250 – 1500  $m/z$ . The ion charge control was 200 000 and the maximum accumulation time was 200 ms. A total of two scans were averaged to obtain an MS spectrum. For tandem MS experiment the

system was operated with automatic switching between MS and MS/MS modes. The six most abundant peptide ions, preferring doubly charged ones (absolute threshold of 5000 and a relative of 5 %), were selected on each MS spectrum for further isolation (4 m/z isolation width) and fragmentation. Ions were excluded after 1 spectrum and released after 0.5 min. the MS/MS scanning was performed in the ultrascan resolution mode at a scan rate of 32 000 m/z per second. The ion charge control was set at 300 000. A total of 5 scans were averaged to obtain a MS/MS spectrum. The complete system was fully controlled by Compass Hystar (Bruker Daltonics).

### **Database search and interpretation**

Mass Data collected during analysis were processed, converted into mgf files using DataAnalysis software (Bruker Daltonics). The mgf files were submitted to two different search engines, Mascot (Matrix Sciences, Boston, MA, USA) in its 2.3 version on a local server; and OMSSA (NCBI, USA) on the EGEE grid (<http://msda.u-strasbg.fr>). Searches were performed against a composite target-decoy database generated using internally developed tools (<http://msda.u-strasbg.fr>) and containing the Swissprot protein sequences of Homo sapiens and common contaminants (Trypsin) downloaded on the 28<sup>th</sup> of January 2010. Two different searches were performed using both search engines: a first search for global protein identification by specific enzyme cleavage and a second search by semispecific enzymatic cleavage for N-termini identification for transit peptide characterization.

For the Q-TOF Maxis data, searches were performed with a 5 ppm Da error tolerance on mass measurements for precursor and 0.02 Da error tolerance for fragment ions. For the Amazon ion trap data searches were performed with a 0.2 Da error tolerance on mass measurements for both precursor and fragment ions.

In specific enzyme searches, oxidation of methionine residues, acetylation of protein N-terminal residues and carbamidomethylation of cysteine residues were searched as variable modifications. Up to one trypsin and 2 AspN miss cleavage sites were allowed. In semispecific enzymes searches, TMPP-Ac peptide N-terminal residues, lysine, tyrosine and cysteine residues were searched as variable modifications along with the above cited modifications. Up to 2 trypsin and 3 AspN miss cleavage sites were allowed.

The searches' results were then processed and interpreted using Scaffold 3 (Proteome Software, Portland, OR, USA). In the global identification searches, a target-decoy strategy is applied using search engine's (OMSSA and Mascot) score thresholds in order to have a protein false discovery rate inferior to 1 %. For the N-termini characterization searches, all spectra yielding a TMPP labeled peptide identification are manually inspected and interpreted in order to validate the identification.

## 2. Chapitre 2

### 2.1. Etude 1

La dérivation au TMPP dans le gel et la digestion trypsique sont effectuées comme décrit dans le protocole du chapitre 2 de la partie II. L'analyse nanoLC-MS/MS est effectuée comme décrit dans le chapitre 1 de la partie III paragraphe 2. Les détails de l'interprétation des données et l'identification des protéines et des peptides sont décrites dans le paragraphe stratégie analytique de cette étude.

### 2.2. Etude 2

Les détails des expériences et des protocoles expérimentaux sont décrits dans le paragraphe stratégie analytique de l'étude.

### 2.3. Etude 3

Les détails des expériences et des protocoles expérimentaux sont décrits dans la publication des résultats.

## 3. Chapitre 3

Les détails des expériences et des protocoles expérimentaux sont décrits dans la publication des résultats.

## 4. Chapitre 4

Pour les analyses nanoLC-MS/MS, les conditions expérimentales sont les suivantes :

- Les peptides générés par la digestion trypsique sont analysés sur un système nanoLC Chip-Cube (Agilent) couplé à un spectromètre de masse se type trappe d'ions (Amazon, Bruker). La pente du gradient utilisée est une pente douce de 10 à 60 % de méthanol dans l'eau en 120 minutes. Des gradients lents avec des pentes douces sont des gradients de choix pour la stratégie N-TOP en général, surtout lorsqu'il s'agit de l'analyse d'échantillon complexe. Ces gradients permettent une meilleure séparation des peptides marqués au TMPP et par la suite, favorisent d'avantage leur sélection par le spectromètre de masse pour la MS/MS.

Le méthanol en tant que modificateur de phase a été utilisé au laboratoire suite à la pénurie mondiale d'acétonitrile en 2009. Des tests ont été effectués pour évaluer le méthanol et le comparer à l'acétonitrile pour l'analyse nanoLC-MS/MS. Ces travaux sont présentés dans la thèse d'Agnès Hovasse. Les résultats de ces tests montrent que l'utilisation du méthanol ne pénalise pas les performances analytiques du système après adaptations

des gradients chromatographiques. De plus, l'efficacité de l'ionisation électrospray s'est trouvée améliorée.

Les paramètres de la trappe d'ions sont identiques à ceux décrits dans le chapitre 1 de la partie III.

## **5. Chapitre 5**

Les détails des expériences et des protocoles expérimentaux sont décrits dans la publication des résultats.



# Travaux supplémentaires

---



## Travaux supplémentaires

Dans le cadre de différentes collaborations au cours de ma thèse avec l'Institut de Biologie Moléculaire des Plantes de Strasbourg, les outils de l'analyse protéomique nous ont permis d'apporter des éléments de réponses à plusieurs problématiques de biologie végétale. Nous avons ainsi pu caractériser un site de phosphorylation de la protéine GIP d'*Arabidopsis thaliana* et de celle de son homologue humaine. Nous avons également identifié certaines protéines impliquées dans la biosynthèse des glycanes liés à l'isoprényle.

Dans cette partie « Travaux supplémentaires », sont jointes une publication dans *The Plant Cell* parue en 2011, le contexte de l'étude de la protéine GIP, ainsi que la partie matériel et méthodes d'un article en cours de rédaction et les spectres qui ont permis la caractérisation des sites de phosphorylation de cette protéine GIP.

Au congrès SMAP 2009 à Dijon, nous avons présenté dans un poster des travaux effectués en collaboration avec le laboratoire de Chimie Analytique et Sciences Séparatives de l'IPHC de Strasbourg. Ce poster est aussi joint dans cette partie. Ces travaux portent sur la caractérisation de métalloprotéines par une approche HR-CN-PAGE (« High Resolution Clear Native PolyAcrylamide Gel Electrophoresis ») et par spectrométrie de masse supramoléculaire. Le poster présenté a été récompensé par le prix du meilleur poster du congrès SMAP 2009.



## **Publication**

**DOLICHOL PHOSPHATE MANNOSE SYNTHASE1 Mediates the Biogenesis of Isoprenyl-Linked Glycans and Influences Development, Stress Response, and Ammonium Hypersensitivity in Arabidopsis**

Nurul Jadid, Alexis Samba Mialoundama, Dimitri Heintz, Daniel Ayoub, Mathieu Erhardt, Jérôme Mutterer, Denise Meyer, Abdelmalek Alioua, Alain Van Dorsselaer, Alain Rahier, Bilal Camara and Florence Bouvier



# DOLICHOL PHOSPHATE MANNOSE SYNTHASE1 Mediates the Biogenesis of Isoprenyl-Linked Glycans and Influences Development, Stress Response, and Ammonium Hypersensitivity in *Arabidopsis* <sup>W</sup>

Nurul Jadid,<sup>a,b,1</sup> Alexis Samba Mialoundama,<sup>a,1</sup> Dimitri Heintz,<sup>a</sup> Daniel Ayoub,<sup>c</sup> Mathieu Erhardt,<sup>a</sup> Jérôme Mutterer,<sup>a</sup> Denise Meyer,<sup>a</sup> Abdelmalek Alioua,<sup>a</sup> Alain Van Dorsselaer,<sup>c</sup> Alain Rahier,<sup>a</sup> Bilal Camara,<sup>a</sup> and Florence Bouvier<sup>a,2</sup>

<sup>a</sup> Institut de Biologie Moléculaire des Plantes du Centre National de la Recherche Scientifique, Université de Strasbourg, 67084 Strasbourg Cedex, France

<sup>b</sup> Department of Biology, Botanical and Plant Tissue Culture Laboratory, Sepuluh Nopember Institut of Technology (Its), Gedung H Kampus Its Sukolilo, Surabaya 60111, East-Java, Indonesia

<sup>c</sup> Laboratoire de Spectrométrie de Masse Bio-Organique, Département des Sciences Analytiques, Institut Pluridisciplinaire Hubert Curien du Centre National de la Recherche Scientifique, Unité Mixte de Recherche 7178, Université de Strasbourg, 67087 Strasbourg Cedex, France

The most abundant posttranslational modification in nature is the attachment of preassembled high-mannose-type glycans, which determines the fate and localization of the modified protein and modulates the biological functions of glycosylphosphatidylinositol-anchored and *N*-glycosylated proteins. In eukaryotes, all mannose residues attached to glycoproteins from the luminal side of the endoplasmic reticulum (ER) derive from the polyprenyl monosaccharide carrier, dolichol P-mannose (Dol-P-Man), which is flipped across the ER membrane to the lumen. We show that in plants, Dol-P-Man is synthesized when Dol-P-Man synthase1 (DPMS1), the catalytic core, interacts with two binding proteins, DPMS2 and DPMS3, that may serve as membrane anchors for DPMS1 or provide catalytic assistance. This configuration is reminiscent of that observed in mammals but is distinct from the single DPMS protein catalyzing Dol-P-Man biosynthesis in bakers' yeast and protozoan parasites. Overexpression of DPMS1 in *Arabidopsis thaliana* results in disorganized stem morphology and vascular bundle arrangements, wrinkled seed coat, and constitutive ER stress response. Loss-of-function mutations and RNA interference-mediated reduction of DPMS1 expression in *Arabidopsis* also caused a wrinkled seed coat phenotype and most remarkably enhanced hypersensitivity to ammonium that was manifested by extensive chlorosis and a strong reduction of root growth. Collectively, these data reveal a previously unsuspected role of the prenyl-linked carrier pathway for plant development and physiology that may help integrate several aspects of candidate susceptibility genes to ammonium stress.

## INTRODUCTION

The biological functions of more than 50% of proteins are determined by the covalent attachment of carbohydrates referred to as glycans, which constitute the most diverse and abundant posttranslational modifications found in nature (Apweiler et al., 1999). Data from human studies reveal that most virus and pathogenic microbes bind to cells via glycoproteins (Hart and Copeland, 2010). Furthermore, several clinical markers and therapeutic targets of cancer cells are glycoproteins (Zhao et al., 2006; Hart and Copeland, 2010). Four different types of protein glycosylation have been characterized. These include

*N*-glycosylation, C- and O-mannosylation, and the glycosyl phosphatidylinositol (GPI) anchor attachment (Figure 1). Although *N*-glycosylations occur in different taxa, including Eukarya, Bacteria, and Archaea (Weerapana and Imperiali, 2006; Jones et al., 2009), plants apparently do not perform protein C-mannosylation and protein O-mannosylation (Furmanek and Hofsteenge, 2000; Lommel and Strahl, 2009).

In eukaryotic organisms, *N*-glycosylation is achieved by co- or posttranslational attachment of the preassembled core glycan, Glc<sub>3</sub>Man<sub>9</sub>GlcNAc<sub>2</sub>-Asn (Glc, Man, and GlcNAc refer to Glc, Man, and *N*-acetylglucosamine) to the  $\gamma$ -amido group of a specific Asn in the glycosylation consensus sequence Asn-X-Ser/Thr. The reaction is catalyzed by an oligosaccharyltransferase multisubunit complex in the endoplasmic reticulum (ER) membranes (Yan et al., 2005). This modification favors the folding of *N*-glycoproteins due to increased solubility (Molinari, 2007; Hanson et al., 2009) and represents a key signal used by the ER-associated protein degradation machinery to determine the fate of misfolded proteins (Helenius and Aebi, 2001). Indeed, just after the transfer, the glycan is subject to trimming processes by glucosidases I and II to allow

<sup>1</sup> These authors contributed equally to this work.

<sup>2</sup> Address correspondence to florence.bouvier@ibmp-cnrs.unistra.fr.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) is: Florence Bouvier (florence.bouvier@ibmp-cnrs.unistra.fr).

<sup>W</sup> Online version contains Web-only data.

www.plantcell.org/cgi/doi/10.1105/tpc.111.083634



addition to their biological functions are key constituents of plant gums (Pettolino et al., 2006; Seifert and Roberts, 2007).

The presence of several Man residues represents a conserved feature of the four types of protein-linked glycans (Figure 1). In the Golgi complex, only guanosine 5'-diphosphate-Man (GDP-Man) is used as the Man donor, whereas in the ER, GDP-Man and prenyl-linked mannoses are used (Helenius and Aebi, 2001). The prenyl monosaccharide dolichol P-Man (Dol-P-Man) acts as the main carrier of transferred Man residues. The synthesis of Dol-P-Man proceeds via two main steps: formation of dolichyl phosphate (Dol-P) and addition of Man from GDP-Man. In yeast and mammalian cells, Dol-P is synthesized exclusively from the mevalonic pathway between the cytosol and the ER compartments (Schenk et al., 2001). In plants, the initial steps of Dol-P synthesis operate in the plastids from deoxy xylulose phosphate-derived isopentenyl diphosphate (IPP) (40 to 50% of total IPP incorporated) and IPP derived from the cytosolic mevalonic pathway (Skorupinska-Tudek et al., 2008). Subsequently, dolichol P-Man synthase (DPMS) located on the cytoplasmic face of the ER catalyzes the transfer of Man from GDP-Man to Dol-P, and the resulting Dol-P-Man is translocated to the luminal side of the ER membrane by an unknown flippase (Sanyal and Menon, 2010). Based on available data from mammalian cells, the translocated Dol-P-Man donates all of the Man residues used on the luminal side of the ER for the synthesis of the GPI anchor and the four mannoses of *N*-glycans (the 6th to the 9th Man residues) (Burda et al., 1999; Frank and Aebi, 2005) (Figure 1).

Accumulating evidence supports the existence of two classes of DPMS. In *Saccharomyces cerevisiae* (Orlean et al., 1988), *Trypanosoma brucei* (Mazhari-Tabrizi et al., 1996), *Leishmania mexicana* (Ilgoutz et al., 1999), and *Thermoplasma acidophilum* (Zhu and Laine, 1996), DPMS functions as a homomeric enzyme termed DPMS1. By contrast, in mammals (Maeda et al., 2000; Maeda and Kinoshita, 2008), *Schizosaccharomyces pombe* (Colussi et al., 1997), and *Trichoderma reesei* (Kruszewska et al., 2000), the synthesis of Dol-P-Man requires the assembly of a heteromeric complex, including a catalytic DPMS1 and noncatalytic DPMS2 and DPMS3 proteins. The identification, organization, and biological role of DPMS in plants are not yet known. In this study, we analyzed how Dol-P-Man is synthesized in plants and evaluated its role using overexpressing and knock-out lines. We provide evidence that in plants, Dol-P-Man is synthesized as in mammals by three protein components referred to as DPMS1, DPMS2, and DPMS3. DPMS1 is the catalytic module and is tethered or assisted by DPMS2 and DPMS3. Our study based on the analysis of the loss of function and the overexpression of DPMS1 demonstrates the central role played by the Dol-P-Man pathway during *Arabidopsis* growth and development and highlights the unexpected hypersensitivity of *dpms1* mutant plants to ammonium.

## RESULTS

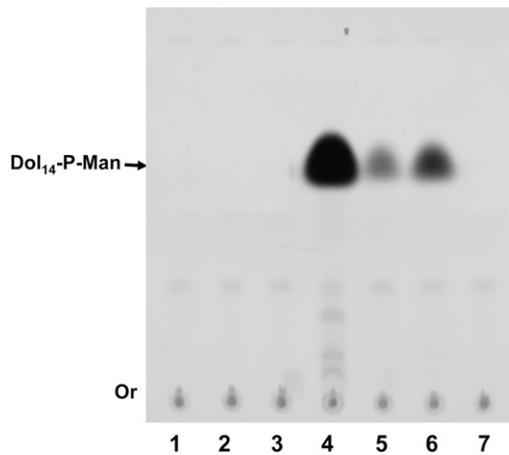
### Characterization and Functional Reconstitution of DPMS

Because Dol-P-Man synthesis is catalyzed by homomeric or heteromeric (Maeda and Kinoshita, 2008) enzymes, we made

use of comparative genomic analysis to identify putative plant homologs of the minimum core enzyme DPMS1 and its potential protein partners, DPMS2 and DPMS3. We performed a BLAST search against the *Arabidopsis* database using yeast DPMS1 and human DPMS1, DPMS2, and DPMS3 as query sequences. We noted a significant match (BLASTP, E value =  $9 \times 10^{-47}$ ) to the peptide sequence encoded by At1g20575 that was subsequently considered to be the putative *Arabidopsis* DPMS1 (At-DPMS1). At-DPMS1 displayed identity 61% to human, *S. pombe* (61%), *S. cerevisiae* (28%), and rice (*Oryza sativa*; 85%) homologs (see Supplemental Figure 1 online). Using the same procedure, we retrieved putative *Arabidopsis* DPMS2 (At1g74340) and DPMS3 (At1g48140) proteins that have E values of  $4 \times 10^{-14}$  and 0.001 and sequence identity 57 and 31% to human DPMS2 and DPMS3, respectively (see Supplemental Figures 2 and 3 online). Using the prediction program HMMTOP (<http://www.enzim.hu/hmmtop/html/submit.html>), we noted the presence of transmembrane domains in DPMS2 (two transmembrane helices, 7-30 and 49-73) and DPMS3 (two transmembrane helices, 7-27 and 36-55). Previous studies based on the relatedness to the bacterial spore coat forming protein SpSA have shown that human DPMS1 and *S. cerevisiae* DPMS1 have similar tertiary structure, in spite of their divergence (Maeda and Kinoshita, 2008). These studies indicate that in human and *S. cerevisiae*, Tyr-12, Asp-44, Asp-97, and Arg-212 are key active residues (Lamani et al., 2006; Maeda and Kinoshita, 2008). These residues are conserved in *Arabidopsis* DPMS1 (see Supplemental Figure 1 online).

To analyze precisely the requirements for *Arabidopsis* Dol-P-Man synthesis, we used purified recombinant DPMS1, DPMS2, and DPMS3 proteins (see Supplemental Figure 4 online). Recombinant DPMS1 alone did not catalyze Dol-P-Man synthesis (Figure 2, lane 1). Similarly, recombinant DPMS2 and DPMS3 did not display DPMS activity either individually or when both proteins were combined (Figure 2, lanes 2, 3, and 7). Dol-P-Man was synthesized when DPMS1 + DPMS2 + DPMS3 were associated (Figure 2, lane 4), and the reaction was linear with respect to time for 45 min. When DPMS1 and DPMS3 alone were coincubated (i.e., in the absence of DPMS2), up to 20% of the activity could be recovered (Figure 2, lane 6), whereas only 5% of the activity was obtained in the presence DPMS1 + DPMS2 alone (Figure 2, lane 5).

Because *Arabidopsis* accumulates dolichols from Dol<sub>14</sub> to Dol<sub>23</sub> (Jozwiak et al., 2009), and due to the paucity of data characterizing Dol-P-Man, scale-up reactions were performed using a reaction mixture containing nonlabeled Dol<sub>14</sub> to analyze further the reaction product of *Arabidopsis* DPMS. The DEAE-cellulose fraction eluted with dichloromethane/methanol (7v/3v) containing ammonia (0.28 N) and 50 mM ammonium acetate was subjected to liquid chromatography–electrospray ionization–mass spectrometry (LC-ESI-MS) analysis to characterize the putative Dol<sub>14</sub>-P-Man product with an expected molecular mass of 1215. LC-ESI-MS analysis revealed under basic conditions (pH 8) the presence of an ion with a mass-to-charge ratio (*m/z*) of 1196 that had the highest intensity, corresponding to the dehydrated form [M-H<sub>2</sub>O- H] (Figure 3A). A [M-H]- form with *m/z* 1214 was also identified, but with a lower intensity, and was nevertheless the best candidate for the identification of the Man residue and the phosphate group that gave a loss of *m/z* 241 corresponding to [Man-PO<sub>3</sub>] (see Supplemental Figure 5 online). These data agree with the fact



**Figure 2.** Reconstitution of DPMS Activity and TLC Analysis of Dol<sub>1,4</sub>-P-Man Produced in Vitro.

Assays were performed using purified recombinant DPMS1, DPMS2, and DPMS3 either alone (lanes 1 to 3) or in different combinations: DPMS1 + DPMS2 + DPMS3 (lane 4), DPMS1 + DPMS2 (lane 5), DPMS1 + DPMS3 (lane 6), and DPMS2 + DPMS3 (lane 7). The TLC plate was developed using the solvent mixture (dichloromethane/methanol/water; 10:10:3, v/v/v), and reaction products were analyzed by autoradiography. The origin (Or) and the position of Dol<sub>1,4</sub>-P-Man (arrow) are indicated.

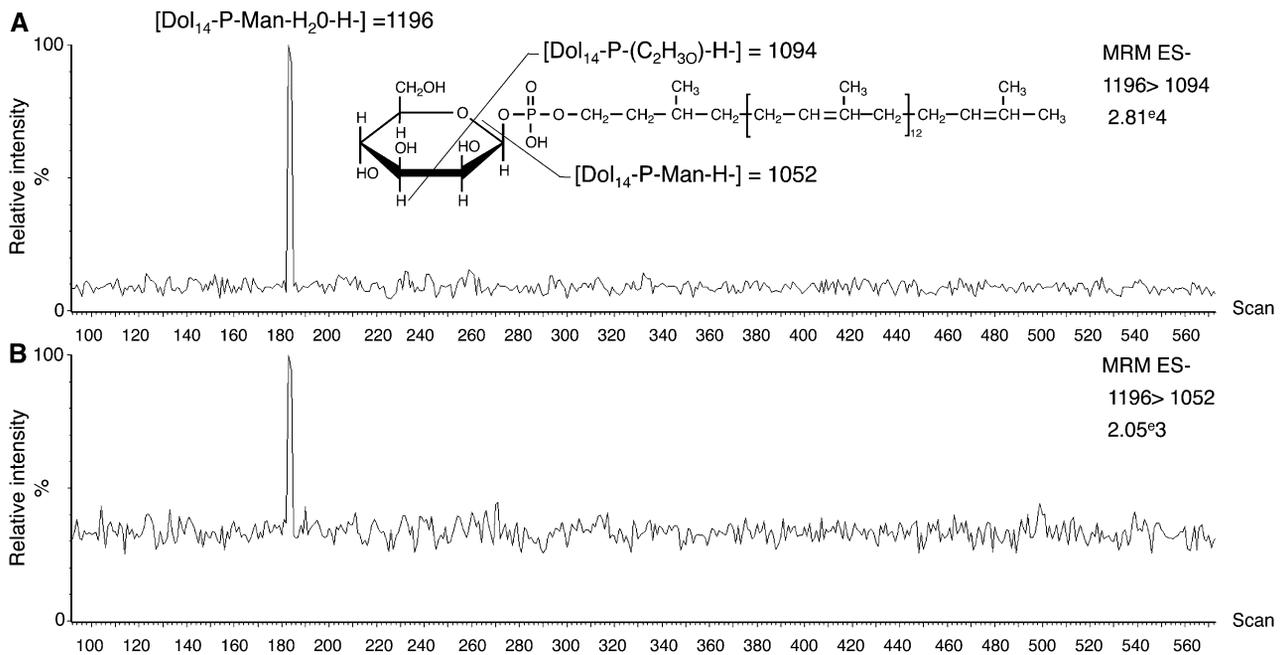
that dolichol-hexose-phosphate molecules are better identified in their dehydrated form and by the characteristic ion at  $m/z$  241 (Rush et al., 1998; Sheeley and Reinhold, 1998; Garrett et al., 2007). It has also been shown that the fragmentation of Dol-P-hexose involves a partial or total loss of the hexose and the phosphate moieties (Rush et al., 1998; Sheeley and Reinhold, 1998; Garrett et al., 2007). Indeed, in our study, the fragmentation of the Dol<sub>1,4</sub>-P-Man generated two intense daughter ions of  $m/z$  1094 and 1052, corresponding to the loss of Man and [Dol-HPO<sub>4</sub>]<sup>-</sup>, respectively, and to partial loss of Man (120 D) [Dol-PO<sub>4</sub>-(C<sub>2</sub>H<sub>3</sub>O)]<sup>-</sup>, which were subsequently used for multiple reaction monitoring (MRM) analysis (Figures 3A and 3B).

All considered, these data demonstrate that DPMS1 is the catalytic module of DPMS, but the enzyme activity requires the coincubation of DPMS1 with DPMS2 and DPMS3. Thus, DPMS2 and DPMS3 are probably orthologs of mammal DPMS2 and DPMS3 (Maeda and Kinoshita, 2008). These data suggest that the catalytic activity of *Arabidopsis* DPMS1 is more stringently modulated by DPMS3 than by DPMS2. These observations are in good agreement with previous data observed in the case of mammalian cells and in contrast with those reported for *S. cerevisiae*, where DPMS1 alone catalyzes the synthesis of Dol-P-Man without the assistance of additional proteins (Orlean et al., 1988).

To characterize further the role of DPMS1, DPMS2, and DPMS3, we searched independent insertion mutants corresponding to each locus among publicly available *Arabidopsis* T-DNA mutant databases. T-DNA insertion lines of DPMS1 (*dpms1-1* and *dpms1-2*), DPMS2 (*dpms2-1* and *dpms2-2*), and DPMS3 (*dpms3-1* and *dpms3-2*) were obtained. The resulting homozygous insertion mutants were identified by PCR, RT-PCR, and immunoblot analyses. *DPMS* transcripts did not accumulate

in the homozygous mutant lines, compared with wild-type plants (see Supplemental Figures 6A to 6C online). Further analysis using anti-DPMS1 antibodies revealed the absence of DPMS1 in *dpms1-1* and *dpms1-2* mutants (see Supplemental Figure 6A online). The capacity of each *dpms* mutant to catalyze the synthesis of Dol-P-Man was tested in vitro using crude microsomal membranes. The synthesis of Dol-P-Man could be demonstrated from microsomal membranes from *dpms2* and *dpms3* but was reduced to ~20 to 30% of the wild-type level (see Supplemental Figure 7A online). By contrast, in microsomal membranes isolated from *dpms1*, the synthesis of Dol-P-Man could not be demonstrated (see Supplemental Figure 7A online). The DPMS deficiency of *dpms1* plants was restored following the expression of DPMS1 in a *dpms1* mutant background (see Supplemental Figure 7A online). Thus, as shown above, in the absence of DPMS1, Dol-P-Man was not synthesized, regardless of the presence of functional DPMS2 and DPMS3.

Our data suggest that *Arabidopsis* DPMS1, DPMS2, and DPMS3 may interact or form a complex in vivo. The binding of DPMS1, DPMS2, and DPMS3 to each other has been previously demonstrated using mammalian cells transfected with epitope-tagged proteins (Maeda et al., 2000). To determine whether *Arabidopsis* DPMS proteins function as parts of a protein complex, we transiently coexpressed green fluorescent protein (GFP)-tagged *Arabidopsis* DPMS1, DPMS2, and DPMS3 in *Nicotiana benthamiana* leaves. As a control, leaves were transfected with *Agrobacterium tumefaciens* harboring an empty GFP vector. Total proteins of transfected leaves were solubilized with digitonin before immunoprecipitation using antibodies directed against *Arabidopsis* DPMS1 (anti-DPMS1) covalently linked to protein A-Sepharose. Following immunoprecipitation, bound proteins were eluted with an acidic buffer and subjected to immunoblot analysis using anti-GFP (Figure 4A). Immunoblot analysis of the immunoprecipitates obtained using anti-DPMS1 indicates that no immunoprecipitate was recovered from leaves transfected with the empty GFP vector (Figure 4A, lane 6). On the other hand, the anti-GFP antibodies revealed interactions between DPMS1, DPMS2, and DPMS3 (Figure 4A, lane 7). Collectively, these data indicate that DPMS1, DPMS2, and DPMS3 coimmunoprecipitate and probably interact in vivo. Using the same protocol, one could note interactions between DPMS1 + DPMS2 (Figure 4A, lane 8), DPMS1 + DPMS3 (Figure 4A, lane 9), and DPMS2 + DPMS3 (Figure 4A, lane 10). To assess the role of DPMS3 further, we analyzed its interaction with DPMS1 and DPMS2 using an in vitro pull-down assay. We used biotin-labeled DPMS3 and ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) as a control to pull down interacting proteins from an *Escherichia coli* extract expressing recombinant DPMS1 and DPMS2 as thioredoxin fusion proteins. After extensive washing to remove nonspecifically bound proteins, the acidic pH buffer eluate was subjected to SDS-PAGE analysis (Figure 4B). DPMS3 pulled down two prominent polypeptides having molecular masses of 43.8 kD (P1) and 25.2 kD (P2) (Figure 4B, lane 3) that were excised and used for in-gel tryptic digestion and sequence analysis by nanoliquid chromatography-MS/MS. The MS analysis of interacting proteins from the pull-down assay revealed that DPMS2 and DPMS1 are both DPMS3 binding proteins (see Supplemental Table 1 online). Further analysis using individually



**Figure 3.** UPLC-MS/MS Analysis of Dol<sub>14</sub>-P-Man Produced by the *Arabidopsis* DPMS Complex.

The reaction product obtained from the nonradioactive DPMS assay was subjected to DEAE-cellulose column chromatography and the fraction eluted with dichloromethane/methanol (7:3, v/v) containing ammonia (0.28 N), and 50 mM ammonium acetate was analyzed using two MRM modes with negative electrospray ionization (ESI-).

**(A)** Dol<sub>14</sub>-P-Man with the transition of  $m/z$  1196 > 1094 corresponding to [M-H<sub>2</sub>O-H-] = 1196 and the characteristic fragment of  $m/z$  = 1094 representing partial loss of Man [Dol<sub>14</sub>-P-(C<sub>2</sub>H<sub>3</sub>O)-H-].

**(B)** MRM transition of  $m/z$  1196 > 1052 corresponding to [M-H<sub>2</sub>O-H-] = 1196 and to another characteristic fragment of  $m/z$  = 1052, representing the loss of Man [Dol<sub>14</sub>-P-Man-H-].

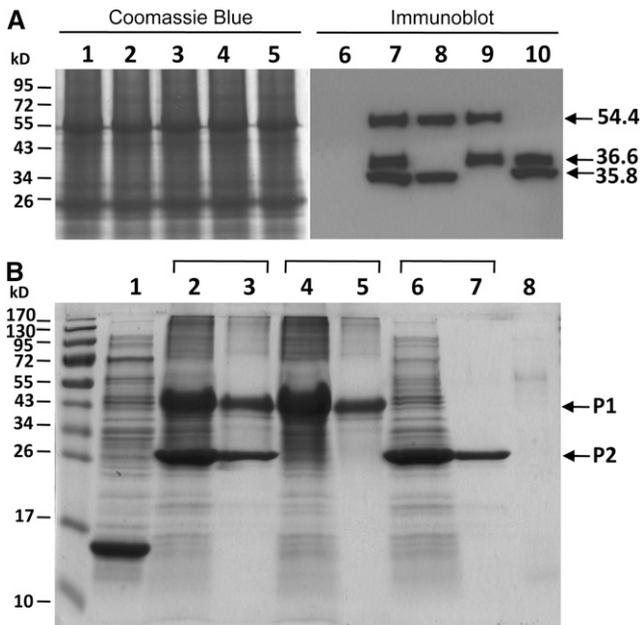
recombinant extracts containing DPMS1 (Figure 4B, lanes 4 and 5) or DPMS2 (Figure 4B, lanes 6 and 7) revealed that both interact with DPMS3. The specificity was ascertained by the fact that these polypeptides were not pulled down by Rubisco (Figure 4B, lane 8). These data are reminiscent of the behavior of mammalian DPMS proteins in which DPMS3 could represent the binding module for DPMS1 and DPMS2 (Ashida et al., 2006).

Finally, the subcellular localization of DPMS protein components was analyzed by confocal microscopy using GFP-tagged DPMS1, DPMS2, and DPMS3 that were transiently expressed in *Arabidopsis* leaves. The results indicate that DPMS1 as well as DPMS2 and DPMS3 were localized in the ER membranes like the GFP-tagged marker At-RTNLB2 (Nziengui et al., 2007) (Figure 5).

### Hypersensitivity of *dpms1* Plants to Ammonium

During the first step in characterizing *dpms* mutants, we observed that *dpms2* (*dpms2-1* and *dpms2-2*) and *dpms3* (*dpms3-1* and *dpms3-2*) plants were phenotypically indistinguishable from wild-type plants when grown on Murashige and Skoog (MS, M0221) medium or soil. On the other hand, *dpms1* (*dpms1-1* and *dpms1-2*) grown on MS (M0221) exhibited a chlorotic phenotype and developed short roots (Figures 6A and 6B). To analyze further the specificity of this phenomenon, we generated additional mutants by RNA interference (RNAi)-mediated reduction of

*DPMS1* expression (*DPMS1*-RNAi). We isolated three *DPMS1*-RNAi lines and confirmed the low expression of *DPMS1* by RT-PCR in lines 2 and 3 (see Supplemental Figure 6D online). Chlorophyll content and root growth were drastically reduced in *DPMS1*-RNAi3 plants (Figure 6B). The extent to which the DPMS activity was reduced in the *DPMS1*-RNAi lines was evaluated using microsomal membranes, revealing background activities in *DPMS1*-RNAi2 and *DPMS1*-RNAi3 lines compared with wild-type plants (see Supplemental Figure 7B online). Thus, the phenotype analysis of *DPMS1*-RNAi plants recapitulated that observed for T-DNA *dpms1* plants. Unexpectedly, upon transfer to soil, *dpms1* and *DPMS1*-RNAi plants progressively lost their chlorotic phenotype and accumulated chlorophylls to near wild-type levels (Figures 6C to 6G). The root growth of soil-grown *dpms1* plants was increased but remained shorter compared with that of wild-type plants. The seed coat of *dpms1* seeds displayed a wrinkled architecture (shown in Figure 8P). Functional complementation of the *dpms1* mutant phenotype was observed following transformation of *dpms1* plants with the *DPMS1* gene (Figure 6G). Taken together, these observations suggested that endogenous factors or the composition of the MS culture medium may play an important role in *dpms1* mutant phenotypes. After several modifications of the mineral composition of the medium, we established a correlation between the presence of 20.6 mM NH<sub>4</sub><sup>+</sup> in MS (M0221) and the appearance of

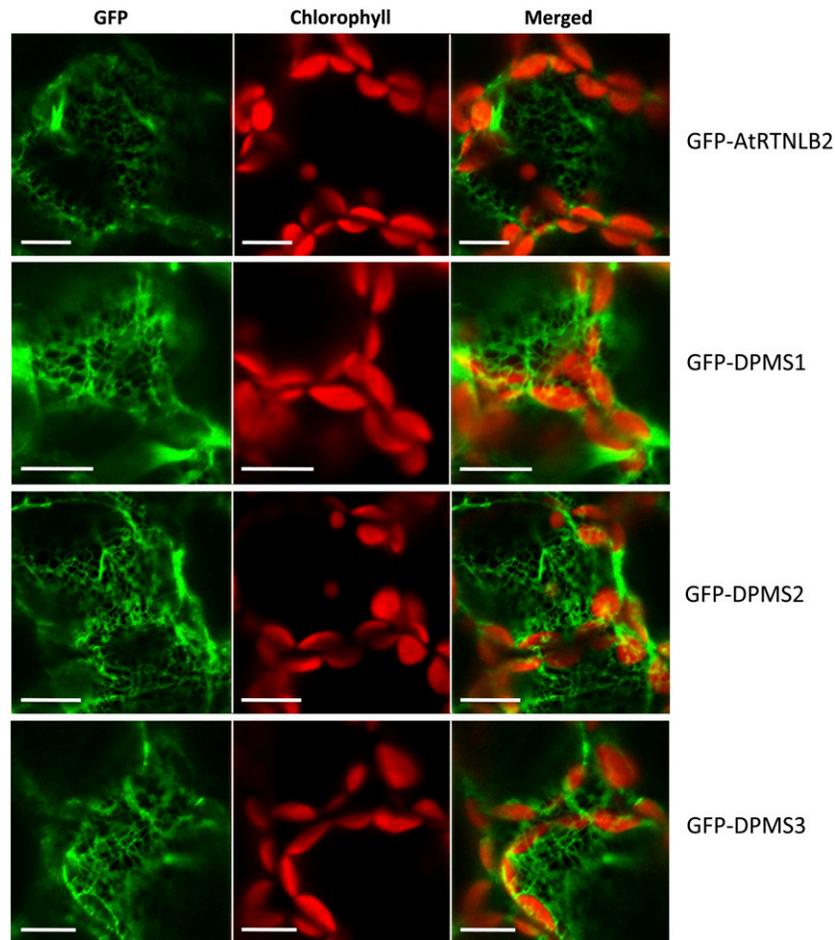


**Figure 4.** Interaction between DPMS Protein Components.

**(A)** *N. benthamiana* leaves were infiltrated with 1:1 or 1:1 mixtures of *Agrobacterium* strain GV3101 individually harboring the GFP-DPMS1, GFP-DPMS2, or GFP-DPMS3 constructs. Leaves were infected similarly with *Agrobacterium* harboring an empty GFP vector control. Infiltrated leaves were collected 72 h later for immunoblot analysis. Coomassie blue-stained protein bands of control (lane 1), GFP-DPMS1 + GFP-DPMS2 + GFP-DPMS3 transfected leaves (lane 2), GFP-DPMS1 + GFP-DPMS2 transfected leaves (lane 3), GFP-DPMS1 + GFP-DPMS3 transfected leaves (lane 4), and GFP-DPMS2 + GFP-DPMS3 transfected leaves (lane 5) were used for immunoprecipitation with anti-DPMS1. The anti-DPMS1 immunoprecipitates were used for immunoblot analysis with anti-GFP as shown: lane 6, immunoprecipitates from control leaves; lanes 7 to 10, immunoadsorbed fractions from leaves expressing GFP-DPMS1 + GFP-DPMS2 + GFP-DPMS3 (lane 7), GFP-DPMS1 + GFP-DPMS2 (lane 8), GFP-DPMS1 + GFP-DPMS3 (lane 9), and GFP-DPMS2 + GFP-DPMS3 (lane 10).

**(B)** Recombinant DPMS3 was biotinylated and used as a bait to pull down recombinant DPMS1 and DPMS2. For a negative control, Rubisco protein was used as bait. The soluble extracts obtained from *E. coli* overexpressing recombinant DPMS1 and DPMS2 as thioredoxin fusion proteins were incubated with purified DPMS3 or Rubisco immobilized on agarose beads for 2 h at 6°C before extensive washing and elution with acidic buffer, pH 2.8). SDS-PAGE and Coomassie blue-stained proteins from total proteins from induced bacteria harboring an empty vector (lane 1; *pBAD/TOPO ThioFusion*), 1:1 mixture of total proteins from induced bacteria expressing DPMS1 and DPMS2 (lane 2), DPMS3-interacting proteins (lane 3; P1, 43.8 kD and P2, 25.2 kD) that were subsequently subjected to nanoLC-MS/MS analysis; total proteins from induced bacteria expressing DPMS1 (lane 4), DPMS3-interacting protein (lane 5; P1, 43.8 kD), total proteins from induced bacteria expressing DPMS2 (lane 6), DPMS3-interacting protein (lane 7; P2, 25.2 kD), and Rubisco interacting proteins (lane 8).

chlorotic leaves. We compared the effects of MS culture media differing predominantly in having variable  $\text{NO}_3^-/\text{NH}_4^+$  ratios. When *dpms1* mutants were cultivated on MS (M 0238) medium (modification number 4,  $\text{NH}_4^+$  free) containing exclusively  $\text{NO}_3^-$  (39.39 mM) as mineral nitrogen, the chlorophyll content was retained, and the visible phenotype was indistinguishable from that observed for the wild-type plants, except that root growth was reduced to ~30 to 50% (see Supplemental Figure 8 online). To test further the sensitivity and the specificity of the response of *dpms1* and *DPMS1*-RNAi plants to  $\text{NH}_4^+$ , we analyzed the effect of increasing concentrations of  $\text{NH}_4^+$  in the culture medium. Cultivation on medium (MS, M 0238) containing increased concentrations of  $\text{NH}_4^+$  decreased both the chlorophyll pigments and root growth of *dpms1* and *DPMS1*-RNAi plants, compared with plants grown on the same medium devoid of  $\text{NH}_4^+$  (Figures 6H and 6I). The severity of the ammonium-induced effects on *dpms1* plants increased and became nonreversible and lethal with longer exposure time. These analyses revealed that the presence of  $\text{NH}_4^+$  in MS medium was responsible for the chlorotic phenotype of *dpms1* plants. The chlorophyll content of wild-type plants was not changed by any of the  $\text{NH}_4^+$  concentrations used; only root growth was slightly reduced at higher  $\text{NH}_4^+$  concentrations (Figure 6I). Thus, our data demonstrate that functional *DPMS1* is vital in the adaptation of *Arabidopsis* to increased external  $\text{NH}_4^+$  concentration. It has been previously shown that  $\text{NH}_4^+$  induces a dramatic decrease of root growth in *Arabidopsis* in the absence of  $\text{K}^+$  (Cao et al., 1993). As the concentration of  $\text{K}^+$  (20.04 mM) in the MS medium usually supports optimal *Arabidopsis* growth, we can conclude that  $\text{K}^+$  could not alleviate the toxic effect of  $\text{NH}_4^+$ . Therefore, our data suggest that *dpms1* mutants and *DPMS1*-RNAi plants are  $\text{NH}_4^+$  hypersensitive, regardless of the presence  $\text{K}^+$ . Our data unveil a previously unsuspected connection between the synthesis of Dol-P-Man and the sensitivity of plant to ammonium. To explore further the connection between the Dol-P-Man pathway and  $\text{NH}_4^+$  in *Arabidopsis*, we analyzed the behavior of *Arabidopsis srd5a3-like* mutants. *SRD5A3* encodes the reductase catalyzing the later steps of dolichol synthesis from polyprenol in mammals and yeast and probably in plants (Cantagrel et al., 2010) (see Supplemental Figure 9A online). The *Arabidopsis* genome contains two *Arabidopsis SRD5A3-LIKE* genes, At2g16530 and At1g72590 (see Supplemental Figure 10 online). We analyzed two independent homozygous knockout lines, *srd5a3-like1-1* and *srd5a3-like1-2*, corresponding to At2g16530 (see Supplemental Figures 9B and 9C online). We found that *srd5a3-like1* and *srd5a3-like2* display chlorotic phenotypes remarkably similar to those observed for *dpms1* (Figures 6A and 6B) when grown in the presence of ammonium (see Supplemental Figures 9D and 9E online). Although additional work is needed to evaluate fully the role of *Arabidopsis SRD5A3-LIKE* genes, these results extend the fact that knockout of *DPMS1* potentiates ammonium sensitivity in *Arabidopsis*. Previous studies established a link between GDP-Man pyrophosphorylase (GMPase) and ammonium sensitivity in *Arabidopsis* using *vtc1* and *hsn*, two allelic GMPase-deficient mutants (Qin et al., 2008; Barth et al., 2010). Therefore, we determined the GMPase activity of wild-type and *dpms1* mutant plants. The GMPase activity of wild-type plants ( $940 \pm 50$  nmol/h/mg protein) and *dpms1* plants ( $915 \pm 30$

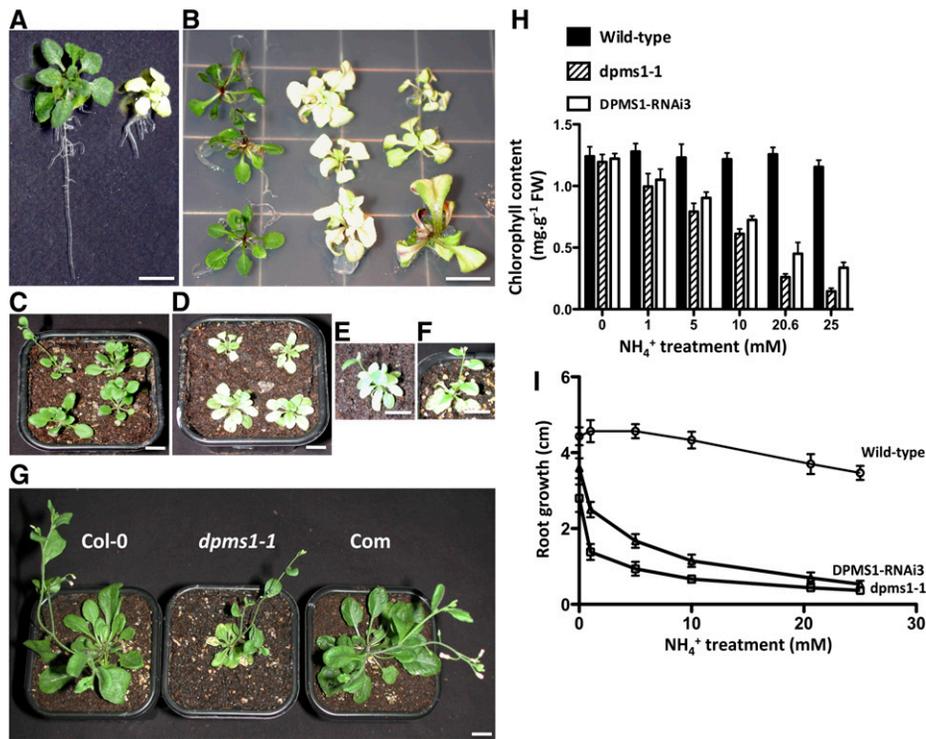


**Figure 5.** ER Localization of *Arabidopsis* DPMS Protein Components.

*Arabidopsis* leaves were transfected using GFP-tagged DPMS1, DPMS2, or DPMS3 (GFP-DPMS1, GFP-DPMS2, and GFP-DPMS3). The reticulon-like GFP-AtRTNLBL2 (Nziengui et al., 2007) was used as a positive marker of the ER membranes. Chloroplast autofluorescence is shown in red. Bars = 10  $\mu$ m.

nmol/h/mg protein) was nearly identical. Because ascorbic acid (AsA) is produced via GMPase activity, we determined the AsA content of wild-type and *dpms1* plants. The AsA content of *dpms1* plants ( $2700 \pm 200$  nmol/g fresh weight) was lower than that of wild-type plants ( $3500 \pm 300$  nmol/g fresh weight). However, when *dpms1* plants were supplemented with 5 mM galactono-1,4-lactone, the proximal AsA precursor, the wild-type phenotype was not restored. Therefore, our data indicate that AsA deficiency could not account for the *dpms1* mutant phenotype. In addition, the data suggest that when Dol-P-Man is not synthesized in *Arabidopsis*, the resulting GDP-Man precursor is not redirected toward the synthesis of AsA. Thus, downstream steps involving the use of Dol-P-Man are probably key determinants of *Arabidopsis*  $\text{NH}_4^+$  sensitivity. To analyze further the consequence of DPMS1 deficiency, total proteins from wild-type and *dpms1* were subjected to affinity detection using concanavalin A, a high Man-specific lectin (Faye and Chrispeels, 1985). Several reactive bands corresponding to glycoproteins bearing *N*-linked glycans could be observed in the wild-type and in the *dpms1* mutant. The different electrophoretic mobility of

reactive polypeptides suggests a modified glycosylation status between wild-type plants and *dpms1* mutants (see Supplemental Figure 11 online). Decreased Dol-P-Man availability makes it likely that *N*-glycosylation and GPI anchor attachment are perturbed or impaired. We therefore tested both types of post-translational modifications using anti-Protein Disulfide Isomerase (PDI) (Lukowitz et al., 2001) and anti-Plasmodesmata Callose Binding (PDCB1) protein (Simpson et al., 2009) antibodies as genuine markers of *N*-glycosylated and GPI-anchored proteins, respectively. Immunoblot analysis revealed that PDI migrated slightly faster in *dpms1-1* and *dpms1-2* mutants than in the wild-type plants, consistent with a modified *N*-glycosylation pattern (Figures 7A and 7B). On the other hand, for PDCB1 we detected in the wild type several polypeptide bands with a molecular mass of 19 to 22 kD (Figures 7C and 7D) due to the fact that PDCB is encoded by a multigene family in *Arabidopsis* (Simpson et al., 2009). In *dpms1-1* and *dpms1-2*, the lower molecular mass PDCBs was not detected, whereas the content of the higher molecular mass species was reduced, thus suggesting a GPI anchor deficiency (Figures 7C and 7D).



**Figure 6.** Ammonium Sensitivity of *DPMS1* Mutants.

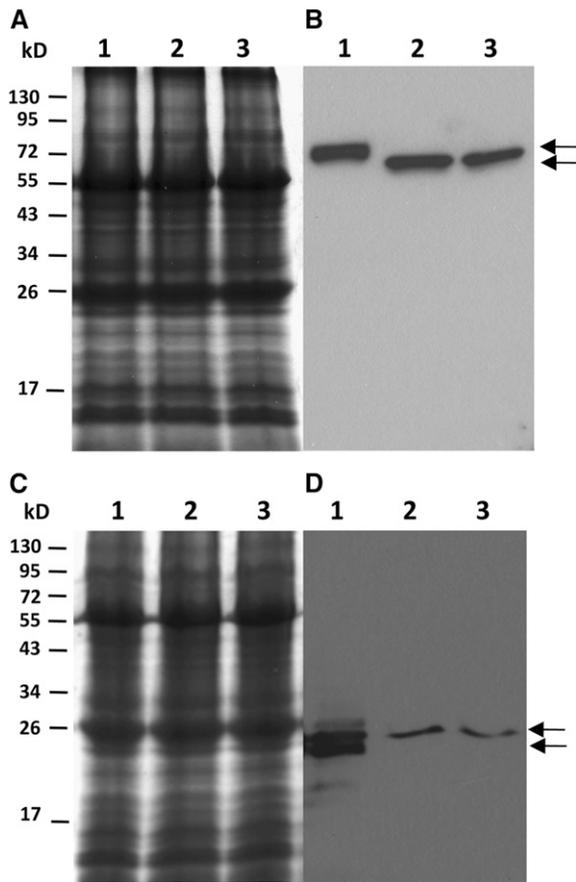
(A) Aerial and root phenotypes of wild-type (left) and *dpms1-1* (right) plants grown in vitro on MS (M0221) agar medium containing 20.6 mM NH<sub>4</sub><sup>+</sup> for 26 d. (B) Phenotypes of wild-type (left column), *dpms1* (middle column, from the top to bottom *dpms1-1*, *dpms1-1*, and *dpms1-2*), and *DPMS1*-RNAi (right column, from the top to bottom, lines 2, 2, and 3) plants grown in vitro on MS (M0221) agar medium containing 20.6 mM NH<sub>4</sub><sup>+</sup> for 20 d. (C) Wild-type plants first grown in vitro on MS (M0221) agar medium containing 20.6 mM NH<sub>4</sub><sup>+</sup> for 27 d were transferred to soil and grown for an additional 7 d. (D) *dpms1-1* and *dpms1-2* (left column, top to bottom), *DPMS1*-RNAi2, and *DPMS1*-RNAi3 (right column, top to bottom) first grown in vitro on MS (M0221) agar medium containing 20.6 mM NH<sub>4</sub><sup>+</sup> for 27 d were transferred to soil and grown for an additional 7 d. (E) and (F) *dpms1-1* (E) and *DPMS1*-RNAi3 (F) first grown in vitro on MS (M0221) agar medium containing 20.6 mM NH<sub>4</sub><sup>+</sup> for 27 d were transferred to soil and grown for an additional 14 d. (G) Wild-type plants (Col-0), *dpms1-1* mutant (*dpms1-1*), and *dpms1-1* plants complemented with 35S:*DPMS1* (Com) first grown in vitro on MS (M0221) agar medium containing 20.6 mM NH<sub>4</sub><sup>+</sup> for 27 d were transferred to soil and grown for an additional 25 d. (H) Effect of NH<sub>4</sub><sup>+</sup> on chlorophyll content in wild-type, *dpms1-1*, and *DPMS1*-RNAi3 plants grown for 21 d. Results represent means values ± SD from triplicates. (I) Effect of NH<sub>4</sub><sup>+</sup> on root growth of wild-type, *dpms1*, and *DPMS1*-RNAi3 plants grown for 21 d. Results represent means values ± SD from triplicates. Bars = 1 cm.

### Effect of Modulated Expression of *DPMS1*

Loss-of-function mutants and functional rescue of a mutant phenotype have been widely used to assess the role of plant gene products. The limitations of this strategy using only insertional mutants have been discussed previously (Hirschi, 2003). Previous data have shown that the availability of Dol-P-Man modulates the structure and the composition of *S. cerevisiae* cell walls (van Berkel et al., 1999; Ortowski et al., 2007). Therefore, we decided to generate *Arabidopsis* lines overexpressing *DPMS1* using a Rubisco promoter that possesses enhancer elements activated by several plant transcription factors. In two selected lines, the overexpression of *DPMS1* occasionally resulted in altered stem branch diameter and morphology (Figures 8A to 8D). This phenomenon was paralleled by the appearance of wrinkled seed coat

architecture and by perturbation of the vascular bundle arrangements very similar to that observed in the *clavata1* mutant that is altered in a receptor kinase gene (Deyoung and Clark, 2008) (Figures 8E to 8P). Notably, the seed coat architecture of overexpressing lines nearly matched that of *dpms1* mutants and is indicative of the formation of aberrant cell walls.

Because Dol-P-Man is at the convergence of several pathways, we examined how overexpression of *DPMS1* might influence the expression of several genes known to encode enzymes implicated in or catalyzing upstream and downstream steps of glycan synthesis. These included the biosynthetic enzyme GDP-Man pyrophosphorylase (also known as VTC1, for vitamin C1) (Conklin et al., 1997), dolichol prenyltransferase (DoIS) or Leaf Wilting 1 (LEW1) (Zhang et al., 2008), UDP-*N*-acetylglucosamine:dolichol



**Figure 7.** Immunoblot Analysis of Total Leaf Proteins Using Antibodies Directed against *N*-Linked Glycan PDI or GPI-Anchored PDCB1.

- (A) Coomassie blue-stained proteins extracted from the wild type (lane 1), *dpms1-1* (lane 2), and *dpms1-2* (lane 3).  
 (B) Immunodetection of PDI in the wild type (lane 1), *dpms1-1* (lane 2), and *dpms1-2* (lane 3). Arrows indicate electrophoretic mobilities of polypeptides.  
 (C) Coomassie blue-stained proteins extracted from the wild type (lane 1), *dpms1-1* (lane 2), and *dpms1-2* (lane 3).  
 (D) Immunodetection of PDCB1 in the wild type (lane 1), *dpms1-1* (lane 2), and *dpms1-2* (lane 3). Arrows indicate electrophoretic mobilities of polypeptides.

phosphate *N*-acetylglucosamine-1-P transferase (GPT) (Koizumi et al., 1999), two mannosyltransferases called Asn-linked glycosylation (ALG) genes, ALG11 (Zhang et al., 2009) and ALG3 (Henquet et al., 2008), and the *N*-linked glycan trimming enzymes,  $\alpha$ -glucosidase I (GCS1) and an uncharacterized  $\alpha$ -glucosidase-like enzyme (GCSL) (Boisson et al., 2001). Real-time RT-PCR using soil-grown plants revealed that *VTC1*, *DoIS*, *DPMS1*, *DPMS2*, *DPMS3*, *ALG11*, *ALG3*, *GPT*, and *GCS1*, which sequentially mediate the synthesis of Dol-P-Man and its downstream processing, were induced in the *DPMS1*-overexpressing lines compared with *dpms1* mutants as shown by the expression profile observed for *DPMS1-Ov2* and *dpms1-1* (Figure 9A). Except for *ALG11* and *ALG3*, a similar trend was observed between the *DPMS1*-overexpressing lines and wild-type plants (Figure 9A).

## DPMS and ER Stress Status

It has been shown in yeast that the alteration in the protein glycosylation pathways alone induces a buildup of misfolded proteins in the ER lumen and leads to ER stress. During ER stress, the most prominent change is represented by the transcriptional induction of genes controlling the pathway known as the unfolded protein response (UPR) (Ng et al., 2000). For instance, the UPR is constitutively triggered in the dolichol phosphate Man-deficient chinese hamster ovary (Foulquier et al., 2002). It has also been reported that in fibroblasts from patients suffering from congenital disorders of glycosylation type-I, the UPR is moderately but constitutively induced (Lecca et al., 2005). Finally, *Arabidopsis lew* mutants affected in the biogenesis of dolichols are more sensitive to tunicamycin (Zhang et al., 2008), which is an inhibitor of *N*-acetylglucosamine transferase that acts as a potent inducer of the UPR (Martinez and Chrispeels, 2003; Iwata and Koizumi, 2005). Thus, due to the central role of Dol-P-Man in protein glycosylation, we tested whether *DPMS1*-deficient or *DPMS1*-overexpressing *Arabidopsis* plants have an altered UPR in the presence or absence of tunicamycin. We analyzed the expression of several ER stress marker genes, including UPR marker genes using soil-grown plants or plants treated with tunicamycin. Marker genes included three bZIP membrane-bound transcription factors (bZIP17, bZIP28, and bZIP60) that globally regulate the induction of ER stress response genes (Iwata and Koizumi, 2005; Che et al., 2010; Liu and Howell, 2010), including the UPR markers Bip2, Bip3, calnexin, and calreticulin. Gene expression monitored by real-time RT-PCR revealed that UPR marker genes were constitutively expressed in *DPMS1*-overexpressing plants compared with *dpms1* and wild-type plants (Figure 9B). The expression profile of *DPMS1*-overexpressing plants was paralleled by the induced expression of two transcriptional regulators of oxidative stress responses, RCD1 (Katiyar-Agarwal et al., 2006) and RD29A (Liu et al., 1998), suggesting that the ER stress in *DPMS1*-overexpressing plants is balanced by concomitant upregulation of antioxidant pathways. When plants were treated with tunicamycin to induce the UPR, the magnitude of the response of the genes encoding bZIP17, bZIP28, and bZIP60 transcription factors, which globally regulate the stress, was lower in *DPMS1*-overexpressing plants compared with *dpms1* (Figure 9C). These data are consistent with ER stress being already present in plants overexpressing *DPMS1* and suggest that *dpms1* and wild-type plants are more responsive to tunicamycin treatment than are *DPMS1*-overexpressing plants.

## DISCUSSION

### Multimodular Organization of Plant DPMS

Man residues transported by Dol-P play a key role in the biosynthesis of *N*-glycoproteins, *O*- and *C*-mannosylated proteins, and GPI-anchored proteins. In addition, the removal of the terminal Man residue donated by Dol-P-Man is the last enzymatic step before glycoproteins can exit the ER. If folding is not complete at this step, the attached glycan dictates the



**Figure 8.** Phenotypic Analysis of *DPMS1* Overexpression in *Arabidopsis*.

(A) and (B) Close-up view of the wild type (A) and a *DPMS1*-overexpressing plant *DPMS1*-Ov1 (B).

(C) Cylindrical stem of wild-type (Col-0) plants and flattened stem of *DPMS1*-Ov1 (Ov1) and *DPMS1*-Ov3 (Ov3).

(D) Higher magnification of (C).

(E) to (G) Sections of the flattened stem of *DPMS1*-Ov1 stained with phloroglucinol (pink color) to detect lignin in the secondary cell walls of vascular tissues.

(H) to (J) Stem sections of *dpms1-1* plants stained as in (E) to (G).

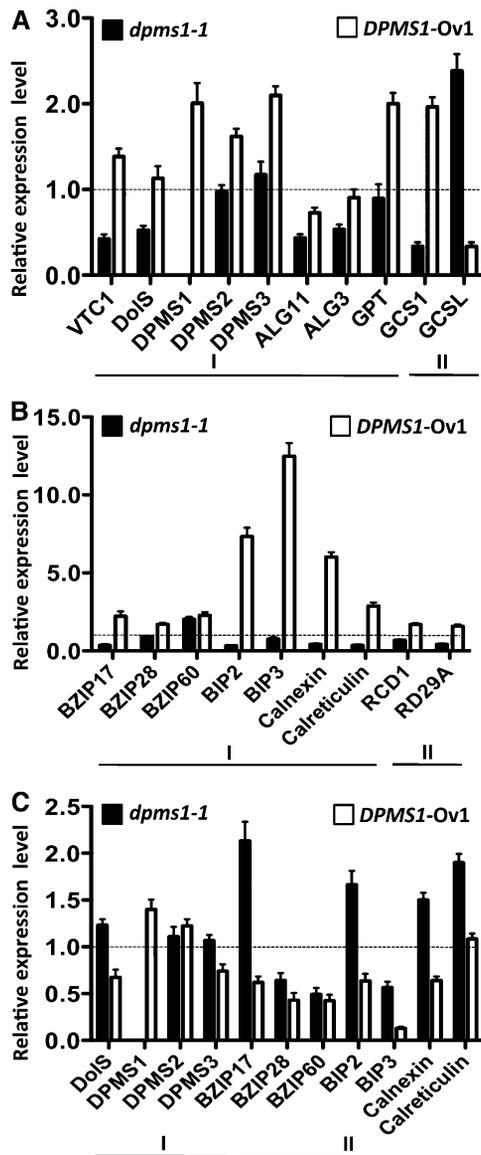
(K) to (M) Stem sections of wild-type plants stained as in (E) to (G).

(N) to (P) Scanning electron micrographs of seed coat from wild-type plants (N) compared with wrinkled seed coats of *DPMS1*-Ov1 (O) and *dpms1-1* (P) plants.

Bars = 1.5 cm in (A) and (B), 0.5 cm in (C) and (D), 250  $\mu$ m for stem sections in (E) to (M), and 100  $\mu$ m for seeds in (N) to (P).

degradation of the polypeptide chain (Tokunaga et al., 2000; Lehrman, 2001; Molinari, 2007) (Figure 1). Although the synthesis of Dol-P-Man has been studied in detail in *S. cerevisiae* and in human cells, the protein components catalyzing its synthesis in plants were unknown. Two prototypic organizations of DPMS have been proposed. The first is represented by the yeast enzyme, which comprises a homomeric protein called DPMS1 (Maeda and Kinoshita, 2008). The second type is represented by the mammalian enzyme, which is composed of three proteins, including DPMS1, DPMS2, and DPMS3 (Maeda and Kinoshita, 2008). We determined the component enzymes involved in the synthesis of Dol-P-Man from *Arabidopsis*. We provide evidence that *Arabidopsis* DPMS1, DPMS2, and DPMS3 are functional orthologs of mammalian genes involved in Dol-P-Man synthesis.

Like human DPMS1, we show that *Arabidopsis* DPMS1 has no activity when expressed in *E. coli*, thereby pointing to a functional difference with *S. cerevisiae* (Colussi et al., 1997; Maeda and Kinoshita, 2008). As observed in mammalian cells (Tomita et al., 1998), the synthesis of Dol-P-Man could be demonstrated in vitro using a reconstituted enzyme system containing DPMS1, DPMS2, and DPMS3 (Figure 2). To examine the contribution of each protein, we tested the activity of DPMS1+DPMS2, DPMS1+DPMS3, and DPMS2+DPMS3. We observed that whereas the combinations of DPMS1+DPMS2 and DPMS1+DPMS3 displayed 5 and 20% reduced synthesis of Dol-P-Man, respectively, compared with DPMS1+DPMS2+DPMS3, the DPMS2+DPMS3 combination was inactive (Figure 2). Taken together, these data suggest that *Arabidopsis* DPMS1 represents the catalytic core of



**Figure 9.** Quantitative Analysis of the Expression of Dolichol and Glycan Biosynthetic Genes and Stress Marker Genes.

**(A)** Real-time PCR quantification of dolichol and glycan biosynthetic genes and stress marker genes from 32-d-old soil-grown wild-type, *dpms1-1*, and *DPMS1*-overexpressing (*DPMS1-Ov1*) plants. The expression values for *dpms1-1* and *DPMS1-Ov1* are shown relative to the wild type, which is set to 1 (dashed line). Results are calculated from means values  $\pm$  SD of triplicates from three independent biological samples. Dol-P-Man and glycan pathway genes (zone I): *GDP-Man* pyrophosphorylase/*Man-1*-pyrophosphatase *VTC1*, *Dolichol* synthase *DolS*, *DPMS1*, *DPMS2*, *DPMS3*, *mannosyltransferases* *ALG11* and *ALG3*, *GPT*. Glycan trimming genes (zone II): *GCS1*,  $\alpha$ -*glucosidase* *GCSL*.

**(B)** Real-time PCR quantification of stress marker genes from 32-d-old soil-grown wild-type, *dpms1-1*, and *DPMS1*-overexpressing (*DPMS1-Ov1*) lines. The expression conditions are presented as in **(A)**. UPR and ER stress markers (zone I): *BZIP17*, *BZIP28*, *BZIP60*, *Bip2*, *Bip3*, *Calnexin*, and *Calreticulin*. Oxidative stress markers (Zone II): *RCD1* and *RD29A*.

the enzyme complex including *DPMS2* and *DPMS3*, as shown in mammalian cells (Maeda and Kinoshita, 2008). Consistent with these data, we show that endogenous *DPMS1*, *DPMS2*, and *DPMS3* are compartmentalized in the ER membranes (Figure 5) and probably form a complex in vivo as shown by pull-down assay and affinity interaction analyses (Figure 4). Although *DPMS2* is associated with catalytically active *DPMS* complex, its precise function is unknown. It has been suggested that it may stabilize the *DPMS* complex (Maeda and Kinoshita, 2008). Our data suggest that the trimerization of *DPMS* could be facilitated by the fact that *DPMS3* interacts with both *DPMS1* and *DPMS2* (Figure 4). Because *DPMS2* enhances up to twofold the activity of *GPI-N*-acetylglucosaminyltransferase, which is involved in the biosynthesis of the *GPI* anchor, *DPMS2* has also been considered as a regulator of the *GPI* pathway (Watanabe et al., 2000; Maeda and Kinoshita, 2008). This led to the suggestion that the biosynthesis of *GPI* and *Dol-P-Man* are probably coregulated (Orlean and Menon, 2007). It has also been shown that the binding of *Dol-P* is increased in the presence of *DPMS2* (Maeda et al., 1998). In mammalian cells, deficiency of *DPMS1* (Sugiyama et al., 1991; Tomita et al., 1998), *DPMS2* (Maeda et al., 1998), and *DPMS3* (Ashida et al., 2006) drastically blocks or reduces the biosynthesis of *Dol-P-Man* and leads to *N*-glycosylation defects. The requirement for *DPMS3* is supported by the fact that mammalian cell *DPMS3* acts as a coiled-coil protein that tethers *DPMS1* to the ER membranes (Ashida et al., 2006; Maeda and Kinoshita, 2008; Lefeber et al., 2009). Furthermore, a single missense mutation in mammalian *DPMS3* drastically reduces the catalytic activity of *DPMS1* and leads to muscular dystrophy (Lefeber et al., 2009). In mammalian cells, it has also been shown that stable expression of *DPMS1* requires *DPMS3*, whereas *DPMS2* is required for the stable expression of *DPMS3* (Maeda et al., 1998; Maeda et al., 2000; Ashida et al., 2006). Notably, mammalian *DPMS1* is degraded by the proteasome in the absence of *DPMS3* (Ashida et al., 2006). Interestingly, database (<http://atted.jp/>) searches revealed that in *Arabidopsis*, the expression of *DPMS1* is correlated with that of *DPMS3*. *DPMS3* ranks third on the list with a Pearson's correlation coefficient of 0.56. However, in *Arabidopsis*, *dpms2* and *dpms3* mutants still retain *DPMS* activity, thus pointing to a key difference between plants and animals (see Supplemental Figure 7A online).

### Multiple Roles for *DPMS* in Plants

The biogenesis of *Dol-P-Man* involves the cooperation of several plant cell compartments. Previous studies led to the isolation of

**(C)** Real-time PCR quantification of UPR-induced marker genes following tunicamycin treatment using 21-d-old MS-grown wild-type, *dpms1-1*, and *DPMS1*-overexpressing (*DPMS1-Ov1*) plants. The expression values are presented relative to the ratio  $R = \text{gene expression in tunicamycin-treated wild type} / \text{gene expression in untreated wild type}$ , which is set to 1 (dashed line). Results are calculated from means values  $\pm$  SD of triplicates from three independent biological samples. Dol-P-Man pathway genes (zone I): *DolS*, *DPMS1*, *DPMS2*, and *DPMS3*. UPR and ER stress markers (zone II): *BZIP17*, *BZIP28*, *BZIP60*, *Bip2*, *Bip3*, *Calnexin*, and *Calreticulin*.

*Arabidopsis LEW1*, which encodes a *cis*-prenyltransferase involved in the biosynthesis of dolichol (Zhang et al., 2008). *lew1* mutants display a leaf-wilting phenotype (Zhang et al., 2008). We show that among the three genes encoding the *Arabidopsis* DPMS enzyme complex, only mutations in the *DPMS1* gene encoding the catalytic module lead to an apparent phenotype, characterized by extensive chlorosis, reduction of root growth, and a wrinkled seed coat under normal culture conditions (Figures 6 and 8). Thus, our data point to nonlethal phenotypes when DPMSs are knocked out. Interestingly, *Arabidopsis* ALG3, which mediates the incorporation of the 6th Man of *N*-glycans from Dol-P-Man, has been identified (Henquet et al., 2008; Kajiura et al., 2010). However, *alg3* mutants are undistinguishable from the wild-type plants when grown under normal or stress conditions, including salt stress and low and high temperatures. *Arabidopsis* ALG12/*EBS4* encodes the enzyme sequentially catalyzing the transfer of the (8th) Man residue from Dol-P-Man to yield the glycan intermediate Man<sub>8</sub>GlcNAc<sub>2</sub>-PP-Dol (Hong et al., 2009). In *alg12/eps4* mutants, incomplete glycans are transferred to polypeptides, and the degradation of defective brassinosteroid receptors is inhibited (Hong et al., 2009). Thus, in plants, mutations affecting the synthesis and the attachment of protein glycans are silent or associated with diverse developmental and physiological deficiencies that could lead to lethal phenotypes (Pattison and Amtmann, 2009; Zhang et al., 2009). Interestingly, short or premature glycan chains could be attached to glycoproteins in plants (Henquet et al., 2008). By contrast, *DPMS1* is essential for cell viability in *S. cerevisiae* (Orlean, 1992), whereas knockout of mice *DPMS2* blocks the synthesis of the GPI anchor and is embryo-lethal (Nozaki et al., 1999). Several human congenital disorders associated with Dol-P-Man biosynthesis and metabolism have been described (Haeuptle and Hennet, 2009). Detailed analysis revealed that in humans, Dol-P-Man deficiency gradually affects two subpathways of protein glycosylation. Mild Dol-P-Man deficiency due to *DPMS3* mutations affects mainly *O*-mannosylation pathways and leads to  $\alpha$ -dystroglycanopathies. On the other hand, severe Dol-P-Man deficiencies perturb *N*-glycosylation and GPI pathways and provoke abnormal clotting and epilepsy (Lefeber et al., 2009). *DPMS1* deficiency caused a reduction of PDCB1 accumulation, consistent with an impairment of GPI anchor biosynthesis (Figures 7C and 7D). This feature is consistent with previous reports showing that the impairment of GPI attachment to phospholipase C in *T. brucei* (Garg et al., 1997) or yeast *Gas1p* (Meyer et al., 2000) leads to luminal ER retention, premature secretion, or degradation of proteins. Dol-P-Man is the key donor for the synthesis of the core glycan of GPI anchors (Orlean and Menon, 2007). *Arabidopsis* mutants affected in the early steps of GPI anchor synthesis have been characterized. These include *peanut*, which encodes GPI-mannosyl transferase I that catalyzes the transfer of the first mannosyl residues to the GPI anchor (Gillmor et al., 2005), and *seth1* and *seth2*, which encode protein components of GPI-*N*-acetylglucosaminyltransferase (Lalanne et al., 2004). In contrast with *dpms1*, *Arabidopsis* *peanut* is lethal (Gillmor et al., 2005), and *seth1* and *seth2* are male sterile (Lalanne et al., 2004). GPI anchor synthesis is complex and requires at least 18 genes (Orlean and Menon, 2007). Due to the complexity of the pathway, few studies have

been made using purified enzymes. For instance, GPI-mannosyltransferase I, which catalyzes the first incorporation of a Man residue into the GPI anchor, is a multicomponent protein complex (Orlean and Menon, 2007). Data from human, yeast, *Plasmodium falciparum*, and *T. brucei* reveal that the protein components are species specific and are not interchangeable (Kim et al., 2007), thus making cross-study comparisons difficult. A potentially attractive possibility is that plants could synthesize dolichol or lipid-linked phosphate Man through an alternative pathway. In this view, it is worth noting that the stringent requirement for the presence of a reduced  $\alpha$ -isoprene unit in Dol-P-Man can be fulfilled by a monoterpenoid derivative, such as citronellol phosphate (Rush et al., 1993), synthetic lipid phosphates (Wilson et al., 1993), and various derivatives (Sprung et al., 2003). Finally, the *Arabidopsis* genome contains several genes encoding Golgi-localized nucleotide sugar transporters (GONSTs) predicted to import GDP-Man in the Golgi (Handford et al., 2004). Thus, one cannot exclude the transport of GDP-Man in the ER lumen via GONST-like transporters and its use in the biosynthesis of *N*-glycoproteins and GPI-anchored proteins in *dpms1* plants.

We show that the overexpression of *DPMS1* could change the architecture of *Arabidopsis* stems and the vascular bundle arrangements and provoke the formation of a wrinkled seed coat (Figure 8), consistent with the important role played by GPI-anchored proteins in cell wall formation (Gillmor et al., 2005). This phenomenon is paralleled by a broad induction of the expression of several genes encoding the enzymes catalyzing individual steps of the *N*-glycan pathway (Figure 9A) and is associated with constitutive expression of the UPR marker genes (Figure 9B). These modifications probably reflect excess or aberrant glycosylations. Interestingly, the incorporation of Man in the cell walls of potato tuber was reduced to 30 to 40% of the wild-type levels following antisense expression of GMPase (Keller et al., 1999). Consistent with these data, the reduction of dolichol-linked oligosaccharides to 20% of the wild type results in altered cell walls in the *S. cerevisiae* mutant (*cwh*) (van Berkel et al., 1999). One could also note that in *S. cerevisiae* overexpressing the dolichol prenyltransferase gene *RER2*, the synthesis of Dol-P-Man and *N*-protein glycosylation were increased through the stimulation of several glycan biosynthetic enzymes ALG7, ALG13, and ALG14. This trend was paralleled by the formation of an aberrant cell wall due to alteration in its composition (Orłowski et al., 2007). Our finding may also apply to the overexpression of *S. cerevisiae* *DPMS1* in the fungus *T. reesei*. This resulted in the formation of a floccular cell wall in the transformants compared with the compact wall structure of the wild type (Kruszewska et al., 1999). This change was associated with increased glycoprotein secretions (Kruszewska et al., 1999). In a similar vein, correlations between increased DPMS activity and *N*-glycosylation have been documented using mammalian cells (Banerjee et al., 1987; Carson et al., 1990). Interestingly, it has been shown that *S. cerevisiae* and mammal DPMSs are upregulated by phosphorylation (Banerjee et al., 2005; Baksi et al., 2008). The Ser residue identified as the site of phosphorylation is conserved from protozoan parasites to human, including in plants, as shown in the *DPMS1* sequence MRKLTS (underlined in Supplemental Figure 1 online).

### DPMS1, a Candidate Susceptibility Gene for Ammonium Toxicity

Although primary nitrogen assimilation in plants proceeds via its reduced form (i.e., ammonium), *Arabidopsis*, like the other members of the Brassicaceae, is sensitive to the presence of ammonium in the culture medium (Britto and Kronzucker, 2002). Interestingly, an allosteric regulation of ammonium transporter has been demonstrated that involves the phosphorylation of the *Arabidopsis* ammonium transporter AMT1 to reduce ammonium absorption (Lanquar et al., 2009). We show that the Dol-P-Man pathway mediated by DPMS1, DPMS2, and DPMS3 is linked to ammonium sensitivity in *Arabidopsis*. Whereas the loss-of-function mutants *dpms2* and *dpms3* retained a wild-type phenotype, we found that the loss-of-function mutant *dpms1* and DPMS1-RNAi exacerbate the hypersensitivity of *Arabidopsis* to ammonium and, as a result, the leaves become strongly chlorotic and the root growth is drastically reduced. The strong depletion of DPMS1 transcript and protein coincided with background DPMS activity in DPMS1-RNAi2 and DPMS1-RNAi3 lines (see Supplemental Figures 6D and 7B online).

On the other hand, *dpms2* and *dpms3* have ~20 to 30% of DPMS activity and do not display a bleached phenotype. Extrapolation from in vitro data to the in vivo situation is difficult because the integration of DPMS1 in the ER membranes is probably not optimal in the absence of DPMS2 or DPMS3. Thus, the resulting topological changes are likely amplified in vitro and could lower the DPMS activities measured from *dpms2* and *dpms3* microsomes. Alternatively, it could be possible that the reduction of DPMS activity to ~20 to 30% of wild-type activity did not confer sensitivity to ammonium. The chlorotic phenotype disappears when *dpms1* mutant plants were grown on soil, consistent with the usually low ammonium concentration, which has been estimated to be 10 to 1000 times lower than that of nitrate (Marshner, 1995). The mechanism underlying the hypersensitivity to NH<sub>4</sub><sup>+</sup> is not well understood. Multiple factors have been proposed to account for ammonium toxicity in plants. The toxicity has been attributed to increased proton efflux and acidification of the external medium and to the reduction of dicarboxylic acid and cation concentrations (Barker et al., 1966; Kirkby, 1968; Raven and Smith, 1976). It has also been proposed that the high demand for carbohydrate skeletons, such as oxaloacetate and ketoglutarate, for nitrogen incorporation may lead carbon shortage (Schjoerring et al., 2002). Perturbation of hormonal homeostasis has also been proposed (Barker, 1999). Finally, it is well known that the toxicity of NH<sub>4</sub><sup>+</sup> is exacerbated by K<sup>+</sup> deprivation (Wall, 1939) and that K<sup>+</sup> can alleviate the toxic effect of NH<sub>4</sub><sup>+</sup> (Cao et al., 1993). In addition to their physicochemical properties (i.e., charge, size, and hydration energy) (Wang et al., 1996; White, 1996), a reciprocal influence between K<sup>+</sup> and NH<sub>4</sub><sup>+</sup> has been described during their absorption (ten Hoopen et al., 2010). Indeed, the *Arabidopsis* potassium transporter KAT1 has affinity for NH<sub>4</sub><sup>+</sup> (Schachtman et al., 1992). This could contribute to the chlorotic phenotype and the severe reduction of root growth of *dpms1* plants in the basal MS medium (M0221), which contains 20.04 mM K<sup>+</sup> and 20.61 NH<sub>4</sub><sup>+</sup> (Figure 6).

*Arabidopsis* GDP-GMPase, also known as VTC1, catalyzes the conversion of D-Man-1-P into GDP-Man. It has been shown recently that mutation in the *Arabidopsis* GDP-GMPase gene causes NH<sub>4</sub><sup>+</sup> hypersensitivity associated with the inhibition of root growth (Qin et al., 2008; Barth et al., 2010). It has been suggested that the strong inhibitory effect of NH<sub>4</sub><sup>+</sup> on GMPase activity could account for NH<sub>4</sub><sup>+</sup> hypersensitivity (Qin et al., 2008). Because, DPMS1 acts downstream of GMPase and due to the fact that DPMS1 and *dpms1* plants have the same GMPase activity, our data indicate the sensitivity of GMPase to NH<sub>4</sub><sup>+</sup> alone is not sufficient to explain the NH<sub>4</sub><sup>+</sup> sensitivity of *Arabidopsis*. Additionally, our data from *dpms1* and those obtained from *vtc1* plants (Barth et al., 2010) suggest that AsA is not directly implicated in the NH<sub>4</sub><sup>+</sup> sensitivity. Thus, the NH<sub>4</sub><sup>+</sup> sensitivity results primarily from the perturbation of Dol-P-Man biogenesis and its downstream utilization, rather than at the level of the branch leading to AsA. Because Dol-P-man is required for the synthesis of N-glycoproteins, GPI-anchored proteins and arabinogalactan proteins, one may reasonably expect that these different pathways are affected. GPI-anchored proteins are overrepresented in lipid rafts that are relatively enriched in phytosterols (Kierszniowska et al., 2009) that could modulate the function of mineral transporters. With respect to the chlorotic phenotype, it is worth noting that some *Arabidopsis* and rice proteins destined for the plastids are posttranslationally modified by N-glycosylation (Villarejo et al., 2005; Nanjo et al., 2006; Radhamony and Theg, 2006). Rice α-amylase, for instance, is a glycoprotein transported from the ER-Golgi compartment to plastids via the secretory pathway (Kitajima et al., 2009). In addition, it has been shown recently that the blockage of the secretory pathway by brefeldin A induces starch accumulation in *Arabidopsis* plastids (Hummel et al., 2010). The above data, coupled with the fact that blocking starch catabolism triggers plastid senescence (Stettler et al., 2009), suggest that further exploration of the link between starch hydrolysis and ammonium susceptibility is needed.

The biological activity of several proteins is dependent on posttranslational modifications, among which the covalent attachment of glycans is highly prevalent. Dol-P-Man is implicated as a Man donor in the formation of N-glycoproteins, GPI-anchored proteins and arabinogalactan proteins, which are collectively predicted to represent more than 50% of proteins in *Arabidopsis*. Our data reveal that the synthesis of Dol-P-Man is mediated by three genes that encode DPMS1, DPMS2, and DPMS3 proteins, which assemble into a functional DPMS complex where DPMS1 represents the catalytic module. We show through loss of function and overexpression studies that DPMS1, in contrast with DPMS2 or DPMS3, exerts a broad influence on various aspects of *Arabidopsis* development, including root growth, seed architecture, vascular bundle organization, activation of UPR, induction of chlorosis, and sensitivity to ammonium. Given the growing implications of glycan-linked proteins in a wide range of biological functions, including receptor-ligand interaction, signal transduction, pathogenesis, endocytosis, and cell wall organization, further understanding of how Dol-P-Man impacts the glycome network and influences plant development remain important challenges for future work.

## METHODS

### Plant Materials and Growth Conditions

*Arabidopsis thaliana* plants in the Columbia background were usually grown on MS basal salt medium (M0221; Duchefa Biochemie) containing Suc (3%) and agar (8 g/L) as described previously (Bouvier et al., 2006). To test the effect of ammonium on plant growth, the seeds were germinated on MS medium (M 0238, modification number 4, NH<sub>4</sub><sup>+</sup> free; Duchefa Biochemie) supplemented with increasing NH<sub>4</sub><sup>+</sup> concentrations as indicated in the text. Alternatively, the seeds were grown on damp soil (Bouvier et al., 2006). For tunicamycin treatment, 21-d-old MS-grown (M 0238) *Arabidopsis* plants were transferred to the same liquid MS medium containing 5 μg/mL of tunicamycin for 5 h. Leaves from 2-month-old *Nicotiana benthamiana* plants grown under greenhouse conditions were used.

### Screening of T-DNA Insertion Mutants

The T-DNA insertion mutants of *Arabidopsis* (Columbia ecotypes) corresponding to the different genes were obtained. The following lines were used: GK\_767A10 (*dpms1-1*), SALK\_030487C (*dpms1-2*), SALK\_111777 (*dpms2-1*), SALK\_111779 (*dpms2-2*), SALK\_037120 (*dpms3-1*), SALK\_051037 (*dpms3-2*), SALK\_006421 (*srd5a3-like1-1*), and SALK\_113221 (*srd5a3-like1-2*). Segregation analysis (resistant:susceptible, 3:1) based on the antibiotic resistance, genomic PCR, and RT-PCR analyses using appropriate primers (see Supplemental Table 2 online) were performed to characterize the homozygous lines as described previously (Bouvier et al., 2006). Antibodies raised against recombinant *Arabidopsis* DPMS1, as described previously (Bouvier et al., 2006), were also used for immunological analysis of *dpms1* mutants.

### Histological and Microscopy Analysis

Hand-cut section of *Arabidopsis* stems from wild and mutant plants were treated with phloroglucinol-HCl (0.1% phloroglucinol in 20% HCl from VWR International) for 2 min to stain lignin. Sections were washed with water and examined under a light microscope. The seed surface morphology was analyzed using scanning electron microscopy without any special sample preparation. Mature seeds were collected and directly placed into a Hitachi Tabletop microscope TM-1000 and examined by scanning electron microscopy at an operating voltage of 15 kV. Images were taken using the software supplied by manufacturer.

### Molecular Cloning and Plant Transformation

Primers used to generate the different constructs are listed in Supplemental Table 2 online. For lines overexpressing *DPMS1*, we cloned *DPMS1* cDNA into the ImpactVector 1.1 (Plant Research International) previously digested by *NcoI* and *BglII*. The resulting expression cassette was subcloned (*AscI*-*PacI*) in the binary vector *pBIN Plus* (Plant Research International) before transformation into *Agrobacterium tumefaciens* strain GV3101 and infiltration according to the floral dip procedure and kanamycin selection (Clough and Bent, 1998). For the complementation of *dpms* mutant plants, genomic fragments corresponding to the coding region were individually inserted into the *XbaI* site of the pCAMBIA-1300 vector (<http://www.cambia.org/daisy/cambia/materials/vectors>). T1 plants were selected on MS containing hygromycin. For the *DPMS1*-RNAi construct, we made use of a short inverted repeat of *DPMS1* cDNA (hairpin). Two 87-bp fragments amplified from the *DPMS1* cDNA containing *NcoI* and *BglII* at their 5' termini and *EcoRI* at their 3' termini were digested by *EcoRI*, self-ligated, and digested with *NcoI* and *BglII*. The 194-bp fragment was then ligated to the *NcoI*-*BglII* sites of the ImpactVector 1.1. Transient expression assays were performed using *GFP-DPMS1*, *GFP-DPMS2*, *GFP-DPMS3*, and *At-RNLB2* inserted into the *NcoI* site of pCATsGFP vector before insertion into the binary vector (Bouvier et al., 2006). Transient coexpression of *GFP-DPMS1*, *GFP-*

*DPMS2*, and *GFP-DPMS3* in *N. benthamiana* was performed as described previously using 1:1:1 or 1:1 mixtures of *Agrobacterium* GV3101 harboring the GFP constructs (Mialoundama et al., 2009). In parallel, control leaves were infected with *Agrobacterium* harboring an empty vector. The leaves were harvested 72 h after inoculation and processed for SDS-PAGE and immunoblot analysis using anti-DPMS1 and anti-GFP antibodies (Invitrogen) (used at a 1/2500 dilution).

### Gene Expression Analysis

Relative expression of the different genes was evaluated by quantitative PCR using total RNA isolated from the different plant lines using the NucleoSpin RNA kit (Macherey-Nagel) according to the manufacturer's instructions. Three independent plants were used, and quantitative PCR was performed in triplicate from each biological sample. The different gene-specific primers were designed using Probe Finder software (Roche) and are listed in Supplemental Table 2 online. Two micrograms of total RNA were reverse transcribed in a total volume of 40 μL with 2.5 μM oligo(dT)<sub>20</sub>, 0.5 mM deoxynucleotide triphosphate, and 400 units of Superscript III reverse transcriptase (Invitrogen). One microliter of cDNA was quantified using gene-specific primers in a total volume of 10 μL SYBR Green I Master Mix (Roche) in a Light Cycler 480 II apparatus (Roche) according to the manufacturer's instructions. The cycle threshold (Ct) was used to determine the relative expression level using the 2<sup>-ΔΔCt</sup> method (Livak and Schmittgen, 2001). The relative expression level of each gene was analyzed using GenEx Pro software (MultiD analyses; [www.gene-quantification.de/datan.html](http://www.gene-quantification.de/datan.html)) after normalization by reference with the internal references TIP41-LIKE and *GAPDH*.

### Subcellular Localization of DPMS

Transient Expression of GFP-tagged *DPMSs* in *Arabidopsis* leaves were performed using the above-described *GFP-DPMS1*, *GFP-DPMS2*, and *GFP-DPMS3* constructs as described previously (Eastmond et al., 2010). The reticulon-like protein *At-RTNLB2* fused with GFP (*GFP-AtRTNLB2*) was used as an ER marker protein (Nziengui et al., 2007). Leaf tissues expressing the individual GFP constructs were analyzed by confocal microscopy as described previously (Bouvier et al., 2006).

### Recombinant DPMS Proteins and in Vitro Reconstitution of DPMS Complex

DPMS cDNAs were subcloned into a *pBAD-TOPO* vector using the *pBAD/TOPO* Thio fusion vector (Invitrogen) using the primers listed in Supplemental Table 2 online, as described previously (Bouvier et al., 2006). After induction with 0.001% Ara, the bacterial cells containing the recombinant His-tagged DPMS proteins were disrupted using a French press in the presence of lysis buffer A (50 mM Tris-HCl, pH 7.6, 1% digitonin, 0.1% dodecylmaltoside, 50 mM KCl, and 10 mM imidazole). The resulting mixture was centrifuged at 10,000g for 15 min. The supernatant was mixed with nitrilotriacetic acid (Ni-NTA) agarose affinity chromatography (Qiagen). Ni-NTA agarose beads were rinsed three times batchwise with buffer B (50 mM Tris-HCl, pH 7.6, 50 mM KCl, 0.1% digitonin, and 10 mM imidazole) to remove nonspecifically bound proteins. The washed beads (200 mL) were subsequently mixed with 1 mL buffer C, which contained 50 mM Tris-HCl, pH 7.6, 50 mM KCl, and 5 mg/mL of liposomes prepared from identical amounts of phosphatidylcholine and phosphatidylethanolamine using a Vibra Cell sonicator instrument (Sonics and Materials) as described previously (Maeda et al., 2000). The His tag-coupled proteins were eluted from the Ni-NTA agarose beads with buffer C containing 150 mM imidazole. In the reconstitution assay, identical amounts of recombinant *DPMS1*, *DPMS2*, or *DPMS3* in buffer C were combined. The purified proteins were used for enzyme assay and SDS-PAGE analysis.

### In-Gel Digestion and NanoLC-MS/MS Analysis of DPMS Component Proteins

The gel bands were predigested using a MassPREP automated system (Waters). The gel slices were washed twice in 25 mM ammonium hydrogen carbonate and acetonitrile. The disulfide bonds were subsequently reduced in 10 mM dithiothreitol (Sigma-Aldrich) at 57°C, and then the Cys residues were alkylated in 55 mM iodoacetamide (Sigma-Aldrich). Following dehydration with acetonitrile, the proteins were cleaved in the gel using 125 ng of modified porcine trypsin (Promega), 125 ng bovine pancreas chymotrypsin (Roche), or 130 ng of *Pseudomonas fragi* AspN (Promega). The peptides were extracted sequentially with 1% trifluoroacetic acid in acetonitrile/water (60:40 v/v) and 100% acetonitrile before concentration using a Speed Vac system. The extracts were analyzed using an ultra-high-pressure LC system (NanoAcquity; Waters) coupled to a high-resolution nanoelectrospray-quadrupole-time of flight type mass spectrometer (Maxis; Bruker Daltonics). The samples were trapped on a 20 × 0.18-mm, 5- $\mu$ m Symmetry C18 precolumn (Waters), and the peptides were separated on a nanoACQUITY UPLC BEH130 C18 column (Waters; 75  $\mu$ m × 200 mm, 1.7  $\mu$ m particle size). 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 5  $\mu$ L/min with 99% of solvent A and 1% of solvent B. Elution was performed at a flow rate of 450 nL/min using 6 to 35% gradient (solvent A to B) over 9 min at 45°C followed by 90% (solvent B) over 1 min before the reconditioning the column using 99% of solvent A over 6 min. Maxis equipped with a nanospray was operated in the positive ion mode. Mass calibration of the TOF was achieved prior to analysis from  $m/z$  50 to 2200 using Tunemix (Agilent Technologies). Online correction of this calibration was performed using two Tunemix ions ( $m/z$  299.2945 and 922.0098) as recalibration masses. The capillary voltage was set to 4500 V and the end plate offset to -500 V. For MS/MS experiments, the system was operated with automatic switching between MS and MS/MS modes. The MS scan was performed over an  $m/z$  range of [50;2200] at 0.2 s/scan. The MS/MS scan time was automatically set depending on the intensity of the selected precursor peak on the MS spectrum (from 0.1 to 1.4 s/scan over the same  $m/z$  range). The three most abundant peptides, preferably doubly and triply charged ions, were selected on each MS spectrum for further isolation and collision-induced dissociation using an optimized collision energy depending on the charge state and the  $m/z$  of the ion. The selected peptides are then excluded for 0.3 min. The complete system was fully controlled by Compass Hystar (Bruker). The MS data collected during the analysis were processed and converted to mascot generic files (mgf) using Compass Data Analysis 4 software (Bruker). The mgf peak list files were then submitted to the Mascot search engine (Matrix Sciences; version 2.3.01) in a local server. Searches were performed against a composite target decoy database generated using internally developed tools (<http://msda.u-strasbg.fr>) and containing the National Center for Biotechnology Information protein sequences of *Arabidopsis* and common contaminants (trypsin and human keratins) downloaded on the 23rd of September 2010. Searches were performed with a tolerance of 5 ppm for precursor ions and 0.02 D for fragment ions, allowing a maximum of one missed cleavage site for trypsin and chymotrypsin and three missed cleavages sites for AspN. Carbamidomethylation of Cys residues, oxidation of Met residues, and acetylation of protein N-terminal residues were searched as variable modifications. The spectra that yielded identifications were manually inspected.

### Purification of Dolichols and Assay of DPMS Activity

*Ginkgo biloba* seeds obtained from the botanical garden (Strasbourg, France) were used to extract and purify dolichols with chain length of 14 to 18 isoprene residues, according to published procedures (Tateyama et al., 1999). The resulting dolichol fraction was further purified by preparative

HPLC using a Zorbax C18 column eluted using a linear gradient of solvent A (methanol/propan-2-ol/water, 12:8:1 v/v/v) to solvent B (hexane/propan-2-ol, 7:3 v/v) (Skorupińska-Tudek et al., 2003).

Dol-P-Man synthase was synthesized by coupling the synthesis of Dol-P to that of Dol-P-Man. First, purified Dol<sub>14</sub> (10  $\mu$ g) was incubated at 25°C for 10 min with (15  $\mu$ g) purified recombinant polyprenyl-dolichol kinase cloned from *Streptococcus mutans* DGK (Hartley et al., 2008) in a reaction mixture (100  $\mu$ L) containing 50 mM Tris-HCl, pH 7.6, 2 mM [ $\gamma$ -<sup>32</sup>P]ATP (10 mCi/mmol; Hartmann Analytic), 1 mM DTT, 5 mM MgCl<sub>2</sub>, and 0.1% Tween 80. Upon completion of this step, 50  $\mu$ L of liposomes containing purified DPMS protein components (equivalent to 20  $\mu$ g) as specified in the text were added to the reaction mixture supplemented with 2 mM GDP-Man, and the reaction was allowed to proceed for 30 min. Unlabeled Dol-P-Man was prepared under the same condition except nonradioactive ATP (2 mM) was used in the incubation medium. In the case of *Arabidopsis* microsomal membranes, the leaves were homogenized in the medium containing 50 mM Tris-HCl, pH 7.6, 0.25 M Suc, 1 mM EDTA, 0.5% polyvinylpyrrolidone, and 5 mM 2-mercaptoethanol. The homogenate was centrifuged at 10,000g for 15 min, and the resulting supernatant was centrifuged at 100,000g for 1 h to pellet the microsomal membrane fraction. The microsomal membrane fraction was suspended in 50 mM Tris-HCl, pH 7.6, containing 0.25 M Suc and 1 mM DTT. Microsomal preparations equivalent to 250  $\mu$ g of protein were incubated at 25°C for 30 min in the reaction mixture containing 50 mM Tris-HCl, pH 7.6, 2 mM [ $\gamma$ -<sup>32</sup>P]ATP (10 mCi/mmol; Hartmann Analytic), 1 mM DTT, 5 mM MgCl<sub>2</sub>, Dol<sub>14</sub> (10  $\mu$ g), 0.1% Tween 80, and 2 mM GDP-Man for 30 min. The reaction products were extracted using dichloromethane/methanol (2:1 v/v). Following centrifugation, the dichloromethane phase was adsorbed onto a mini DEAE-cellulose column prepared as described previously (Camara and Monéger, 1977). The column was sequentially eluted with dichloromethane and dichloromethane/methanol (7:3 v/v), and the fraction containing the Dol-P-Man was eluted with dichloromethane/methanol (7:3 v/v) containing ammonia (0.28 N) and 50 mM ammonium acetate. The eluate was concentrated and subjected to thin layer chromatography (TLC) on silica gel plate developed with dichloromethane/methanol/water (10:10:3 v/v/v), and the position of Dol<sub>14</sub>-P-Man (retention factor = 0.7) was detected by autoradiography. Radioactivity incorporated in the products was quantified by a liquid scintillation counter. For incubation with nonradioactive ATP, the Dol-P-Man band was scraped and eluted with dichloromethane/methanol (2:1 v/v) before ultra-high-pressure liquid chromatography coupled to tandem mass spectrometry (UPLC-MS/MS) analysis.

### UPLC and MS Analysis of Dol-P-Man

Characterization of Dol<sub>14</sub>-P-Man recovered from the DEAE cellulose column was performed using UPLC-MS/MS. All analyses were performed using a Waters Quattro Premier XE equipped with an ESI source and coupled to an Acquity UPLC system (Waters) with diode array detector. UV spectra were recorded from 200 to 500 nm. Chromatographic separation was achieved using an Acquity UPLC BEH C<sub>18</sub> column (100 × 2.1 mm, 1.7  $\mu$ m; Waters), coupled to an Acquity UPLC BEH C<sub>18</sub> precolumn (2.1 × 5 mm, 1.7  $\mu$ m; Waters). The temperature of the column oven was 40°C, and the injection volume was 3  $\mu$ L. Eluents were methanol/acetonitrile/water (60/30/9.99, v/v/v), basified with 1 mM ammonium acetate and 0.01% piperidine (solvent A) and ethanol 99.99%; basified with 1 mM ammonium acetate and 0.03% piperidine (solvent B). The flow rate was 0.31 mL/min. Chromatographic conditions were as follows: 100% (A) for 1 min; gradient from 100% (A) to 100% (B) for 5 min, 100% (B) and an isocratic run of 5 min at 100% of B, followed by a gradient from 100% (B) to 100% (A) for 1 min to the initial conditions that was maintained for 3 min. The total run time was 15 min. Nitrogen generated from pressurized air in a N<sub>2</sub>G nitrogen generator (Mistral) was used as the drying and nebulizing gas. The nebulizer gas flow was set to ~50 L/h, and the desolvation gas flow to 700 L/h. The interface

temperature was set at 300°C and the source temperature at 120°C. The capillary voltage was set at 3 kV, and the cone voltage was optimized for each Dol-P-Man depending on the length of the isoprenoid skeleton. Full-scan and selected ion recording analyses were used to determine the *m/z* of Dol-P-Man. Fragmentation was performed by collision-induced dissociation with argon at  $1.0 \times 10^{-4}$  mbar. The collision energy was optimized using daughter scan monitoring and MRM. MS conditions for Dol<sub>1,4</sub>-P-Man were set after optimization as follows: polarity ES<sup>-</sup>, capillary 3 kV, cone 40 V. Low mass and high mass resolution were 13 for both mass analyzers, ion energies 1 and 2 were 0.5 V, entrance and exit potential were 2 and 1 V, and detector (multiplier) gain was 650 V. Collision-induced dissociation of deprotonated parent ions was accomplished with collision energy of 10 and 15 V. Daughter scan monitoring, neutral loss, and MRM permitted the identification of Dol<sub>1,4</sub>-P-Man. The prominent compound was identified as a deprotonated parent ion [M-H<sub>2</sub>O-H]<sup>-</sup>, and the predominant daughter fragment ions 1094 and 1052 corresponding to partial and total Man loss were obtained by daughter scan and subsequently used for MRM transitions. The combination of chromatographic retention time, parent mass, and the two more intense daughter ions analysis was used to selectively monitor Dol<sub>1,4</sub>-Man-P-Man (1196 > 1094) and (1196 > 1052). Data acquisition and analysis were performed with the MassLynx software (ver.4.1) running under Windows XP professional on a Pentium PC.

#### Affinity Interactions of DPMS Protein Components

Total leaf proteins were prepared from *N. benthamiana* leaves expressing various combinations of GFP-DPMS1, GFP-DPMS2, and GFP-DPMS3 constructs. Leaf tissues were ground into a fine powder in liquid nitrogen before extraction with the buffer (50 mM Tris-HCl, pH 7.6, 2 mM PMSF, 0.2% Triton X-100, and 3% digitonin). Anti-DPMS1-Sepharose beads were prepared using dimethylpiperimidate (Schneider et al., 1982). Proteins derived from leaf extracts were centrifuged, and the supernatant was incubated for 6 h at 4°C to 100 μL of anti-DPMS1-Sepharose beads in the presence of the incubation buffer (50 mM Tris-HCl, pH 7.6, 0.1 M KCl, 2 mM PMSF, and 0.3% digitonin). Immunoabsorbed complexes were washed by centrifugation using the same buffer containing 0.25 M KCl before elution of bound proteins with 0.1 M Gly, pH 2.5. The eluate was neutralized to pH 7.6 and subjected to SDS-PAGE and immunoblot analyses using anti-GFP (Invitrogen) as described previously (Bouvier et al., 2006).

The pull-down assay was performed using DPMS3 as bait to pull down putative DPMS1 and DPMS2 preys. Streptavidin agarose beads (Thermo Scientific) (200 μL) were incubated for 30 min at 25°C with biotin-labeled DPMS3 (250 μg) prepared using the Pierce pull-down biotinylated protein:protein interaction kit (Thermo Scientific). Purified Rubisco protein (Sigma-Aldrich) was used as negative control bait. Next, total *Escherichia coli* lysates (500 μg) obtained after French press disruption and centrifugation of bacterial cells expressing recombinant DPMS1 and DPMS2 were incubated for 2 h at 6°C with biotinylated DPMS3 coupled to streptavidin agarose in a medium containing 50 mM Tris-HCl, pH 7.6, 3% digitonin, 2 mM PMSF, 0.2% Triton X-100, and 0.1 M KCl. The beads were washed by centrifugation with 50 mM Tris-HCl, pH 7.6, buffer containing 2 mM PMSF, 0.3% digitonin, and 0.25 M KCl to remove nonspecifically bound proteins. DPMS3 interacting proteins were eluted with the acidic pH 2.8 buffer.

#### Affinodetection of Glycoproteins

Total proteins were extracted using 25 mM Tris, pH 8.3, containing 250 mM Gly and 0.1% SDS and subjected to SDS-PAGE analysis before electroblot transfer onto nitrocellulose membranes. Concanavalin A binding proteins were detected using jack bean (*Canavalia ensiformis*) Concanavalin A-peroxidase (Sigma-Aldrich) and Amersham enhanced chemiluminescence reagent. PDI and PDCB1 were used as markers for N-glycoproteins and GPI-anchored proteins and probed using anti-PDI

(used at a 1/2500 dilution) (Lukowitz et al., 2001) and anti-PDCB1 (used at a 1/2500 dilution) (Simpson et al., 2009) antibodies.

#### Ascorbic Acid Analysis

*Arabidopsis* leaves were pulverized in liquid nitrogen, and the resulting powder was homogenized in 1 N HClO<sub>4</sub>. Following centrifugation and neutralization with K<sub>2</sub>CO<sub>3</sub>, the supernatant was saved for ascorbate determination as described previously (Foyer et al., 1983).

#### GMPase Activity

Plant leaves were pulverized in liquid nitrogen, and the resulting powder was homogenized in the extraction buffer containing 50 mM Tris-HCl, pH 7.6, and 2 mM 2-mercaptoethanol. The homogenate was centrifuged at 10,000g for 10 min, and the supernatant (100 μg protein) was used for enzyme assay in a reaction mixture (200 μL final volume) containing 50 mM Tris-HCl, pH 7.6, 2 mM 2-mercaptoethanol, 5 mM MgCl<sub>2</sub>, 2 mM GTP, and 2 mM D-Man. The reaction was performed at 30°C for 1 h. The reaction was quenched by boiling, and GDP-Man was analyzed by HPLC as described previously (Albermann et al., 2000). Protein concentration was determined using the Pierce bicinchoninic protein assay kit (Pierce).

#### Chlorophyll Determination

Total chlorophyll content was determined in 80% acetone extracts (Lichtenthaler and Wellburn, 1983).

#### Accession Numbers

Sequence data from this article can be found in the Arabidopsis Genome Initiative databases under the following accession numbers: At-DPMS1 (*Arabidopsis*; At1g20575), Os-DPMS1 (*Oryza sativa*; NP\_001051751), Sp-DPMS1 (*Schizosaccharomyces pombe*; CAB11700), Hs-DPMS1 (*Homo sapiens*; EAW75608), Rn-DPMS1 (*Rattus norvegicus*; NP\_001100014), Tb-DPMS1 (*Trypanosoma brucei*; AJ866775), Lm-DPMS1 (*Leishmania mexicana*; AJ131960), Sc-DPMS1 (*Saccharomyces cerevisiae*; AAB68116), At-DPMS2 (At1g74340), Os-DPMS2 (NP\_001172795), Hs-DPMS2 (AAH15233), Rn-DPMS2 (NP\_062125) Sp-DPMS2 (NP\_595676), At-DPMS3 (At1g48140), Os-DPMS3 (EAY87536), Hs-DPMS3 (BAA96291), Rn-DPMS3 (NP\_001102801), Cd-DPMS3 (*Candida dubliniensis*; XP\_002419846), At-SRD5A3-like1 (At2g16530), At-SRD5A3-like2 (At1g72590), Hs-SRD5A3 (NM\_024592), At-RTNLB2 (At4g11220), TIP41-LIKE (At4g34270), *GAPDH* (At1g3440), and α-Tubulin (At1g04820).

#### Supplemental Data

The following materials are available in the online version of this article.

**Supplemental Figure 1.** Multiple Sequence Alignment of DPMS1 with Homologous Proteins.

**Supplemental Figure 2.** Multiple Sequence Alignment of DPMS2 with Homologous Proteins.

**Supplemental Figure 3.** Multiple Sequence Alignment of DPMS3 with Homologous Proteins.

**Supplemental Figure 4.** Analysis of Recombinant DPMS Proteins Expressed in *E. coli*.

**Supplemental Figure 5.** Neutral Loss Analysis of Dol<sub>1,4</sub>-P-Man.

**Supplemental Figure 6.** Characterization of *Arabidopsis dpms* Mutants and Transgenic Lines.

**Supplemental Figure 7.** Chromatographic Analysis of Dol<sub>1,4</sub>-P-Man Synthesized by Microsomal Membranes Isolated from Wild-Type

*Arabidopsis*, *dpms* Mutants, Complemented Mutants, and *DPMS1*-RNAi Lines.

**Supplemental Figure 8.** Ammonium Sensitivity of *dpms1* Compared with Wild-Type Plants.

**Supplemental Figure 9.** Characterization and Ammonium Sensitivity of *srd5a3-llke1* Mutants.

**Supplemental Figure 10.** Multiple Sequence Alignment of SRD5A3 with Homologous Proteins.

**Supplemental Figure 11.** Affinodetection Analysis of Total Leaf Proteins Using Concanavalin A.

**Supplemental Table 1.** In-Gel Digestion and NanoLC-MS/MS of DPMS3-Interacting Proteins.

**Supplemental Table 2.** Primer Sequences Used for Cloning, Genotyping, and Expression Studies.

## ACKNOWLEDGMENTS

This work was supported by the Agence Nationale de la Recherche Program (Contract ANR-09-BLAN-0067-02) and by a doctoral fellowship from the Research Program of the Ministry of Science and Technology of Indonesia to N.J. We thank Meredith D. Hartley and Barbara Imperiali from the Massachusetts Institute of Technology for the kind gift of *E. coli* overexpressing *Streptococcus mutans* polyprenyl-dolichol kinase. We thank Andrew J. Maule from the John Innes Center for anti-PDCB1 antibodies and Chris R. Somerville from the Carnegie Institution for PDI antibodies. We thank A. Bailly, M. Kerneis, S. Staerck, and R. Wagner for taking care of the plants. We thank the Salk Institute Genomic Analysis Laboratory, the Nottingham Arabidopsis Stock Centre, and the GABI-KAT collection for providing the *Arabidopsis* T-DNA insertion mutants. The UPLC-MS/MS analysis was financed by the Université de Strasbourg, the Centre National de la Recherche Scientifique, the Région Alsace, and Tepral Company. The Région Alsace is acknowledged for financing the Maxis mass spectrometer.

Received January 25, 2011; revised April 19, 2011; accepted May 1, 2011; published May 10, 2011.

## REFERENCES

- Albermann, C., Distler, J., and Piepersberg, W.** (2000). Preparative synthesis of GDP-beta-L-fucose by recombinant enzymes from enterobacterial sources. *Glycobiology* **10**: 875–881.
- Apweiler, R., Hermjakob, H., and Sharon, N.** (1999). On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim. Biophys. Acta* **1473**: 4–8.
- Ashida, H., Maeda, Y., and Kinoshita, T.** (2006). DPM1, the catalytic subunit of dolichol-phosphate mannose synthase, is tethered to and stabilized on the endoplasmic reticulum membrane by DPM3. *J. Biol. Chem.* **281**: 896–904.
- Baksi, K., Tavárez-Pagán, J.J., Martínez, J.A., and Banerjee, D.K.** (2008). Unique structural motif supports mannosylphospho dolichol synthase: An important angiogenesis regulator. *Curr. Drug Targets* **9**: 262–271.
- Banerjee, D.K., Carrasquillo, E.A., Hughey, P., Schutzbach, J.S., Martínez, J.A., and Baksi, K.** (2005). In vitro phosphorylation by cAMP-dependent protein kinase up-regulates recombinant *Saccharomyces cerevisiae* mannosylphosphodolichol synthase. *J. Biol. Chem.* **280**: 4174–4181.
- Banerjee, D.K., Kousvelari, E.E., and Baum, B.J.** (1987). cAMP-mediated protein phosphorylation of microsomal membranes increases mannosylphosphodolichol synthase activity. *Proc. Natl. Acad. Sci. USA* **84**: 6389–6393.
- 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. USA* **104**: 11676–11681.
- Barker, A.V.** (1999). Ammonium accumulation and ethylene evolution by tomato infected with root-knot nematode and grown under different regimes of plant nutrition. *Commun. Soil Sci. Plant Anal.* **30**: 175–182.
- Barker, A.V., Volk, R.J., and Jackson, W.A.** (1966). Root environment acidity as a regulatory factor in ammonium assimilation by the bean plant. *Plant Physiol.* **41**: 1193–1199.
- Barth, C., Gouzd, Z.A., Steele, H.P., and Imperio, R.M.** (2010). A mutation in GDP-mannose pyrophosphorylase causes conditional hypersensitivity to ammonium, resulting in *Arabidopsis* root growth inhibition, altered ammonium metabolism, and hormone homeostasis. *J. Exp. Bot.* **61**: 379–394.
- Boisson, M., Gomord, V., Audran, C., Berger, N., Dubreucq, B., Granier, F., Lerouge, P., Faye, L., Caboche, M., and Lepiniec, L.** (2001). *Arabidopsis* glucosidase I mutants reveal a critical role of N-glycan trimming in seed development. *EMBO J.* **20**: 1010–1019.
- Borner, G.H., Lilley, K.S., Stevens, T.J., and Dupree, P.** (2003). Identification of glycosylphosphatidylinositol-anchored proteins in *Arabidopsis*. A proteomic and genomic analysis. *Plant Physiol.* **132**: 568–577.
- Bouvier, F., Linka, N., Isner, J.C., Mutterer, J., Weber, A.P., and Camara, B.** (2006). *Arabidopsis* SAMT1 defines a plastid transporter regulating plastid biogenesis and plant development. *Plant Cell* **18**: 3088–3105.
- Britto, D.T., and Kronzucker, H.J.** (2002). NH<sub>4</sub><sup>+</sup> toxicity in higher plants: A critical review. *J. Plant Physiol.* **159**: 567–584.
- Burda, P., Jakob, C.A., Beinhauer, J., Hegemann, J.H., and Aebi, M.** (1999). Ordered assembly of the asymmetrically branched lipid-linked oligosaccharide in the endoplasmic reticulum is ensured by the substrate specificity of the individual glycosyltransferases. *Glycobiology* **9**: 617–625.
- Camara, B., and Monéger, R.** (1977). Les lipides du fruit vert et du fruit mûr de Poivron (*Capsicum annuum* L.). *Physiol. Vég.* **15**: 711–722.
- Cantagrel, V., et al.** (2010). SRD5A3 is required for converting poly-prenol to dolichol and is mutated in a congenital glycosylation disorder. *Cell* **142**: 203–217.
- Cao, Y., Glass, A.D., and Crawford, N.M.** (1993). Ammonium inhibition of *Arabidopsis* root growth can be reversed by potassium and by auxin resistance mutations *aux1*, *axr1*, and *axr2*. *Plant Physiol.* **102**: 983–989.
- Carson, D.D., Farrar, J.D., Laidlaw, J., and Wright, D.A.** (1990). Selective activation of the N-glycosylation apparatus in uteri by estrogen. *J. Biol. Chem.* **265**: 2947–2955.
- Che, P., Bussell, J.D., Zhou, W., Estavillo, G.M., Pogson, B.J., and Smith, S.M.** (2010). Signaling from the endoplasmic reticulum activates brassinosteroid signaling and promotes acclimation to stress in *Arabidopsis*. *Sci. Signal.* **3**: ra69.
- Clough, S.J., and Bent, A.F.** (1998). Floral dip: A simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16**: 735–743.
- Colussi, P.A., Taron, C.H., Mack, J.C., and Orlean, P.** (1997). Human and *Saccharomyces cerevisiae* dolichol phosphate mannose synthases represent two classes of the enzyme, but both function in *Schizosaccharomyces pombe*. *Proc. Natl. Acad. Sci. USA* **94**: 7873–7878.
- Conklin, P.L., Pallanca, J.E., Last, R.L., and Smirnov, N.** (1997).

- L-ascorbic acid metabolism in the ascorbate-deficient *Arabidopsis* mutant *vtc1*. *Plant Physiol.* **115**: 1277–1285.
- Debono, A., Yeats, T.H., Rose, J.K., Bird, D., Jetter, R., Kunst, L., and Samuels, L.** (2009). *Arabidopsis* LTPG is a glycosylphosphatidylinositol-anchored lipid transfer protein required for export of lipids to the plant surface. *Plant Cell* **21**: 1230–1238.
- Deyoung, B.J., and Clark, S.E.** (2008). BAM receptors regulate stem cell specification and organ development through complex interactions with *CLAVATA* signaling. *Genetics* **180**: 895–904.
- Eastmond, P.J., Quettier, A.L., Kroon, J.T., Craddock, C., Adams, N., and Slabas, A.R.** (2010). Phosphatidic acid phosphohydrolase 1 and 2 regulate phospholipid synthesis at the endoplasmic reticulum in *Arabidopsis*. *Plant Cell* **22**: 2796–2811.
- Ellis, M., Egelund, J., Schultz, C.J., and Bacic, A.** (2010). Arabinogalactan-proteins: Key regulators at the cell surface? *Plant Physiol.* **153**: 403–419.
- Faye, L., and Chrispeels, M.J.** (1985). Characterization of N-linked oligosaccharides by affinity blotting with concanavalin A-peroxidase and treatment of the blots with glycosidases. *Anal. Biochem.* **149**: 218–224.
- Ferguson, M.A.J., Kinoshita, T., and Hart, G.W.** (2009). Glycosylphosphatidylinositol anchors. In *Essentials of Glycobiology*, A. Varki, R.D. Cummings, J.D. Esko, H.H. Freeze, P. Stanlet, C.R. Bertozzi, G.W. Hart, and M.E. Etzler, eds (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press), pp. 143–161.
- Foulquier, F., Harduin-Lepers, A., Duvet, S., Marchal, I., Mir, A.M., Delannoy, P., Chirat, F., and Cacan, R.** (2002). The unfolded protein response in a dolichyl phosphate mannose-deficient Chinese hamster ovary cell line points out the key role of a demannosylation step in the quality-control mechanism of N-glycoproteins. *Biochem. J.* **362**: 491–498.
- Foyer, C.H., Rowell, J., and Walker, D.** (1983). Measurements of the ascorbate content of spinach leaf protoplasts and chloroplasts during illumination. *Planta* **157**: 239–244.
- Frank, C.G., and Aebi, M.** (2005). ALG9 mannosyltransferase is involved in two different steps of lipid-linked oligosaccharide biosynthesis. *Glycobiology* **15**: 1156–1163.
- Furmanek, A., and Hofsteenge, J.** (2000). Protein C-mannosylation: Facts and questions. *Acta Biochim. Pol.* **47**: 781–789.
- Garg, N., Tarleton, R.L., and Mensa-Wilmot, K.** (1997). Proteins with glycosylphosphatidylinositol (GPI) signal sequences have divergent fates during a GPI deficiency. GPIs are essential for nuclear division in *Trypanosoma cruzi*. *J. Biol. Chem.* **272**: 12482–12491.
- Garrett, T.A., Guan, Z., and Raetz, C.R.** (2007). Analysis of ubiquinones, dolichols, and dolichol diphosphate-oligosaccharides by liquid chromatography-electrospray ionization-mass spectrometry. *Methods Enzymol.* **432**: 117–143.
- Gillmor, C.S., Lukowitz, W., Brininstool, G., Sedbrook, J.C., Hamann, T., Poindexter, P., and Somerville, C.** (2005). Glycosylphosphatidylinositol-anchored proteins are required for cell wall synthesis and morphogenesis in *Arabidopsis*. *Plant Cell* **17**: 1128–1140.
- Haeuptle, M.A., and Hennet, T.** (2009). Congenital disorders of glycosylation: an update on defects affecting the biosynthesis of dolichol-linked oligosaccharides. *Hum. Mutat.* **30**: 1628–1641.
- Handford, M.G., Sicilia, F., Brandizzi, F., Chung, J.H., and Dupree, P.** (2004). *Arabidopsis thaliana* expresses multiple Golgi-localised nucleotide-sugar transporters related to GONST1. *Mol. Genet. Genomics* **272**: 397–410.
- Hanson, S.R., Culyba, E.K., Hsu, T.L., Wong, C.H., Kelly, J.W., and Powers, E.T.** (2009). The core trisaccharide of an N-linked glycoprotein intrinsically accelerates folding and enhances stability. *Proc. Natl. Acad. Sci. USA* **106**: 3131–3136.
- Hart, G.W., and Copeland, R.J.** (2010). Glycomics hits the big time. *Cell* **143**: 672–676.
- Hartley, M.D., Larkin, A., and Imperiali, B.** (2008). Chemoenzymatic synthesis of polyprenyl phosphates. *Bioorg. Med. Chem.* **16**: 5149–5156.
- Häweker, H., Rips, S., Koiwa, H., Salomon, S., Saijo, Y., Chinchilla, D., Robatzek, S., and von Schaewen, A.** (2010). Pattern recognition receptors require N-glycosylation to mediate plant immunity. *J. Biol. Chem.* **285**: 4629–4636.
- Helenius, A., and Aebi, M.** (2001). Intracellular functions of N-linked glycans. *Science* **291**: 2364–2369.
- Henquet, M., Lehle, L., Schreuder, M., Rouwendal, G., Molthoff, J., Helsper, J., van der Krol, S., and Bosch, D.** (2008). Identification of the gene encoding the alpha1,3-mannosyltransferase (*ALG3*) in *Arabidopsis* and characterization of downstream n-glycan processing. *Plant Cell* **20**: 1652–1664.
- Hirschi, K.D.** (2003). Insertional mutants: A foundation for assessing gene function. *Trends Plant Sci.* **8**: 205–207.
- Hong, Z., Jin, H., Fitchette, A.C., Xia, Y., Monk, A.M., Faye, L., and Li, J.** (2009). Mutations of an alpha1,6 mannosyltransferase inhibit endoplasmic reticulum-associated degradation of defective brassinosteroid receptors in *Arabidopsis*. *Plant Cell* **21**: 3792–3802.
- Hummel, E., Osterrieder, A., Robinson, D.G., and Hawes, C.** (2010). Inhibition of Golgi function causes plastid starch accumulation. *J. Exp. Bot.* **61**: 2603–2614.
- Ilgoutz, S.C., Mullin, K.A., Southwell, B.R., and McConville, M.J.** (1999). Glycosylphosphatidylinositol biosynthetic enzymes are localized to a stable tubular subcompartment of the endoplasmic reticulum in *Leishmania mexicana*. *EMBO J.* **18**: 3643–3654.
- Iwata, Y., and Koizumi, N.** (2005). An *Arabidopsis* transcription factor, AtbZIP60, regulates the endoplasmic reticulum stress response in a manner unique to plants. *Proc. Natl. Acad. Sci. USA* **102**: 5280–5285.
- Jones, M.B., Rosenberg, J.N., Betenbaugh, M.J., and Krag, S.S.** (2009). Structure and synthesis of polyisoprenoids used in N-glycosylation across the three domains of life. *Biochim. Biophys. Acta* **1790**: 485–494.
- Jozwiak, A.P., Skorupska-Tudek, K., Maluszynska, J., Kania, W., Danikiewicz, W., and Swiezewska, E.** (2009). Polyisoprenoid alcohols are accumulated in *Arabidopsis* roots. *Chem. Phys. Lip.* **160**: S47.
- Kajiura, H., Seki, T., and Fujiyama, K.** (2010). *Arabidopsis thaliana* ALG3 mutant synthesizes immature oligosaccharides in the ER and accumulates unique N-glycans. *Glycobiology* **20**: 736–751.
- Katiyar-Agarwal, S., Zhu, J., Kim, K., Agarwal, M., Fu, X., Huang, A., and Zhu, J.K.** (2006). The plasma membrane Na<sup>+</sup>/H<sup>+</sup> antiporter SOS1 interacts with RCD1 and functions in oxidative stress tolerance in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* **103**: 18816–18821.
- Keller, R., Renz, F.S., and Kossmann, J.** (1999). Antisense inhibition of the GDP-mannose pyrophosphorylase reduces the ascorbate content in transgenic plants leading to developmental changes during senescence. *Plant J.* **19**: 131–141.
- Kierszniowska, S., Seiwert, B., and Schulze, W.X.** (2009). Definition of *Arabidopsis* sterol-rich membrane microdomains by differential treatment with methyl-beta-cyclodextrin and quantitative proteomics. *Mol. Cell. Proteomics* **8**: 612–623.
- Kim, Y.U., Ashida, H., Mori, K., Maeda, Y., Hong, Y., and Kinoshita, T.** (2007). Both mammalian PIG-M and PIG-X are required for growth of GPI14-disrupted yeast. *J. Biochem.* **142**: 123–129.
- Kirkby, E.A.** (1968). Influence of ammonium and nitrate nutrition on the cation-anion balance and nitrogen and carbohydrate metabolism of white mustard plants grown in dilute nutrient solutions. *Soil Sci.* **105**: 133–141.
- Kitajima, A., Asatsuma, S., Okada, H., Hamada, Y., Kaneko, K., Nanjo, Y., Kawagoe, Y., Toyooka, K., Matsuoka, K., Takeuchi, M., Nakano, A., and Mitsui, T.** (2009). The rice alpha-amylase glycoprotein is targeted from the Golgi apparatus through the secretory pathway to the plastids. *Plant Cell* **21**: 2844–2858.

- Koizumi, N., Ujino, T., Sano, H., and Chrispeels, M.J. (1999). Overexpression of a gene that encodes the first enzyme in the biosynthesis of asparagine-linked glycans makes plants resistant to tunicamycin and obviates the tunicamycin-induced unfolded protein response. *Plant Physiol.* **121**: 353–361.
- Kruszewska, J.S., Butterweck, A.H., Kurzatkowski, W., Migdalski, A., Kubicek, C.P., and Palamarczyk, G. (1999). Overexpression of the *Saccharomyces cerevisiae* mannosylphosphodolichol synthase-encoding gene in *Trichoderma reesei* results in an increased level of protein secretion and abnormal cell ultrastructure. *Appl. Environ. Microbiol.* **65**: 2382–2387.
- Kruszewska, J.S., Saloheimo, M., Migdalski, A., Orlean, P., Penttilä, M., and Palamarczyk, G. (2000). Dolichol phosphate mannose synthase from the filamentous fungus *Trichoderma reesei* belongs to the human and *Schizosaccharomyces pombe* class of the enzyme. *Glycobiology* **10**: 983–991.
- Lalanne, E., Honys, D., Johnson, A., Borner, G.H., Lilley, K.S., Dupree, P., Grossniklaus, U., and Twell, D. (2004). SETH1 and SETH2, two components of the glycosylphosphatidylinositol anchor biosynthetic pathway, are required for pollen germination and tube growth in *Arabidopsis*. *Plant Cell* **16**: 229–240.
- Lamani, E., et al. (2006). Structural studies and mechanism of *Saccharomyces cerevisiae* dolichyl-phosphate-mannose synthase: insights into the initial step of synthesis of dolichyl-phosphate-linked oligosaccharide chains in membranes of endoplasmic reticulum. *Glycobiology* **16**: 666–678.
- Lanquar, V., Loqué, D., Hörmann, F., Yuan, L., Bohner, A., Engelsberger, W.R., Lalonde, S., Schulze, W.X., von Wirén, N., and Frommer, W.B. (2009). Feedback inhibition of ammonium uptake by a phospho-dependent allosteric mechanism in *Arabidopsis*. *Plant Cell* **21**: 3610–3622.
- Lecca, M.R., Wagner, U., Patrignani, A., Berger, E.G., and Hennet, T. (2005). Genome-wide analysis of the unfolded protein response in fibroblasts from congenital disorders of glycosylation type-I patients. *FASEB J.* **19**: 240–242.
- Lefeber, D.J., et al. (2009). Deficiency of Dol-P-Man synthase subunit DPM3 bridges the congenital disorders of glycosylation with the dystroglycanopathies. *Am. J. Hum. Genet.* **85**: 76–86.
- Lehle, L., Strahl, S., and Tanner, W. (2006). Protein glycosylation, conserved from yeast to man: a model organism helps elucidate congenital human diseases. *Angew. Chem. Int. Ed. Engl.* **45**: 6802–6818.
- Lehrman, M.A. (2001). Oligosaccharide-based information in endoplasmic reticulum quality control and other biological systems. *J. Biol. Chem.* **276**: 8623–8626.
- Lichtenthaler, H.K., and Wellburn, A.R. (1983). Determinations of total carotenoids and chlorophylls *a* and *b* of leaf extracts in different solvents. *Biochem. Soc. Trans.* **11**: 591–592.
- Liu, J.X., and Howell, S.H. (2010). Endoplasmic reticulum protein quality control and its relationship to environmental stress responses in plants. *Plant Cell* **22**: 2930–2942.
- Liu, Q., Kasuga, M., Sakuma, Y., Abe, H., Miura, S., Yamaguchi-Shinozaki, K., and Shinozaki, K. (1998). Two transcription factors, DREB1 and DREB2, with an EREBP/AP2 DNA binding domain separate two cellular signal transduction pathways in drought- and low-temperature-responsive gene expression, respectively, in *Arabidopsis*. *Plant Cell* **10**: 1391–1406.
- Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. *Methods* **25**: 402–408.
- Lommel, M., and Strahl, S. (2009). Protein O-mannosylation: Conserved from bacteria to humans. *Glycobiology* **19**: 816–828.
- Lukowitz, W., Nickle, T.C., Meinke, D.W., Last, R.L., Conklin, P.L., and Somerville, C.R. (2001). *Arabidopsis* *cyt1* mutants are deficient in a mannose-1-phosphate guanylyltransferase and point to a requirement of N-linked glycosylation for cellulose biosynthesis. *Proc. Natl. Acad. Sci. USA* **98**: 2262–2267.
- Maeda, Y., and Kinoshita, T. (2008). Dolichol-phosphate mannose synthase: Structure, function and regulation. *Biochim. Biophys. Acta* **1780**: 861–868.
- Maeda, Y., Tanaka, S., Hino, J., Kangawa, K., and Kinoshita, T. (2000). Human dolichol-phosphate-mannose synthase consists of three subunits, DPM1, DPM2 and DPM3. *EMBO J.* **19**: 2475–2482.
- Maeda, Y., Tomita, S., Watanabe, R., Ohishi, K., and Kinoshita, T. (1998). DPM2 regulates biosynthesis of dolichol phosphate-mannose in mammalian cells: correct subcellular localization and stabilization of DPM1, and binding of dolichol phosphate. *EMBO J.* **17**: 4920–4929.
- Marshner, P. (1995). *Mineral Nutrition of Higher Plants*. (London: Academic Press).
- Martínez, I.M., and Chrispeels, M.J. (2003). Genomic analysis of the unfolded protein response in *Arabidopsis* shows its connection to important cellular processes. *Plant Cell* **15**: 561–576.
- Mazhari-Tabrizi, R., Eckert, V., Blank, M., Müller, R., Mumberg, D., Funk, M., and Schwarz, R.T. (1996). Cloning and functional expression of glycosyltransferases from parasitic protozoans by heterologous complementation in yeast: the dolichol phosphate mannose synthase from *Trypanosoma brucei brucei*. *Biochem. J.* **316**: 853–858.
- Meyer, U., Benghezal, M., Imhof, I., and Conzelmann, A. (2000). Active site determination of Gpi8p, a caspase-related enzyme required for glycosylphosphatidylinositol anchor addition to proteins. *Biochemistry* **39**: 3461–3471.
- Mialoundama, A.S., Heintz, D., Debayle, D., Rahier, A., Camara, B., and Bouvier, F. (2009). Abscisic acid negatively regulates elicitor-induced synthesis of capsidiol in wild tobacco. *Plant Physiol.* **150**: 1556–1566.
- Molinari, M. (2007). N-glycan structure dictates extension of protein folding or onset of disposal. *Nat. Chem. Biol.* **3**: 313–320.
- Nanjo, Y., Oka, H., Ikarashi, N., Kaneko, K., Kitajima, A., Mitsui, T., Muñoz, F.J., Rodríguez-López, M., Baroja-Fernández, E., and Pozueta-Romero, J. (2006). Rice plastidial N-glycosylated nucleotide pyrophosphatase/phosphodiesterase is transported from the ER-golgi to the chloroplast through the secretory pathway. *Plant Cell* **18**: 2582–2592.
- Ng, D.T., Spear, E.D., and Walter, P. (2000). The unfolded protein response regulates multiple aspects of secretory and membrane protein biogenesis and endoplasmic reticulum quality control. *J. Cell Biol.* **150**: 77–88.
- Nozaki, M., Ohishi, K., Yamada, N., Kinoshita, T., Nagy, A., and Takeda, J. (1999). Developmental abnormalities of glycosylphosphatidylinositol-anchor-deficient embryos revealed by Cre/loxP system. *Lab. Invest.* **79**: 293–299.
- Nziengui, H., Bouhidel, K., Pillon, D., Der, C., Marty, F., and Schoefs, B. (2007). Reticulon-like proteins in *Arabidopsis thaliana*: Structural organization and ER localization. *FEBS Lett.* **581**: 3356–3362.
- Orlean, P. (1992). Enzymes that recognize dolichols participate in three glycosylation pathways and are required for protein secretion. *Biochem. Cell Biol.* **70**: 438–447.
- Orlean, P., Albright, C., and Robbins, P.W. (1988). Cloning and sequencing of the yeast gene for dolichol phosphate mannose synthase, an essential protein. *J. Biol. Chem.* **263**: 17499–17507.
- Orlean, P., and Menon, A.K. (2007). Thematic review series: Lipid posttranslational modifications. GPI anchoring of protein in yeast and mammalian cells, or: How we learned to stop worrying and love glycosphospholipids. *J. Lipid Res.* **48**: 993–1011.
- Orłowski, J., Machula, K., Janik, A., Zdebska, E., and Palamarczyk, G. (2007). Dissecting the role of dolichol in cell wall assembly in the

- yeast mutants impaired in early glycosylation reactions. *Yeast* **24**: 239–252.
- Pattison, R.J., and Amtmann, A.** (2009). N-glycan production in the endoplasmic reticulum of plants. *Trends Plant Sci.* **14**: 92–99.
- Pettolino, F., Liao, M.L., Ying, Z., Mau, S.L., and Bacic, A.** (2006). Structure, function and cloning of arabinogalactan-proteins (AGPs): An overview. *Foods Food Ingrid. J. Jpn.* **211**: 12–25.
- Qin, C., Qian, W., Wang, W., Wu, Y., Yu, C., Jiang, X., Wang, D., and Wu, P.** (2008). GDP-mannose pyrophosphorylase is a genetic determinant of ammonium sensitivity in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **105**: 18308–18313.
- Radhamony, R.N., and Theg, S.M.** (2006). Evidence for an ER to Golgi to chloroplast protein transport pathway. *Trends Cell Biol.* **16**: 385–387.
- Raven, J.A., and Smith, F.A.** (1976). Nitrogen assimilation and transport in the vascular land plants in relation to intracellular pH regulation. *New Phytol.* **76**: 415–431.
- Roudier, F., Fernandez, A.G., Fujita, M., Himmelspach, R., Borner, G. H., Schindelman, G., Song, S., Baskin, T.I., Dupree, P., Wasteneys, G.O., and Benfey, P.N.** (2005). COBRA, an *Arabidopsis* extracellular glycosyl-phosphatidyl inositol-anchored protein, specifically controls highly anisotropic expansion through its involvement in cellulose microfibril orientation. *Plant Cell* **17**: 1749–1763.
- Rush, J.S., Shelling, J.G., Zingg, N.S., Ray, P.H., and Waechter, C.J.** (1993). Mannosylphosphoryldolichol-mediated reactions in oligosaccharide-P-P-dolichol biosynthesis. Recognition of the saturated alpha-isoprene unit of the mannosyl donor by pig brain mannosyltransferases. *J. Biol. Chem.* **268**: 13110–13117.
- Rush, J.S., van Leyen, K., Ouerfelli, O., Wolucka, B., and Waechter, C.J.** (1998). Transbilayer movement of Glc-P-dolichol and its function as a glucosyl donor: protein-mediated transport of a water-soluble analog into sealed ER vesicles from pig brain. *Glycobiology* **8**: 1195–1205.
- Saijo, Y., Tintor, N., Lu, X., Rauf, P., Pajeroska-Mukhtar, K., Häweker, H., Dong, X., Robatzek, S., and Schulze-Lefert, P.** (2009). Receptor quality control in the endoplasmic reticulum for plant innate immunity. *EMBO J.* **28**: 3439–3449.
- Sanyal, S., and Menon, A.K.** (2010). Stereoselective transbilayer translocation of mannosyl phosphoryl dolichol by an endoplasmic reticulum flippase. *Proc. Natl. Acad. Sci. USA* **107**: 11289–11294.
- Schachtman, D.P., Schroeder, J.I., Lucas, W.J., Anderson, J.A., and Gaber, R.F.** (1992). Expression of an inward-rectifying potassium channel by the *Arabidopsis KAT1* cDNA. *Science* **258**: 1654–1658.
- Schenk, B., Fernandez, F., and Waechter, C.J.** (2001). The ins(ide) and out(side) of dolichyl phosphate biosynthesis and recycling in the endoplasmic reticulum. *Glycobiology* **11**: 61R–70R.
- Schjoerring, J.K., Husted, S., Mäck, G., and Mattsson, M.** (2002). The regulation of ammonium translocation in plants. *J. Exp. Bot.* **53**: 883–890.
- Schneider, C., Newman, R.A., Sutherland, D.R., Asser, U., and Greaves, M.F.** (1982). A one-step purification of membrane proteins using a high efficiency immunomatrix. *J. Biol. Chem.* **257**: 10766–10769.
- Seifert, G.J., and Roberts, K.** (2007). The biology of arabinogalactan proteins. *Annu. Rev. Plant Biol.* **58**: 137–161.
- Sheeley, D.M., and Reinhold, V.N.** (1998). Structural characterization of carbohydrate sequence, linkage, and branching in a quadrupole ion trap mass spectrometer: Neutral oligosaccharides and N-linked glycans. *Anal. Chem.* **70**: 3053–3059.
- Simpson, C., Thomas, C., Findlay, K., Bayer, E., and Maule, A.J.** (2009). An *Arabidopsis* GPI-anchor plasmodesmal neck protein with callose binding activity and potential to regulate cell-to-cell trafficking. *Plant Cell* **21**: 581–594.
- Skorupińska-Tudek, K., Bienkowski, T., Olszowska, O., Furmanowa, M., Chojnacki, T., Danikiewicz, W., and Swiezewska, E.** (2003). Divergent pattern of polyisoprenoid alcohols in the tissues of *Coluria geoides*: A new electrospray ionization MS approach. *Lipids* **38**: 981–990.
- Skorupińska-Tudek, K., et al.** (2008). Contribution of the mevalonate and methylerythritol phosphate pathways to the biosynthesis of dolichols in plants. *J. Biol. Chem.* **283**: 21024–21035.
- Sprung, I., Carmès, L., Watt, G.M., and Flitsch, S.L.** (2003). Synthesis of novel acceptor substrates for the dolichyl phosphate mannose synthase from yeast. *ChemBioChem* **4**: 319–332.
- Stettler, M., Eicke, S., Mettler, T., Messerli, G., Hörtensteiner, S., and Zeeman, S.C.** (2009). Blocking the metabolism of starch breakdown products in *Arabidopsis* leaves triggers chloroplast degradation. *Mol. Plant* **2**: 1233–1246.
- Sugiyama, E., DeGasperi, R., Urakaze, M., Chang, H.M., Thomas, L.J., Hyman, R., Warren, C.D., and Yeh, E.T.** (1991). Identification of defects in glycosylphosphatidylinositol anchor biosynthesis in the Thy-1 expression mutants. *J. Biol. Chem.* **266**: 12119–12122.
- Tateyama, S., Wititsuwannakul, R., Wititsuwannakul, D., Sagami, H., and Ogura, K.** (1999). Dolichols of rubber plant, ginkgo and pine. *Phytochemistry* **51**: 11–15.
- ten Hoopen, F., Cuin, T.A., Pedas, P., Hegelund, J.N., Shabala, S., Schjoerring, J.K., and Jahn, T.P.** (2010). Competition between uptake of ammonium and potassium in barley and *Arabidopsis* roots: molecular mechanisms and physiological consequences. *J. Exp. Bot.* **61**: 2303–2315.
- Tokunaga, F., Brostrom, C., Koide, T., and Arvan, P.** (2000). Endoplasmic reticulum (ER)-associated degradation of misfolded N-linked glycoproteins is suppressed upon inhibition of ER mannosidase I. *J. Biol. Chem.* **275**: 40757–40764.
- Tomita, S., Inoue, N., Maeda, Y., Ohishi, K., Takeda, J., and Kinoshita, T.** (1998). A homologue of *Saccharomyces cerevisiae Dpm1p* is not sufficient for synthesis of dolichol-phosphate-mannose in mammalian cells. *J. Biol. Chem.* **273**: 9249–9254.
- van Berkel, M.A., Rieger, M., te Heesen, S., Ram, A.F., van den Ende, H., Aebi, M., and Klis, F.M.** (1999). The *Saccharomyces cerevisiae CWH8* gene is required for full levels of dolichol-linked oligosaccharides in the endoplasmic reticulum and for efficient N-glycosylation. *Glycobiology* **9**: 243–253.
- van der Hoorn, R.A., Wulff, B.B., Rivas, S., Durrant, M.C., van der Ploeg, A., de Wit, P.J., and Jones, J.D.** (2005). Structure-function analysis of cf-9, a receptor-like protein with extracytoplasmic leucine-rich repeats. *Plant Cell* **17**: 1000–1015.
- Villarejo, A., et al.** (2005). Evidence for a protein transported through the secretory pathway en route to the higher plant chloroplast. *Nat. Cell Biol.* **7**: 1224–1231.
- Wall, M.E.** (1939). The role of potassium in plants. I. Effects of varying amounts of potassium on nitrogenous, carbohydrate and mineral metabolism in the tomato plant. *Soil Sci.* **47**: 143–161.
- Wang, M.Y., Siddiqi, M.Y., and Galss, A.D.M.** (1996). Interactions between K<sup>+</sup> and NH<sub>4</sub><sup>+</sup>: Effects on ion uptake by rice roots. *Plant Cell Environ.* **19**: 1037–1046.
- Watanabe, R., Murakami, Y., Marmor, M.D., Inoue, N., Maeda, Y., Hino, J., Kangawa, K., Julius, M., and Kinoshita, T.** (2000). Initial enzyme for glycosylphosphatidylinositol biosynthesis requires PIG-P and is regulated by DPM2. *EMBO J.* **19**: 4402–4411.
- Weerapana, E., and Imperiali, B.** (2006). Asparagine-linked protein glycosylation: From eukaryotic to prokaryotic systems. *Glycobiology* **16**: 91R–101R.
- White, P.J.** (1996). The permeation of ammonium through a voltage-independent K<sup>+</sup> channel in the plasma membrane of rye roots. *J. Membr. Biol.* **152**: 89–99.
- Wilson, I.B., Taylor, J.P., Webberley, M.C., Turner, N.J., and Flitsch,**

- S.L.** (1993). A novel mono-branched lipid phosphate acts as a substrate for dolichyl phosphate mannose synthetase. *Biochem. J.* **295**: 195–201.
- Yan, A., Wu, E., and Lennarz, W.J.** (2005). Studies of yeast oligosaccharyl transferase subunits using the split-ubiquitin system: Topological features and in vivo interactions. *Proc. Natl. Acad. Sci. USA* **102**: 7121–7126.
- Zhang, H., Ohyama, K., Boudet, J., Chen, Z., Yang, J., Zhang, M., Muranaka, T., Maurel, C., Zhu, J.K., and Gong, Z.** (2008). Dolichol biosynthesis and its effects on the unfolded protein response and abiotic stress resistance in *Arabidopsis*. *Plant Cell* **20**: 1879–1898.
- Zhang, M., Henquet, M., Chen, Z., Zhang, H., Zhang, Y., Ren, X., van der Krol, S., Gonneau, M., Bosch, D., and Gong, Z.** (2009). LEW3, encoding a putative alpha-1,2-mannosyltransferase (ALG11) in N-linked glycoprotein, plays vital roles in cell-wall biosynthesis and the abiotic stress response in *Arabidopsis thaliana*. *Plant J.* **60**: 983–999.
- Zhao, J., Simeone, D.M., Heidt, D., Anderson, M.A., and Lubman, D. M.** (2006). Comparative serum glycoproteomics using lectin selected sialic acid glycoproteins with mass spectrometric analysis: Application to pancreatic cancer serum. *J. Proteome Res.* **5**: 1792–1802.
- Zhu, B.C., and Laine, R.A.** (1996). Dolichyl-phosphomannose synthase from the archae *Thermoplasma acidophilum*. *Glycobiology* **6**: 811–816.

**DOLICHOL PHOSPHATE MANNOSE SYNTHASE1 Mediates the Biogenesis of Isoprenyl-Linked Glycans and Influences Development, Stress Response, and Ammonium Hypersensitivity in *Arabidopsis***

Nurul Jadid, Alexis Samba Mialoundama, Dimitri Heintz, Daniel Ayoub, Mathieu Erhardt, Jérôme Mutterer, Denise Meyer, Abdelmalek Alioua, Alain Van Dorsselaer, Alain Rahier, Bilal Camara and Florence Bouvier

*Plant Cell* 2011;23:1985-2005; originally published online May 10, 2011;  
DOI 10.1105/tpc.111.083634

This information is current as of January 10, 2013

<b>Supplemental Data</b>	<a href="http://www.plantcell.org/content/suppl/2011/05/10/tpc.111.083634.DC1.html">http://www.plantcell.org/content/suppl/2011/05/10/tpc.111.083634.DC1.html</a>
<b>References</b>	This article cites 129 articles, 73 of which can be accessed free at: <a href="http://www.plantcell.org/content/23/5/1985.full.html#ref-list-1">http://www.plantcell.org/content/23/5/1985.full.html#ref-list-1</a>
<b>Permissions</b>	<a href="https://www.copyright.com/ccc/openurl.do?sid=pd_hw1532298X&amp;issn=1532298X&amp;WT.mc_id=pd_hw1532298X">https://www.copyright.com/ccc/openurl.do?sid=pd_hw1532298X&amp;issn=1532298X&amp;WT.mc_id=pd_hw1532298X</a>
<b>eTOCs</b>	Sign up for eTOCs at: <a href="http://www.plantcell.org/cgi/alerts/ctmain">http://www.plantcell.org/cgi/alerts/ctmain</a>
<b>CiteTrack Alerts</b>	Sign up for CiteTrack Alerts at: <a href="http://www.plantcell.org/cgi/alerts/ctmain">http://www.plantcell.org/cgi/alerts/ctmain</a>
<b>Subscription Information</b>	Subscription Information for <i>The Plant Cell</i> and <i>Plant Physiology</i> is available at: <a href="http://www.aspb.org/publications/subscriptions.cfm">http://www.aspb.org/publications/subscriptions.cfm</a>

## Caractérisation du site de phosphorylation de la protéine GiP

### Contexte

La base du vivant est la division cellulaire. Ce processus hautement régulé utilise chez les cellules à noyau des complexes protéiques faisant appel aux tubulines et à tout un ensemble d'autres partenaires. Parmi ceux-ci, la protéine GiP (Gamma-tubulin Complex Protein 3 Interacting Protein) découverte tout d'abord dans la plante *Arabidopsis thaliana* (At), semble jouer un rôle particulièrement vital. Une étude bioinformatique a montré que cette protéine semble être présente chez la quasi-totalité des eucaryotes. Elle présente aussi au moins un site de phosphorylation potentiel. Dans les kinases impliquées dans la régulation du cycle cellulaire et de ces complexes à tubulines, on peut citer Aurora 1 chez les plantes. Des expériences *in vitro* utilisant les protéines AtGiP1 et AtAurora 1 en présence de gamma-ATP radioactif ont permis de montrer qu'AtGiP1 est bien un substrat de la kinase considérée. Nous avons donc proposé de vérifier que le site effectif de phosphorylation correspondait bien au site prédit par les algorithmes bioinformatiques.

### Matériels et méthodes de l'analyse de phosphorylation de la protéine GiP

Les protéines GiP d'*Arabidopsis* (AtGiP1) et humaine (HsGiP1) sont exprimées dans *E. coli* et purifiées. Chacune d'entre elles est ensuite incubée avec la kinase AtAurora1 en présence et en absence d'ATP (contrôle). Les protéines ont été aussi incubées avec de l'ATP sans la kinase pour faire un deuxième contrôle. Les protéines sont ensuite digérées en solution avec de la trypsine et les peptides de digestion sont analysés par LC-MS/MS comme indiqué dans l'extrait du manuscrit de la publication en cours de rédaction.

#### Nano LC-MS/MS analysis

Samples were analyzed using an Agilent 1200 series HPLC-Chip/MS system (Agilent Technologies, Palo Alto, CA, USA) coupled to an HCT Plus ion trap mass spectrometer (Bruker Daltonics, Bremen, Germany). The Chip contained a Zorbax 300SB-C18 (43 mm x 75  $\mu$ m, with a 5  $\mu$ m particle size) column and a Zorbax 300SB-C18 (40nL, 5 $\mu$ m) enrichment column

(Agilent Technologies). The solvent system consisted of 2 % acetonitrile and 0.1 % formic acid in water (solvent A); 2 % water and 0.1 % formic acid in acetonitrile (solvent B). 5  $\mu$ L of sample were loaded onto the enrichment column at a 3.75  $\mu$ L/min flow rate using solvent A. Elution was performed at a flow rate of 0.3  $\mu$ L/min with a 0-40% solvent B linear gradient over 9 min followed by a 70% stage (solvent B) over 2 min before reconditioning the column over 7 min. The HCT ion trap was operated in the positive ion mode. The peptides were fragmented by Collision-Induced Dissociation (CID). The MS scanning was performed in the standard enhanced resolution mode at a scan rate of 8 100 m/z per second. The mass range was 250 – 1500 m/z. The ion charge control was 100 000 and the maximum accumulation time was 200 ms. A total of three scans were averaged to obtain an MS spectrum. For tandem MS experiment the system was operated with automatic switching between MS and MS/MS modes. The three most abundant peptide ions, preferring doubly charged ones (absolute threshold of 10 000 and a relative of 5 %), were selected on each MS spectrum for further isolation (4 m/z isolation width) and fragmentation. Ions were excluded after 1 spectrum and released after 0.12 min. the MS/MS scanning was performed in the ultrascan resolution mode at a scan rate of 26 000 m/z per second. The ion charge control was set at 300 000. A total of five scans were averaged to obtain an MS/MS spectrum. The complete system was fully controlled by Chemstation (Agilent Technologies) and EsquireControl (Bruker Daltonics).

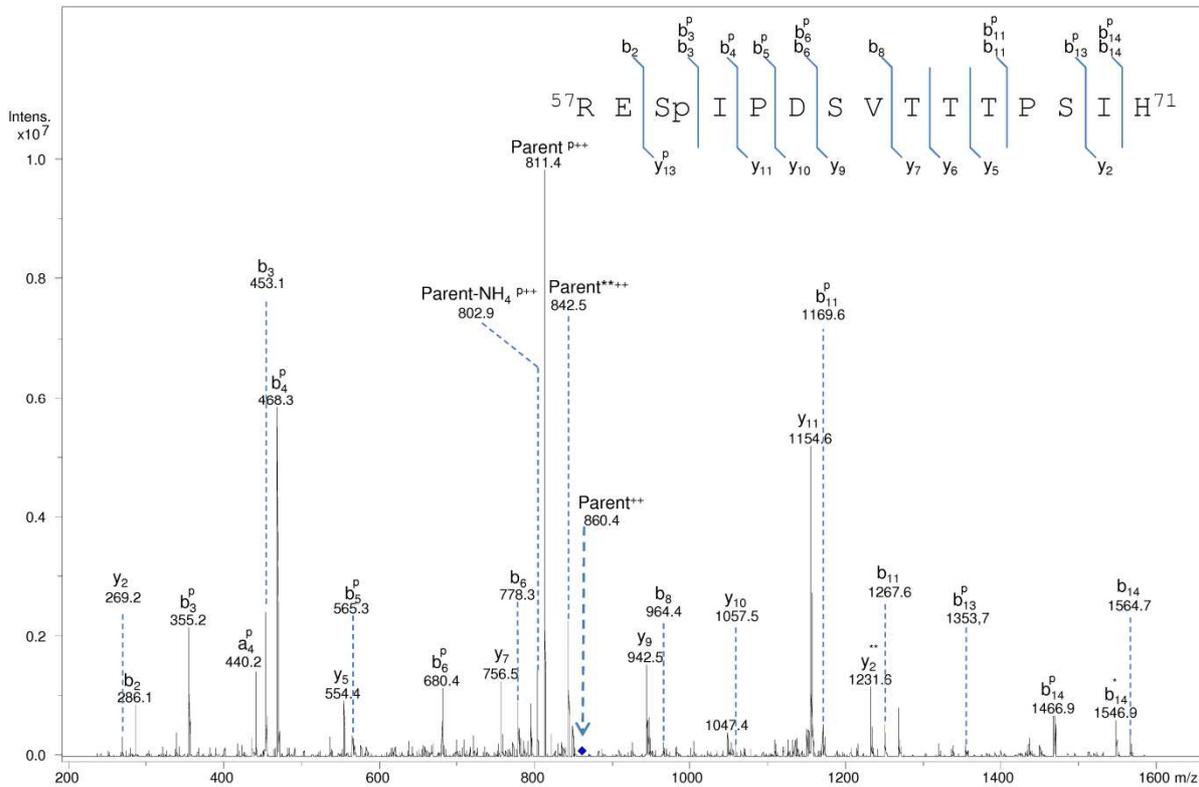
### Protein Identification

Mass Data collected during analysis were processed, converted into mgf files using DataAnalysis software (Bruker Daltonics). The mgf files were then submitted to Mascot search engine (Matrix Sciences, Boston, MA, USA) in its 2.3 version on a local server. Searches were performed against a composite target-decoy database generated using internally developed tools (<http://msda.u-strasbg.fr>) and containing the NCBI protein sequences of Arabidopsis thaliana and common contaminants (trypsin and human keratins) downloaded on the 23<sup>rd</sup> of September 2010 for the Arabidopsis samples, and against Swissprot database for human samples. Searches were performed with a 0.2 Da error tolerance on mass measurements for both precursor and fragment ions. Oxidation of methionine residues, carbamidomethylation of cysteine residues and phosphorylation of serine and threonine residues were searched as variable modifications. Up to one trypsin miss cleavage site was allowed. The spectra of the phosphorylated peptides were manually reinspected and annotated.

### Results

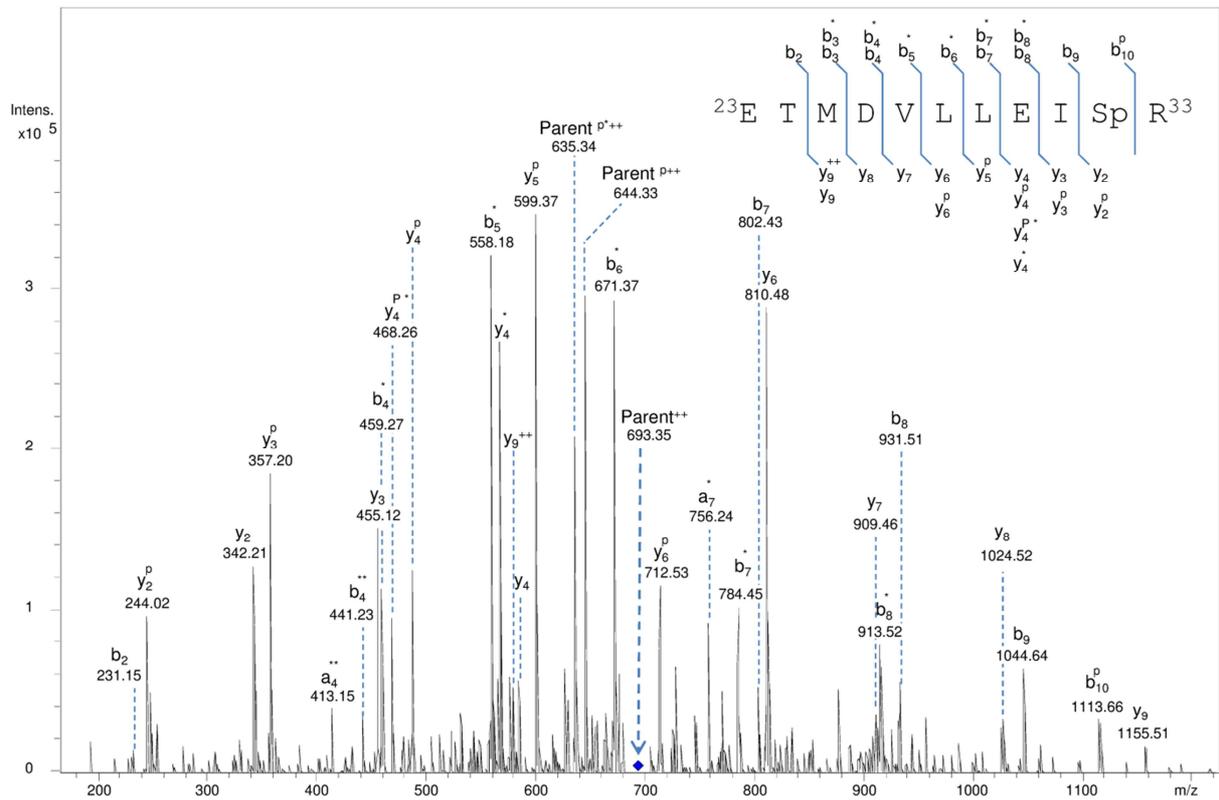
For the Arabidopsis GIP protein incubated with the kinase and ATP, the peptide R<sup>57</sup>-H<sup>71</sup> is identified with a phosphorylation on the serine 59 residue by a good quality MS/MS spectrum and a Mascot ion score of 63. This spectrum is manually reinspected and annotated providing unambiguous information for phosphorylation site assignment on the

serine residue 59 (Figure 1). The same peptide is identified without phosphorylation which suggests that both forms, phosphorylated and non phosphorylated exist. The extracted ion chromatograms of the same phosphorylated peptide in the control analysis show clearly that the phosphorylation is inexistent in the controls. This suggests that both kinase and ATP are required for the phosphorylation of GIP.



**Figure 1: Fragmentation spectrum of the protein AtGIP doubly charged phosphopeptide R57-H71. Fragment ions b<sub>3</sub> - b<sub>6</sub> permitted the unambiguous identification of the peptide with a phosphorylated S59. Fragments y<sub>2</sub> - y<sub>11</sub> provide evidence that the phosphorylation is not on the other potential sites. (p) indicates an ion with a neutral loss of 98 corresponding to H<sub>3</sub>PO<sub>4</sub>. (\*) indicates a neutral loss of a water molecule.**

For the human GIP protein incubated with the kinase and ATP, the peptide E<sup>23</sup>-R<sup>33</sup> is identified with a phosphorylation on the serine 32 residue by a good quality MS/MS spectrum and a Mascot ion score of 39. This spectrum is manually reinspected and annotated providing unambiguous information for phosphorylation site assignment on the serine residue 32. The same peptide is identified without phosphorylation which suggests that both forms, phosphorylated and non phosphorylated exist. The extracted ion chromatograms of the same phosphorylated peptide in the control analysis show clearly that the phosphorylation is inexistent in the controls. This suggests that both kinase and ATP are required for the phosphorylation of GIP.



**Figure 2 : Fragmentation spectrum of the protein HsGIP doubly charged phosphopeptide E23-R33. Fragment ions y permitted the unambiguous identification of the peptide with a phosphorylated S32. Fragment ions b provide evidence that the phosphorylation is not on the threonine 24. (p) indicates an ion with a neutral loss of 98 corresponding to H<sub>3</sub>PO<sub>4</sub>. (\*) indicates a neutral loss of a water molecule.**

## **Poster**

### Apport de la Spectrométrie de Masse Supramoléculaire pour la Caractérisation des Métalloprotéines après Séparation sur Gel d'Electrophorèse en conditions non-dénaturantes

Daniel Ayoub, Bertrand Chaumande, Agnès Hagège, Alain Van Dorsselaer, Laurence Sabatier,  
Sarah Sanglier-Cianférani

Ce poster a été récompensé par le prix du meilleur poster du congrès SMAP 2009



# Apport de la Spectrométrie de Masse Supramoléculaire pour la Caractérisation des Métalloprotéines après Séparation sur Gel d'Electrophorèse en conditions non-dénaturantes

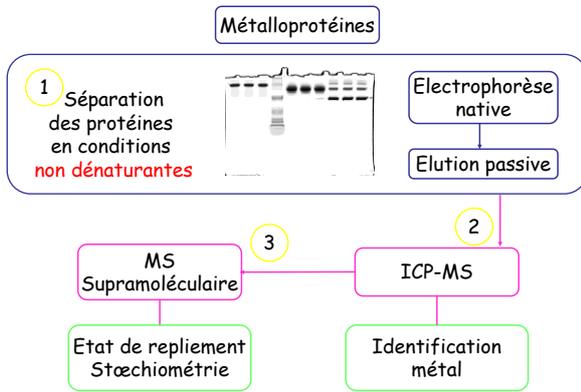
Daniel Ayoub<sup>1</sup>, Bertrand Chaumande<sup>2</sup>, Agnès Hagège<sup>2</sup>, Alain Van Dorsselaer<sup>1</sup>, Laurence Sabatier<sup>2</sup>, Sarah Sanglier-Cianféroni<sup>1</sup>

1- Laboratoire de Spectrométrie de masse Bio-Organique, IPHC-DSA, Uds, CNRS, UMR 7178; 25 Rue Becquerel F-67087 Strasbourg  
2- Laboratoire de Chimie Analytiques et Sciences Séparatives, IPHC-DSA, Uds, CNRS, UMR 7178; 25 Rue Becquerel F-67087 Strasbourg

## Introduction

Dans une cellule, on estime qu'un tiers des protéines contient au moins un élément métallique. Ces éléments sont nécessaires au bon repliement de la protéine ainsi qu'à sa fonction et son activité. C'est pourquoi, la caractérisation de ces interactions protéine-métal revêt une importance particulière. Dans ce contexte, la préservation de ces interactions lors des différentes étapes analytiques permettant la caractérisation de cet ensemble constitue encore un réel défi. Nous présentons ici une évaluation d'une nouvelle approche par gel d'électrophorèse en conditions non-dénaturantes HR-CN-PAGE pour séparer les mélanges protéiques en conservant les interactions protéine-métal. La SM supramoléculaire est utilisée ici pour vérifier le maintien et caractériser les complexes protéine-métal issus du gel HR-CN-PAGE.

## Stratégie analytique



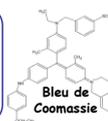
## 1 Electrophorèse sur gel en conditions non dénaturantes

### Le Blue Native Page ou BN PAGE:

Coloration au bleu de Coomassie → protéines chargées (-)

☹️ Bruit de fond élevé en ICP MS

☹️ Décoloration → dénaturation → Perte du métal.



### Le Clear Native Page ou CN PAGE:

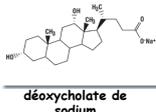
Pas de coloration.

☹️ les protéines migrent sous le seul effet de leurs charges, résolution insuffisante

### Le High resolution CN PAGE ou Hr CN PAGE:

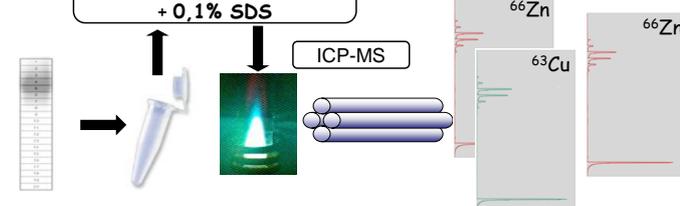
déoxycholate de sodium → Protéines chargées (-)

😊 Résolution identique à BN PAGE



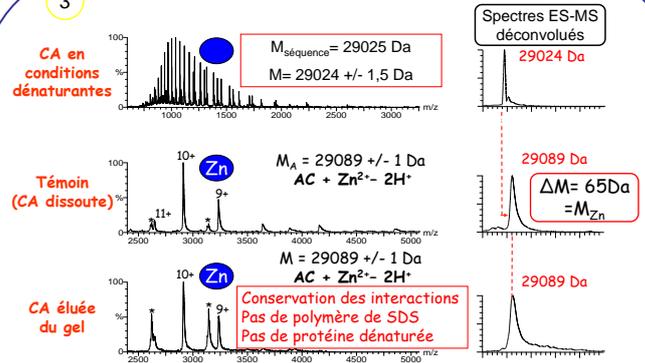
## 2 Analyse en ICP-MS exemple de la SuperOxyde Dismutase (SOD) bovine et de l'Anhydrase carbonique (CA) bovine

Elution passive dans 10mM Acétate d'ammonium + 0,1% SDS

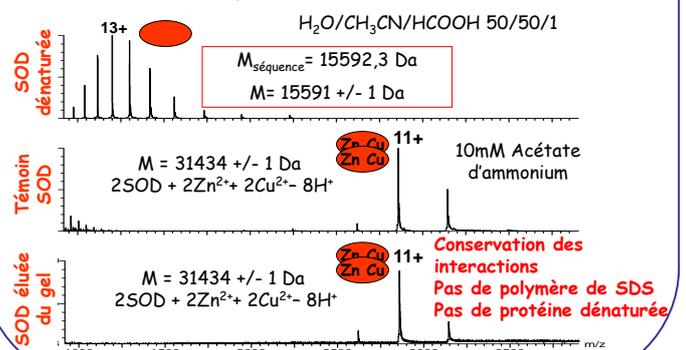


Identification du métal et de la bande contenant des métalloprotéines  
La SOD lie du cuivre et du zinc  
L'Anhydrase Carbonique lie du Zinc

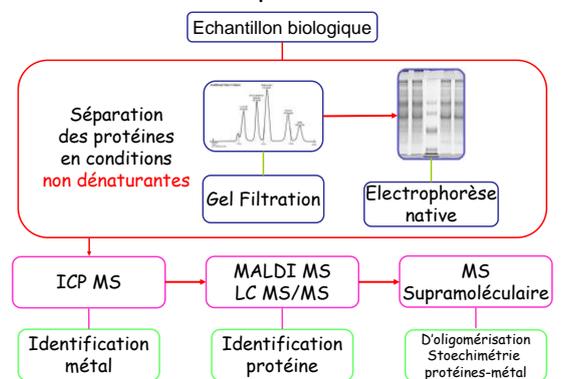
## 3 Analyse de la bande anhydrase carbonique (CA)



## Analyse de la bande de SOD



## Perspectives



## Matériel et méthodes

Les protéines sont achetées chez Sigma. Pour les protéines issues de l'élué du gel, des échanges de tampon par gel filtration (Zeba, Pierce) et dialyse (Slides A lyser, Pierce) sont effectués pour éliminer le SDS. Les expériences sont effectuées sur un ESI-TOF (LCT, Waters) équipé d'une source nanoES-Chip (Advion Biosystems) et sur un ICP-Quad (Agilent 7500ce).









## VERS UNE ÉTUDE APPROFONDIE DES PROTÉOMES : CARACTÉRISATION DES ÉXTREMITÉS N-TERMINALES DES PROTÉINES

### Résumé

Dans ce travail de thèse, nous avons développé et optimisé une stratégie originale pour la caractérisation des extrémités N-terminales des protéines et des sites de clivages protéolytiques. Elle s'appuie sur la dérivation chimique spécifique des amines N-terminales et nous l'avons adapté à différents types d'échantillons biologiques. L'application de cette stratégie dans des études en biologie nous a permis d'apporter plusieurs éléments de réponse à différentes problématiques. Nous avons ainsi caractérisé les peptides de transit des protéines mitochondriales humaines et ainsi validé/corrigé expérimentalement leurs prédictions dans les banques de données. Nous avons aussi appliqué cette stratégie à l'étude du protéome du parasite *P. falciparum*. La mise au point de la dérivation N-terminale de protéines immobilisées dans un gel SDS PAGE nous a permis d'étudier le mécanisme d'export des protéines de ce parasite vers sa cellule hôte et de déterminer le rôle des acides aminés impliqués dans cet export. Un réactif de dérivation marqué aux isotopes stables permet d'effectuer des études différentielles des processus protéolytiques que subissent les protéines. Cette stratégie quantitative a été appliquée à l'étude du protéome hépatique du rat soumis au jeûne expérimental. D'autres applications de l'analyse protéomique en biologie sont aussi présentées dans ce manuscrit.

Mots Clés : protéomique, protéolyse, clivage protéolytique, modifications post-traductionnelles, TMPP, mitochondrie, Plasmodium *falciparum*, paludisme, jeûne expérimentale.

### Abstract

In this manuscript, we describe the development and the optimization of an original strategy for the characterization of protein N-termini and protease cleavage sites. The strategy is based on specific chemical derivation of alpha-amines. We applied this method to the characterization of mitochondrial proteins' transit peptides which allowed us to experimentally validate/correct their prediction in protein databases. In another study, the strategy was applied to the analysis of the proteome of the malaria parasite Plasmodium *falciparum*. The optimization of *ingel* N-terminal derivation and its application to the study of parasite exported proteins allowed us to determine the role of implicated amino acid residues in the signaling and export mechanism of these proteins to the host cell. To enable differential studies of proteolysis, we introduced an isotope labeled derivation reagent. This quantitative method was applied in the context of a study of the rat liver proteome after experimental long-term fasting. Other applications of proteomics in biology are also presented in this manuscript.

Key words: proteomics, proteolysis, protease cleavage, post-translational modifications, TMPP, mitochondria, Plasmodium *falciparum*, malaria, long term fasting.